

**Exploring the Single-Trial-Type-Dominance-Effect in the IRAP: Developing a  
Differential Arbitrarily Applicable Relational Responding Effects (DAARRE) Model**

Martin Finn, Dermot Barnes-Holmes, and Ciara McEnteggart

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

**Correspondence should be addressed to Martin Finn, Experimental, Clinical and Health  
Psychology, Ghent University, Ghent, Belgium, [Martin.Finn@UGent.be](mailto:Martin.Finn@UGent.be)**

**Acknowledgements**

The data for the current manuscript was collected, and prepared with the support of the FWO  
Type I Odysseus Programme at Ghent University, Belgium.

## Abstract

The Implicit Relational Assessment Procedure (IRAP) has been used as a measure of implicit cognition and to analyze the dynamics of arbitrarily applicable relational responding. The current study employs the IRAP for the latter purpose. Specifically, the current research focuses on a pattern of responding observed in a previously published IRAP study that was difficult to explain using existing conceptual analyses. The pattern is referred to as the single-trial-type-dominance-effect because one of the IRAP trial-types produces an effect that is significantly larger than the other three. Based on a post-hoc explanation provided in a previously published article, the first experiment in the current series explored the impact of prior experimental experience on the single-trial-type-dominance-effect. The results indicated that the effect was larger for participants who reported high levels of experimental experience ( $M = 32.3$  previous experiments) versus those who did not ( $M = 2.5$  previous experiments). In the second experiment, participants were required to read out loud the stimuli presented on each trial and the response option they chose. The effect of experimental experience was absent but the single-trial-type-dominance-effect remained. In the third experiment, a different set of stimuli to those used in the first two was employed in the IRAP, and a significant single-trial-type-dominance-effect was no longer observed. The results obtained from the three experiments led inductively to the development of a new model of the variables involved in producing IRAP effects, the Differential Arbitrarily Applicable Relational Responding Effects (DAARRE) model, which is presented in the General Discussion.

Key words: RFT, Relational network, IRAP, trial-type, differential

The study of derived stimulus relations has been used by many behavior analysts as a conceptual basis for analyzing behaviors that appear to be closely related to human language and cognition. Perhaps the clearest and most self-conscious example of this approach is provided by Relational Frame Theory (RFT; Hayes, Barnes-Holmes, & Roche, 2001). Drawing on the seminal work of Sidman (1971; see 1994, for a book length treatment) on equivalence relations, RFT argued that the functional units of human language and cognition involve a wide range of generalized relational operants, known as relational frames, each possessing three core properties. The first property is mutual entailment and involves a bi-directional relation between two stimuli, such that if A is related to B then B is related to A. The second property is combinatorial entailment and involves three or more stimuli, such that if A is related to B and B is related to C, then A is related to C and C is related to A. The third property is the transformation of functions, which recognizes that any mutual or combinatorial entailment will involve specific behavioral functions. Thus, if A is related to B, and B acquires a mildly appetitive function, the function of A may be transformed based on the type of relation between A and B. For example, if you are told that “A is better than B”, then the appetitive function acquired by A may be larger than the appetitive function that was initially acquired by B.

According to RFT, relational frames are always under two types of contextual control. One type of contextual control is denoted as Crel, which refers to any contextual cues that determine the specific mutual and combinatorial entailed relations. In natural language, these may be words or phrases such as “bigger than”, “smaller than”, “same”, “opposite”, “different” and so on. The other type of contextual control is Cfunc, which refers to any contextual cues that select the behavioral functions that are transformed in accordance with an entailed relation. For example, if two stimuli (A and B) are entailed in a “same” relation, a Cfunc determines the specific behavioral function that “expresses” the entailed relation. If A

is an actual chocolate brownie and B is the phrase “chocolate brownie” then different Cfuncs will evoke different responses, although the entailed relation remains the same. For example, the Cfunc “tastes like” and the Cfunc “looks like” will evoke the gustatory and visual properties of chocolate brownies, respectively.

According to RFT, many of the functions of stimuli that we encounter in the natural environment may appear to be relatively basic or simple but have acquired those properties due, at least in part, to a history of relational framing. Even a simple tendency to orient more strongly towards one stimulus rather than another in your visual field may be based on relational framing. Identifying the name of your home town or city from a random list of place names may occur more quickly or strongly because it coordinates with other stimuli that control strong orienting functions (e.g., the many highly familiar stimuli that constitute your home town). Such functions may be defined as Cfunc properties because they are examples of specific stimulus functions (i.e., orienting) that are acquired based on, but are separate from, the entailed relations among the relevant stimuli. The reader should note that the use of the term “orienting” (as an example of a Cfunc property) is one that we will employ frequently later in the current article.

Since the publication of the Hayes et al. (2001) volume on RFT, a large number of studies have explored and tested the basic account (see Dymond & Roche, 2013; Hughes & Barnes-Holmes, 2016a, 2016b, 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, 2016). The development of the Implicit Relational Assessment Procedure (IRAP), which grew directly out of the theory, was instrumental in pursuing this line of research (see Hughes & Barnes-Holmes, 2011). The IRAP is a computer-based task that presents label stimuli at the top of the screen, such as

pictures of spiders and puppies, and target stimuli that appear in the middle of the screen, such as, “Scares me”, “Frightens me”, “I like it”, and “Makes me calm”. 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 crossover design that yields four different trial-types on the IRAP (in the current example: Spider-Negative; Spider-Positive; Puppy-Negative; and Puppy-Positive). 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, a spider picture and “Scares me” might appear on a given trial with the response options “Yes” and “No”, and in this case, participants would be required to confirm (i.e., pick Yes) or deny (i.e., pick No) that spiders scare me.

The IRAP operates by requiring opposing patterns of responding across successive blocks. For example, a spider picture and “Scares me” would require the response “Yes” on one block and “No” 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; this basic assumption was formalized in the context of the relational elaboration and coherence (REC) model, the details of which have been presented in numerous other articles (see Barnes-Holmes, Barnes-Holmes, Stewart, & Boles, 2010, for a detailed treatment of the REC model). According to the REC model, an individual with a history of relatively high fear responses towards spiders is more likely to emit patterns of responding on an IRAP indicative of this history. 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 history involved in the two patterns of responding. In this case, the difference score should indicate that the spider-fearful individual responded more quickly

when asked to confirm (respond “Yes”) rather than disconfirm (respond “No”) that the pictures of spiders are scary.

In concluding that an IRAP performance is indicative of specific histories of relational responding it might be assumed that each of the four trial-types of the IRAP would be equally sensitive to these histories. However, growing evidence suggests that this is not the case. Perhaps one of the clearest examples of the way in which the four trial-types of the IRAP may not be equally sensitive to pre-experimental history was provided in a series of experiments reported by Finn, Barnes-Holmes, Hussey, and Graddy (2016). Across three experiments, they employed what we will call here a Shapes-and-Colors IRAP. Across trials, the two label stimuli were “Color” and ‘Shape” and the target words were “Red”, “Green”, “Blue”, “Square”, “Circle”, and “Triangle”. As such, the IRAP involved presenting four different trial-types that could be designated as (i) *Color-Color*, (ii) *Color-Shape*, (iii) *Shape-Color* and (iv) *Shape-Shape*. The experiments they reported focused on the impact of specific types of instructions on the individual trial-type effects, the details of which are not critical here. The important finding was a persistent finding, which we will refer to as the single-trial-type-dominance-effect for the *Color-Color* trial-type. That is, even when all four trial-type effects were history-consistent, the size of the *Color-Color* trial-type effect was significantly larger than for the *Shape-Shape* trial-type. The puzzling issue about this difference is that these two trial-types share the same response option within blocks of trials (i.e., “True” during history-consistent blocks and “False” during history-inconsistent blocks) and thus the difference between the trial-types cannot be explained by a simple tendency to respond “True” more readily than “False”, which is common in natural language (Dodds et al., 2015).

In speculating about the single-trial-type-dominance-effect, Finn et al. (2016) suggested that perhaps prior experimental experience with the IRAP might play a role in generating the effect. Since the Finn et al. data were collected, our research team moved to a

different university, which runs a large research participant pool populated with volunteers who vary considerably in the number of experiments they have completed previously using a range of latency-based measures. Access to this participant pool afforded us the opportunity to explore the potential impact of prior experimental experience on the single-trial-type-dominance-effect. On the grounds of intellectual honesty, we have presented the studies here as they were conducted, and contextualized them with a narrative that reflects our thinking as we moved from one experiment to the next. Although the research may appear initially as somewhat trivial, it served to generate a new conceptual model of the key variables that appear to be involved in generating IRAP performances. Again, on the grounds of intellectual honesty, this new model is presented towards the end of the current paper rather than at the beginning (i.e., because it emerged inductively from the experimental work).

Before proceeding, it seems important to emphasize that research employing the IRAP may be separated into two categories. One category contains studies in which the IRAP has been used largely as a type of psychometric instrument (e.g., to predict some criterion variable, such as treatment outcome; e.g., Carpenter, Martinez, Vadhan, Barnes-Holmes, & Nunes, 2012). The other category contains studies that have used the IRAP as an experimental context for exploring the dynamics of relational framing (e.g., Finn et al., 2016; Maloney & Barnes-Holmes, 2016). The current study is very much in the latter category. Although it is important for the reader to bear this distinction in mind, developing an improved understanding of the dynamics of behavior that are typically produced by the IRAP may be of benefit in using it as a psychometric instrument in future research.

### **Experiment 1**

The first study aimed to test the extent to which different levels of experimental experience impacted upon the relative sizes of the four trial-types in an IRAP. Generally speaking, the methodologies employed by the various research groups at the Department of

Experimental, Clinical, and Health Psychology are latency based procedures ostensibly similar to the IRAP. Examples of these procedures include the Implicit Association Test (IAT), Affect Misattribution Procedure (AMP), variations of the Stroop task, and reaction time based procedures delivered as part of EEG studies. While not all experiments involve procedures of this kind, they are the predominant mode of investigation. In this context, amassing a sample of completely experimentally naïve individuals would have been extremely difficult. We therefore decided to compare differences in IRAP performances after minimal exposure versus repeated exposure to these tasks. In this context, it should be noted that participants sometimes complete more than one experiment within a single visit to the research laboratory, and thus it would also have been difficult to limit a “Low-Experience” group to just one or two experiments. Therefore, we operationalized experimental experience as follows; participants who had taken part in 5 or fewer experiments were categorized as Low Experience (LE) while participants who had taken part in 6 or more experiments were categorized as being High Experience (HE).

In Experiment 1, all participants completed a Shapes-and-Colors IRAP. The performance of each participant on the IRAP was established initially using a verbal feedback procedure rather than through the provision of block-specific rules (described in more detail below). The purpose of the verbal feedback procedure was to avoid providing any rules or instructions that might specify or emphasize any of the relations the IRAP was assessing because, as noted above, Finn et al. (2016) showed that such instructions may impact upon IRAP performances.

## **Method**

### **Participants**

Sixty-five individuals with an age range of 18 to 48 years ( $M = 22.6$  years) participated in the experiment in return for payment of €5. Informed consent was provided by

all participants. The experiment was approved by the ethics committee of the Faculty of Psychology and Educational Sciences at Ghent University.

## **Materials**

**IRAP.** The IRAP was presented on a standard laptop computer. The IRAP software was used to present the instructions, the stimuli, and to record responses.<sup>1</sup> Each trial presented one of two labels; “Color” or “Shape.” The label stimulus was presented along with one of twelve target stimuli. The target stimuli were all words, six denoting colors, “Red,” “Green,” “Blue,” “Pink,” “Yellow,” and “Orange” and the other six denoting shapes, “Triangle,” “Circle,” “Square,” “Rectangle,” “Octagon” and “Oval”. The following instruction appeared on screen between each block of trials “Important: During the next phase the previously correct and wrong answers are reversed. This is part of the experiment. Please try to make as few errors as possible – in other words, avoid the red X”.

**Questionnaire.** A questionnaire comprising of demographic information (age) was presented to participants after they had completed the IRAP. In addition, the questionnaire asked participants to estimate approximately the number of previous psychology experiments in which they had participated.

## **Procedure**

The experiment was conducted on a one-to-one basis in soundproof cubicles at the Department of Educational, Clinical, and Health Psychology at Ghent University. Upon entering the cubicle, participants were seated in front of a laptop which had the IRAP program initiated. The first stage of the experiment involved shaping participants’ performance on the IRAP in a manner similar to that employed in research that was conducted prior to the introduction of specific pre-block rules that were made possible with the 2012 version of the

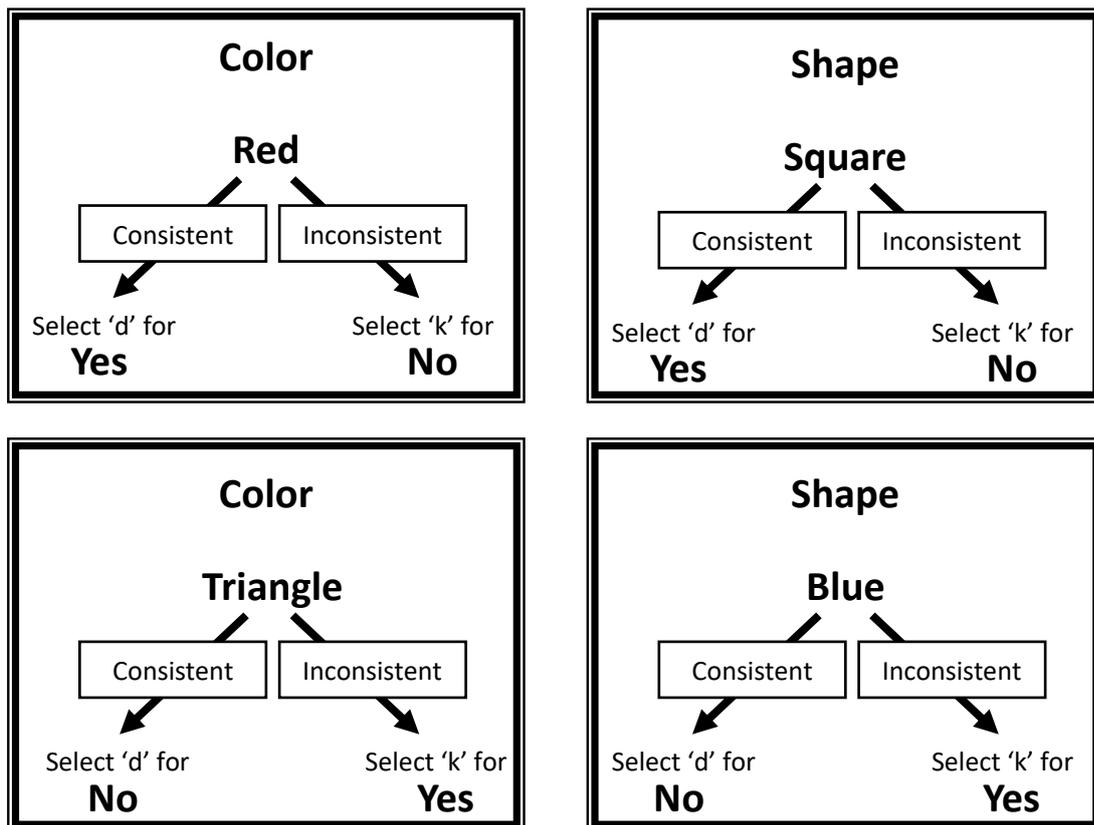
---

<sup>1</sup> All stimuli used in the current IRAP were presented to participants in Dutch. For the purposes of the current article the English translations will be used. The original Dutch versions of all on screen instructions and a full list of stimuli is available from the first author on request.

IRAP program. Before interacting with the IRAP, the experimenter described to participants what they would see during the task. Specifically, they were told that on each trial they would see a word appear at the top of the screen, a second word would appear at the middle of the screen, and that they would be required to respond to the combination of words by choosing one of two response options, “Yes” or “No” that appeared at the bottom of the screen, one on the left and one on the right. Participants were also informed that pressing the “d” key would select the response at the bottom left and pressing the “k” key would select the response option at the bottom right of the screen.

After the basic details of the task had been outlined, participants were told that their goal or objective in the task was to discover the pattern of responses required by the computer program. They were told that this meant for each trial one response would be accurate and the other inaccurate. Selecting the accurate response would allow them to progress to the next trial while selecting an inaccurate response would produce a red X below the target stimulus. At this point, participants were encouraged to press the spacebar and interact with the task, being told that all they could do on the first trial was to guess which response option was the accurate one. Following their initial response, the experimenter noted the feedback for that trial. If they emitted a response deemed inaccurate according to the program (i.e. the response produced a red X), the experimenter said that particular combination of labels and targets did not go together on this occasion. If participants emitted an accurate response the experimenter provided verbal reinforcement (e.g., saying, “well done”) and then invited them to continue with the next trial, which appeared on screen. The aim of the verbal feedback procedure was to ensure participants attained the latency and accuracy criteria required on the IRAP without specifying a formal rule or instruction (e.g. “Please respond as if shapes are shapes and colors are colors”).

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 (e.g., “Red”, “Green”, “Blue”, “Circle”, “Square”, or “Triangle”), and the two response options “Yes” and “No” 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, “Yes” and “No”) 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 1.



*Figure 1.* 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 1).

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 79 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 a maximum of 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. The test phase of the IRAP involved six test blocks 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.

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 “Yes” on *Color-Color* and *Shape-Shape* trial-types and to choose “No” on *Color-Shape* and *Shape-Color* trial-types. The contingencies deemed inconsistent required participants to respond in an opposite pattern, choosing “No” on *Color-Color* and *Shape-Shape* trial-types and choosing “Yes” 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. In other words, half of the participants 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. After participants completed the IRAP they completed the questionnaire with the experimenter.

## **Results and Discussion**

### **Questionnaire Data**

As noted previously, completing five or fewer psychology experiments was used as the criterion for defining the Low Experience (LE) group whereas completing 6 or more psychology experiments defined the High Experience (HE) group. These criteria were chosen because it was deemed impractical to attempt gathering data from a completely experimentally naïve participant group within Ghent University's participant recruitment system. In the LE group, 28 of 29 participants met the IRAP performance criteria and the estimated average number of experiments participated in was 2.3, while in the HE group 29 of 35 participants met the IRAP performance criteria and the estimated average number of experiments participated in was 32.5. Note again that the vast majority of these experiments involved exposure to some form latency-based measure.

### **IRAP Data Processing**

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 has been employed previously to avoid higher attrition rates (Leech, Barnes-Holmes, & Madden, 2016; Nicholson & Barnes-Holmes, 2012). Application of these criteria resulted in the exclusion of 7 of the 65 participants from the analyses.

Each participant who completed the current IRAP produced 24 response latencies for each test block. 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).

### **IRAP Data Analyses**

The foregoing calculations yielded four mean *D*-IRAP scores for each participant, one for each trial-type (no main effect for block-sequence was identified). The overall mean *D*-IRAP scores, divided according to level of experience, are presented in Figure 2. All eight scores 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 “Yes” more quickly than “No” 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 “No” more quickly than “Yes” 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. In comparing the two levels of experience, the HE group produced an effect for the *Color-Color* trial-type that differed dramatically from the remaining three trial-type scores; the LE group produced a broadly similar pattern but the differences between the *Color-Color* and other three trial types was far less pronounced. A 2x4 mixed repeated measures ANOVA failed to produce a main effect for experience ( $p > .6$ ), but did yield a significant interaction between trial-type and experience,  $F(3,55) = 3.39, p = .02, \eta_p^2 = .06$ . Two follow-up one-way repeated measures ANOVAs yielded a large effect size for the HE group,  $F(3,28) = 18.34, p < .001, \eta_p^2 = 0.4$ , with a much smaller effect size for the LE group  $F(3,27) = 5.82, p = .001, \eta_p^2 = 0.18$ . Four follow-up independent  $t$ -tests, one for each trial-type, were non-significant ( $ps > .1$ ). Eight one-sample  $t$ -tests indicated that each of the *D-IRAP* effects differed significantly from zero ( $ps < .03$ ). The inferential statistics therefore confirmed the descriptive analyses presented in Figure 2, in that the effect size for trial-type differences for the HE group was over twice that of the LE group.

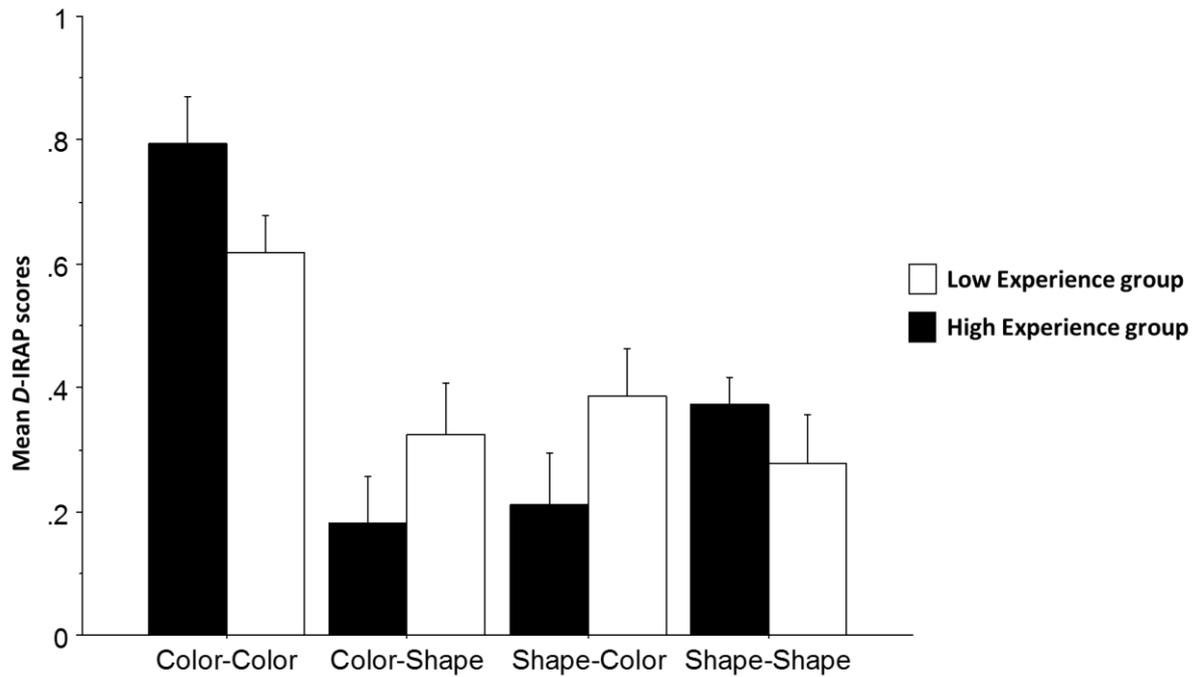


Figure 2. Mean D-IRAP scores and standard errors for the four trial-types for each of the experience groups.

How might we explain the interaction between experimental experience and trial-types on the IRAP? As noted previously, Finn et al. (2016) examined the impact of providing rules before each block of trials on the IRAP. The study revealed that rules that specified particular relations (e.g., “respond as if colors are colors and shapes are shapes”) generated more pronounced differences among the trial-types than general rules (e.g., “respond correctly to the stimuli”). Although speculative, perhaps large amounts of experience in broadly similar latency-based experimental studies facilitated the production of self-generated rules for completing the current IRAP that were more specific than those generated by participants with less experience. Or more informally, perhaps generating and following relatively specific rules for completing reaction-time tasks had been established across the many experiments completed by the HE group. Indeed, in many of these previous studies it is likely that participants had been presented with quite specific rules for completing latency-based measures, and thus they simply continued to do in the current study what they had been

trained to do in very similar contexts (i.e., in the same laboratory). If this was the case, then the relatively large differences between the *Color-Color* trial-type and the other three trial-types for the HE groups would be expected (because experience functioned as a proxy for the provision of relatively specific rules).

In the research reported by Finn et al. (2016) the relatively large differences between the *Color-Color* trial-type and the other three trial-types were attenuated considerably when the specific rules, which focused on specific trial-types, were replaced with more general rules. The rationale for employing more general rules was to avoid emphasizing specific trial-types before participants started responding on each block of trials. A reasonable question at this point therefore was how to attenuate the impact of experience on differential trial-type effects on the IRAP? Given that it was not possible to “undo” the effects of experience simply by providing general rules (because we suspected that experienced participants were self-generating *specific* rules), it seemed important to create a context for performance on the IRAP which served to undermine the dominance of a single trial-type over the other three. One way that this might be achieved would be to require participants to engage with each trial-type in a broadly similar manner during exposure to the task. We were unaware at the time of any obvious method that would work in this regard, but one method did suggest itself. Specifically, requiring participants to verbally report exactly what appears on screen on each IRAP trial, and the response they make, may reduce any tendency to treat one trial-type differently from the others. We referred to this method as the read-aloud procedure. On the grounds of intellectual honesty, we must acknowledge that adopting the read-aloud procedure was to some extent an intuitive leap, rather than a systematic solution to the problem we faced. We shall, however, return to this issue in greater detail in the General Discussion.

## **Experiment 2**

The design of Experiment 2 mirrored that of Experiment 1, except for the addition of a read-aloud procedure. Experimental experience was operationalized in the same way as it was in Experiment 1. The IRAP in the second experiment was identical to the IRAP used previously.

### **Method**

#### **Participants.**

Fifty-five individuals with an age range of 18-36 years ( $M = 21.7$ ) participated in the experiment in return for payment of €5. Informed consent was provided by all participants. The experiment was approved by the Ethics committee of the Faculty of Psychology and Educational Sciences at Ghent University.

#### **Materials**

The materials used in Experiment 2 were the same as those used in Experiment 1.

#### **Procedure**

The procedures of Experiment 2 differed from Experiment 1 only with regard to the verbal feedback procedure for the IRAP. The instructions given to participants before they engaged with the IRAP were similar to Experiment 1 up until the point at which participants began interacting directly with the program. At this point, the verbal feedback procedure introduced the read-aloud requirement (in Experiment 1 no reference was made to the read-aloud procedure because it was not employed in the previous experiment). The extra instructions given to participants were: first, to read out loud the word they saw at the top of the screen; second, to read out loud the word they saw in the middle of the screen; and third, to say out loud what response they were selecting as they pressed one of the response keys. So, for example, if participants selected the response “Ja” they were required to say the word “Ja” (i.e., the Dutch word for “Yes”) as they pressed the key related to this response.

If participants at any point failed to engage in the read-aloud procedure they were prompted by the experimenter to do so (e.g., “Please remember to say out loud what appears on screen and which response option you are choosing”). After participants completed the IRAP they were administered the questionnaire. Participants were then compensated, thanked for their time, and dismissed. For a random sample of participants a digital recording of them engaging in the read-aloud procedure was taken. Three independent raters were then asked to listen to a sample of the recordings and judge whether or not participants did in fact engage in the procedure as specified. All three raters agreed independently that participants, on the basis of the recordings, were fully engaging in the read-aloud procedure.

### **Results and Discussion**

The IRAP data were prepared for analysis in the same manner as was described for Experiment 1. The IRAP performance criteria were maintained during the test blocks by 24 of the 28 participants in the LE group and 25 of the 28 participants in the HE group. Only data from the remaining 49 participants are presented here. The mean number of experiments that participants reported completing before the current study, calculated across the remaining participants, was 36.7 for the HE group and 1.5 for the LE group.

The effect for each trial-type divided according to experimental group are presented in Figure 3 (no main effect for block-sequence was identified). All eight *D*-IRAP effects were in a direction consistent with natural verbal relations. Unlike Experiment 1, the effects for the two conditions did not appear to differ dramatically from each other. A 2x4 ANOVA revealed no main or interaction effects for experimental experience ( $ps > .5$ ). There was an effect for trial-type  $F(3,48) = 3.35, p < .01, \eta_p^2 = 0.1$ . Fisher’s PLSD tests conducted on the data, collapsed across the two conditions, indicated that the *D*-IRAP scores for the *Color-Color* trial-type were significantly different from the remaining three trial types ( $ps < .01$ ), with no differences among the remaining three ( $ps > .4$ ). Eight one-sample *t*-tests indicated that each

of the *D*-IRAP effects differed significantly from zero ( $ps < .01$ ). The inferential statistics therefore confirmed the descriptive analyses presented in Figure 3, in that experimental experience did not impact significantly on the IRAP performances when a read-aloud procedure was implemented.

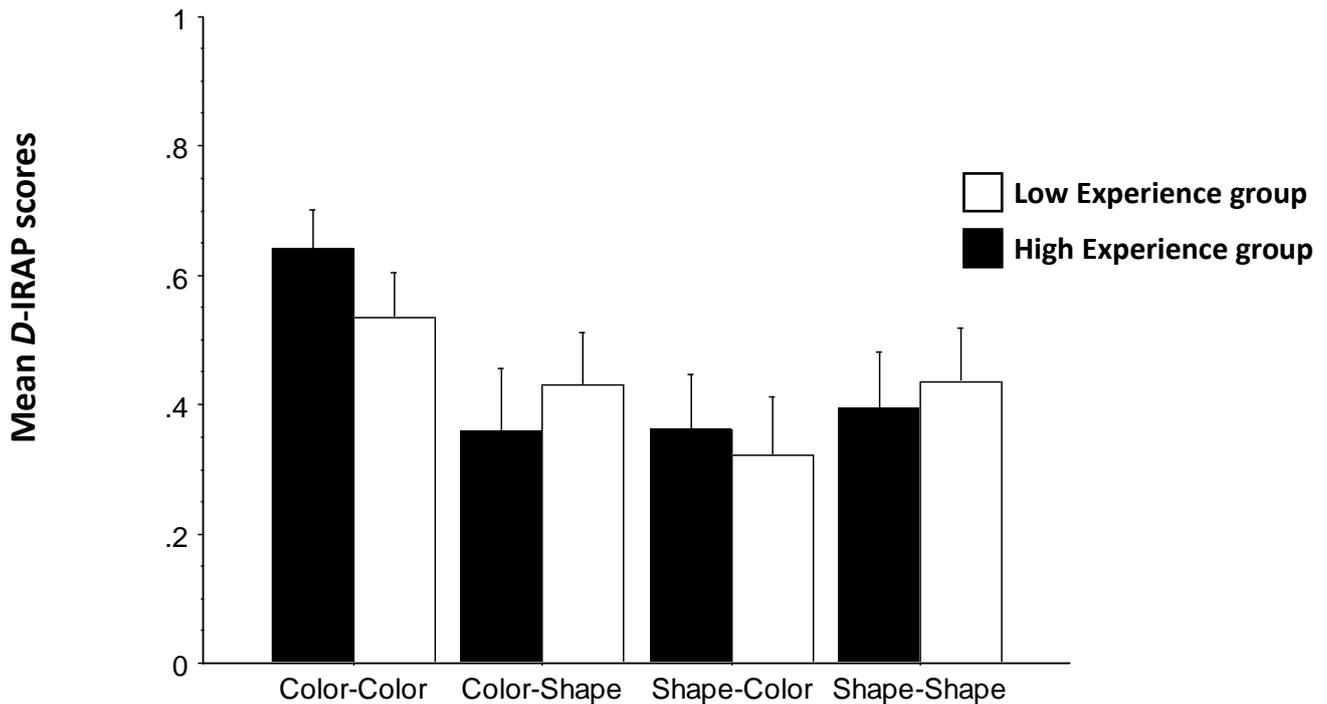


Figure 3. *D*-IRAP scores and standard errors for each trial-type for both IRAP block order in both experience groups.

### Experiment 3

At this point we had produced the single-trial-type-dominance-effect for the *Color-Color* trial-type across two experiments. Critically, although the effect appeared to be moderated by experimental experience and the requirement to engage in a concurrent read-aloud procedure, the dominance of the *Color-Color* trial-type remained statistically significant (across the two experiments). The question we continued to struggle with was why the *Color-Color* trial-type dominated over the *Shape-Shape* trial-type? No instructions were

provided that emphasized the former over the latter trial-type and both trial-types required the same response option (e.g., “Yes” rather than “No” during history-consistent trials). One possibility that we considered at this point was the potential role that verbal history with regard to colors versus shapes might have. Specifically, the SUBTELX-NL Lexical Database, containing 14,089 Dutch words, indicates that the color words used in Experiments 1 and 2 of the current study occur with relatively high frequencies in comparison with the shape words (Keuleers, Diependaele, & Brysbaert, 2010). That is, the raw frequency of each word in the database is as follows: Red = 2125; Blue = 1439; Green 1227; Yellow = 418; Pink = 835; Orange = N/A; Circle = 544; Square = 115; Triangle = 105; Oval = 14; Rectangle = 13; Octagon = N/A. Perhaps this differential in word frequencies may help to explain why the *Color-Color* trial-type effect consistently dominated over the *Shape-Shape* trial-type. As will become clear, this intuitive leap changed the way we approached the analysis of responding on the IRAP and it seems important to elaborate a little on this below.

We now assumed that the color words evoked relatively strong orienting responses relative to the shape words because the former occur more frequently in natural language.<sup>2</sup> Or more informally, participants may experience a type of orienting response to the color stimuli that is stronger than for the shape stimuli. Critically, a functionally similar orienting response may be likely for the “Yes” relative to the “No” response option (because “Yes” frequently functions as a confirmatory response in natural language). Coherence thus emerges on the *Color-Color* trial-type among the orienting functions of the label and target stimuli, and the “Yes” response option.<sup>3</sup> During consistent blocks this coherence extends to the relational

---

<sup>2</sup> As noted in the Introduction, the term “orienting” is used to denote a type of Cfunc property. We did consider using alternative terms, such as salience, but we felt that “orienting” evokes the involvement of a response function for a stimulus. In contrast, salience seems to imply that a stimulus may “stand out” independently of a behavioral history that is attached to it.

<sup>3</sup> Note that coherence refers here to the functional overlap of the Cfunc properties (in this case the orienting functions for the label, target and “Yes” response option) that have been established by the participant’s pre-experimental history of relational responding with regards to those specific stimuli.

response that is required between the label and target stimuli (e.g., “Color-Red-Yes”). In this sense, during consistent blocks this trial-type could be defined as involving a maximum level of coherence because all of the responses to the stimuli, both orienting and relational, are stronger relative to other trial-types. During inconsistent blocks, however, participants are required to choose the “No” response option, which does not cohere with any of the other orienting or relational responses on that trial-type, and this difference in coherence across blocks of trials yields relatively large *D*-scores. We shall return to this complex issue in the context of the General Discussion. At this point, however, we decided that it was important to determine if a single-trial-type-dominance-effect would be observed if the two categories of stimuli were roughly equal in terms of what we will define here as their orienting functions.

To identify two such categories we “brain-stormed” with colleagues to agree on two extremely bland categories of everyday objects, which are used with relatively equal frequency in the natural environment, and settled on “Forks” and “Spoons”. To avoid presenting participants with an IRAP that involved simply matching identical labels and targets (e.g., the word “Fork” to the word “Fork”) we employed three pictures of each of the two categories as label stimuli, while the words “Fork” and “Spoon” appeared as target stimuli. We recognize that employing pictures as stimuli in Experiment 3, rather than just words (as in Experiments 1 and 2), constitutes the manipulation of two variables (orienting functions and stimulus-type) across the experiments. However, our primary focus at this stage was on simply determining if the single-trial-type-dominance-effect would be observed when we presented two categories of stimuli in the IRAP that occurred with relatively equal frequency in the natural environment. We also sought to determine if we could manipulate the effect itself. Consequently, Experiment 3 involved three conditions. In one condition, we presented no instruction, in a second, we instructed participants to focus on forks, and in a third, we presented a spoon-focused instruction. Because both high- and low-experienced

participants in the previous experiments produced a single-trial-type-dominance-effect, albeit with a significant interaction in Experiment 1, the current experiment did not include experience as a variable. Given the on-going inductive and exploratory nature of the research we refrained from making formal predictions.

## Method

### Participants

The sample was composed of 44 participants with an age range of 18 – 33 ( $M = 22.8$ ). Participants were divided randomly into three separate groups, with 15 participants in two of the groups and 14 in the third. All participants gave their informed consent and were paid €5 for their participation. The experiment was approved by the ethics committee of the Faculty of Psychology and Educational Sciences at Ghent University.

### Materials

**Strategy sheets.** Participants in two of the groups were provided with different strategies for completing the IRAP, which were presented on A4-size sheets of paper entitled “Task Instructions”. One sheet presented spoon-focused instructions and the other sheet presented fork-focused instructions. Each sheet contained instructions on both sides: one side contained the history-consistent instructions (i.e., for responding in accordance with natural verbal relations); and the other side contained instructions for responding in a manner that was inconsistent with such relations. As an example, both sides of the spoon-focused instruction sheet contained three screenshots of the *Spoon-Spoon* trial-type, with a set of instructions appearing directly beneath. In each case, the instructions first outlined the generic nature of the IRAP, including the locations of the response keys, and stated that the goal of the task was to respond both quickly and accurately. The history-consistent instructions were as follows:

If you see a screen like the examples above where there are two items related to spoon select the key indicating “Yes”. If you see a screen where one item is related to spoon and one item related to fork select the key indicating “No”. Remember the best strategy is to search for screens that contain spoons.

The history-inconsistent instructions (which appeared on the opposite side of the sheet) were similar, except they told participants to respond in a history-inconsistent manner:

If you see a screen like the examples above where there are two items related to spoon select the key indicating “No”. If you see a screen where one item is related to spoon and one item related to fork select the key indicating “Yes”. Remember the best strategy is to search for screens that contain spoons.

The other sheet was similar to the first, except the three screen shots were of the *Fork-Fork* trial-type and the written instructions focused on forks rather than spoons.

**IRAP.** The labels presented on the IRAP in Experiment 2 were one of six pictures, three of forks and three of spoons, and the targets were the words “Fork” and “Spoon”. The four trial-types (*Fork-Fork*, *Fork-Spoon*, *Spoon-Fork*, and *Spoon-Spoon*) were presented in blocks of 24 trials. The response options displayed to participants were “Yes” and “No”. A history-consistent block of trials involved participants selecting “Yes” on *Fork-Fork*, and *Spoon-Spoon* trial-types and selecting “No” on *Fork-Spoon*, and *Spoon-Fork* trial types. History-inconsistent blocks of trials involved selecting the opposite response on each of the four trial-types. All other features of the task were the same as in the previous two experiments.

## **Procedure**

The experiment took place on an individual basis in sound proof cubicles at the Department of Experimental, Clinical, and Health Psychology at Ghent University. The experimental procedure did not begin until participants had read and signed the informed consent form.

Participants were seated before a laptop, which presented the IRAP program. All participants commenced the IRAP with a history-consistent block of trials. Participants in the

two strategy groups were provided with one of two strategy sheets by the experimenter and asked to read the instructions before interacting with the program. Participants in the third condition were exposed to a verbal feedback procedure similar to that employed in Experiment 1, except the verbal feedback referred to spoons and forks rather than shapes and colors. The IRAP program was set up such that participants who received the spoon-focused instructions received a *Spoon-Spoon* trial-type on the first trial, whereas participants who received fork-focused instructions were presented with a *Fork-Fork* trial-type on the first trial. In the verbal feedback condition, the initial trial on the IRAP was counterbalanced between *Fork-Fork* and *Spoon-Spoon* trial-types. Following the IRAP, participants were paid, thanked for their time and debriefed.

### Results and Discussion

The IRAP data were processed and converted to *D*-Scores in the same manner as the data from Experiments 1 and 2. All 44 participants maintained the performance criteria during the test blocks that were applied in the previous two experiments. The four *D*-IRAP scores for each participant were entered into a 4x3 mixed repeated measures ANOVA with instruction-type as the independent variable and IRAP trial-type as the within-participant variable. The ANOVA failed to yield a main or interaction effect for instruction-type ( $p > .37$ ), but the main effect for trial-type was significant,  $F(3,123) = 7.3, p < .001, \eta_p^2 = .15$ . Post-hoc tests (Fisher's PLSD), applied to the data collapsed across the three instruction-types (see Figure 4), indicated that the *Fork-Fork* and *Spoon-Spoon* trial-types both differed significantly from the remaining two trial-types ( $p < .03$ ) but not from each other ( $p > .71$ ). Finally, the comparison between the *Fork-Spoon* and *Spoon-Fork* trial-types was also non-significant ( $p > .17$ ). Specifically, the effects for the *Fork-Fork* and *Spoon-Spoon* trial-types were relatively even but significantly larger than the two remaining trial-types. As such, the properties of the two response options, rather than the orienting functions of the stimulus categories, appeared

to be the driving factor in producing the pattern of differential arbitrarily applicable relational responding effects. The three different types of instructions that were provided appeared to have little impact on the IRAP performances<sup>4</sup>.

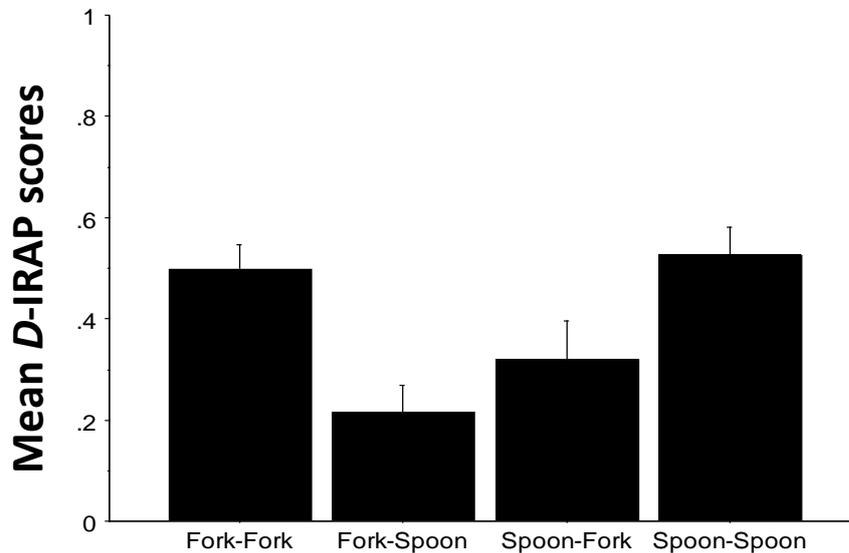


Figure 4. Means and standard errors for each trial-type in Experiment 3.

### General Discussion

Recent research has shown that the IRAP is sensitive to variables that extend beyond those specified by the REC model (e.g., Finn et al., 2016; Maloney & Barnes-Holmes, 2016). Specifically, the type of introductory rules or instructions used with the IRAP appear to impact upon the relative sizes of the individual trial-type effects; in addition, the types of response options employed (i.e., True/False versus Similar/Different) may also impact upon IRAP effects. The three experiments reported here arose from an attempt to explore the impact of prior experimental experience on the single-trial-type-dominance-effect. The results of Experiment 1 indicated that this effect was particularly pronounced for the highly experienced group. The effect was replicated in Experiment 2, although the requirement to

<sup>4</sup> The mean effect size and standard error for each trial-type for each group is available upon request.

engage in a read-aloud procedure appeared to reduce the differential impact of experimental experience. In the third and final experiment we employed a stimulus set that differed from the shapes and colors stimuli by using two categories that we suspected would possess roughly equal orienting functions. In this final experiment, a significant single-trial-type-dominance-effect failed to emerge in the data. In addition, our efforts to instruct specific single-trial-type-dominance-effects did not succeed.

Why did the instructions used in Experiment 3 appear to have such a limited impact on the IRAP performances? Although further research will be needed to address this question, post-hoc verbal reports suggest that many of the participants simply did not follow the instructions.<sup>5</sup> The fact that a single-trial-type-dominance-effect failed to emerge when it was explicitly instructed with the Forks-and-Spoons stimuli, but appeared to be so robust with stimuli that likely differed in terms of orienting functions (based on prior frequency of use), indicates that this historical variable (i.e., frequency) was relatively powerful in producing the effect. Notwithstanding that the effect appeared to be moderated by experimental experience and the read-aloud procedure, it seems important to attempt to explain how orienting functions impact upon IRAP performances. To this end, we will propose the beginnings of an explanation here, which we call the differential arbitrarily applicable relational responding effects (DAARRE) model.

In the introduction, an outline of the REC model was provided, and as noted the model is very much focused on the relations between the label and target stimuli. The basic prediction is that, all things being equal, participants will respond more quickly during trials that require relational responses that cohere with those that are consistent with their pre-

---

<sup>5</sup> Although perhaps a minor procedural issue, it may be worth noting that the instructions to focus on forks or spoons in Experiment 3 were presented on sheets of paper that participants were asked to physically flip between blocks of trials. In the study reported by Finn et al. (2016), which did show an effect for different types of instructions on IRAP performances, these were presented on the computer screen between blocks, which may have enhanced their impact.

experimental histories. The REC model always allowed for the potential impact of the functions of the response options on IRAP performances. For example, Barnes-Holmes, Murphy, Barnes-Holmes, and Stewart (2010) pointed out that, “It is possible. . . that a bias toward responding “True” over “False,” per se, interacted with the. . . stimulus relations presented in the IRAP” (p. 62). However, the REC model did not allow for the impact of the orienting functions of the stimuli, and how they may interact with the functions of the response options, in determining a pattern of differential trial-type effects observed on an IRAP. The DAARRE model proposed here constitutes a fresh attempt to identify the key variables involved in producing IRAP performances.

Before continuing, the reader should note that response options, such as “True” and “False”, are referred to as relational coherence indicators (RCIs) because they are often used to indicate the coherence or incoherence between the label and target stimuli that are presented within an IRAP (see Maloney & Barnes-Holmes, 2016). A core assumption of the DAARRE model is that differential trial-type effects may be explained by the extent to which the Cfunc and Crel properties cohere with the RCI properties of the response options across blocks of trials. The basic model as it applies to the stimuli employed in Experiments 1 and 2 of the current study is presented in Figure 5.

The model identifies three key sources of behavioral influence: (1) the relationship between the label and target stimuli (labeled as Crels); (2) the orienting functions of the label and target stimuli (labeled as Cfuncs); and (3) the coherence functions of the two response options (e.g., “Yes” and “No”). Consistent with the earlier suggestion that color-related stimuli likely possess stronger orienting functions relative to shape-related stimuli (based on differential frequencies in natural language), the Cfunc property for colors is labelled as positive and the Cfunc property for shapes is labelled as negative. The negative labelling for shapes should not to be taken to indicate a negative orienting function but simply an orienting

function that is weaker than that of colors. The labelling of the relations between the label and target stimuli indicates the extent to which they cohere or do not cohere based on the participant's relevant history. Thus, a *Color-Color* relation is labelled with a plus sign (i.e., coherence) whereas a color-shape relation is labelled with a minus sign (i.e., incoherence). Finally, the two response options are each labelled with a plus or minus sign to indicate their functions as either coherence or incoherence indicators (see Maloney & Barnes-Holmes, 2016). In the current example, "Yes" (+) would typically be used in natural language to indicate coherence and "No" (-) to indicate incoherence. These positive and negative labels merely indicate the *relative* dominance of the Crel, Cfunc, and RCI properties for each stimulus, or relationship between stimuli, in the context of the other stimuli or relationships presented within the IRAP. Note, also, that these and all of the other functions labelled in Figure 5 are behaviorally determined, by the past and current contextual history of the participant, and should not be seen as absolute or inherent in the stimuli themselves. The precise pattern of IRAP effects observed for a participant completing an IRAP emerges from the dynamic interplay of these functions in the ongoing act-in-context that is an IRAP performance.

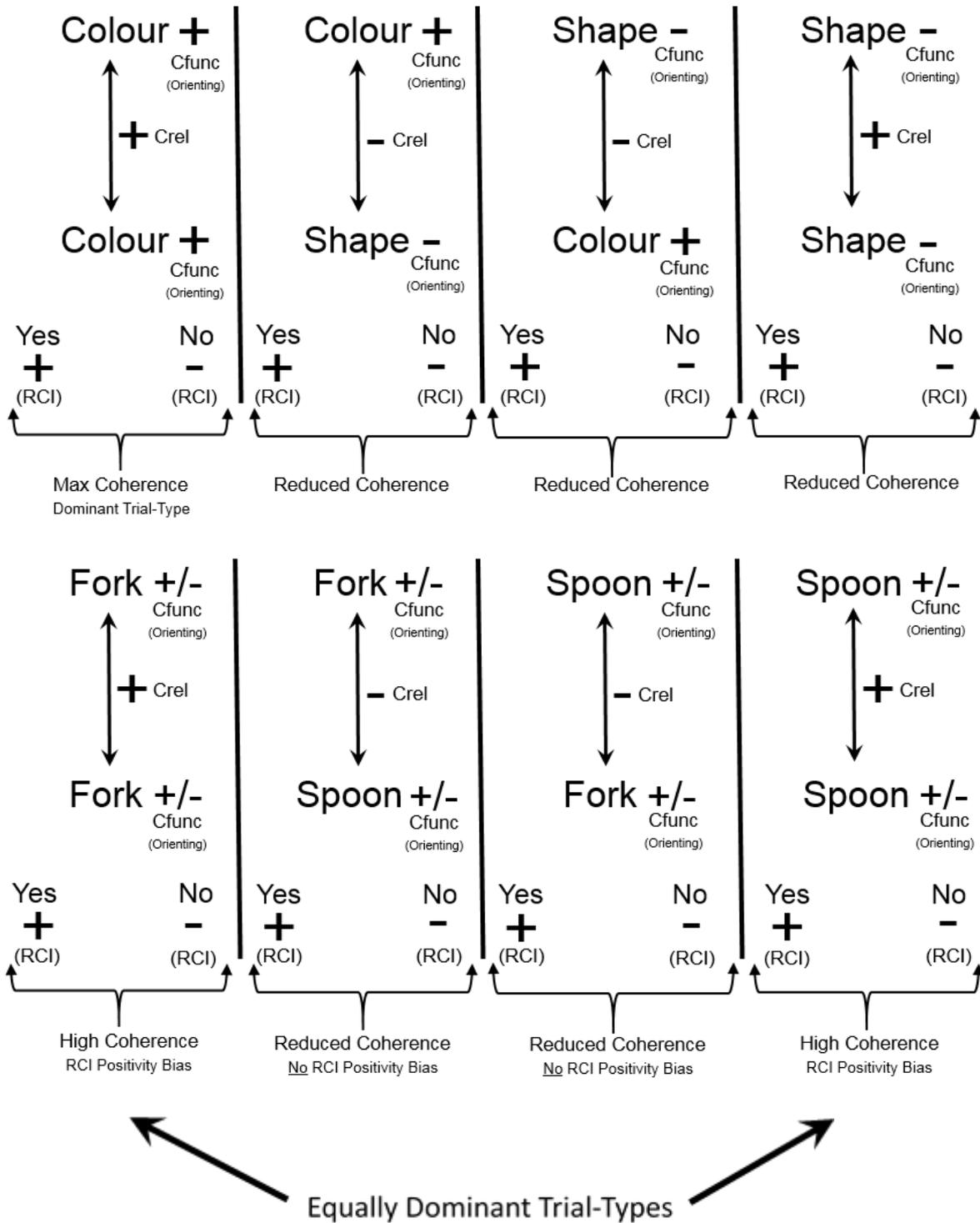


Figure 5. The DAARRE model as it applies to the Shapes-and-Colors stimulus set (upper panel) and Forks-and-Spoons stimulus set (lower panel). The positive and negative labels refer to the relative positivity of the Cfuncs, for each label and target, the relative positivity of the Crels and the relative positivity of the RCIs in the context of the other Cfuncs, Crels and RCIs in that stimulus set.

As can be seen from Figure 5 (upper panel), each trial-type differs in its pattern of Cfuncs and Crels, in terms of plus and minus properties, that define the trial-type for the Shapes-and-Colors IRAP. The single-trial-type-dominance-effect for the *Color-Color* trial-type may be explained, as noted above, by the DAARRE model based on the extent to which the Cfunc and Crel properties cohere with the RCI properties of the response options across blocks of trials. To appreciate this explanation, note that the Cfunc and Crel properties for the *Color-Color* trial-type are all labelled with plus signs; in addition, the RCI that is deemed correct for history-consistent trials is also labelled with a plus sign (the only instance of four plus signs in the diagram). In this case, therefore, according to the model this trial-type may be considered as maximally coherent during history-consistent trials. In contrast, during history-inconsistent trials there is no coherence between the required RCI (minus sign) and the properties of the Cfuncs and Crel (all plus signs). According to the DAARRE model, this stark contrast in levels of coherence across blocks of trials serves to produce a relatively large IRAP effect. Now consider the *Shape-Shape* trial-type, which requires that participants choose the same RCI as the *Color-Color* trial-type during history-consistent trials, but here the property of the RCI (plus sign) does *not* cohere with the Cfunc properties of the label and target stimuli (both minus signs). During history-inconsistent trials the RCI *does* cohere with the Cfunc properties but not with the Crel property (plus sign). Thus, the differences in coherence between history-consistent and history-inconsistent trials across these two trial-types is not equal (i.e., the difference is greater for the *Color-Color* trial-type) and thus favors the single-trial-type-dominance-effect (for *Color-Color*). Finally, as becomes apparent from inspecting the Figure for the remaining two trial-types (*Color-Shape* and *Shape-Color*) the differences in coherence across history-consistent and history-inconsistent blocks is reduced

relative to the *Color-Color* trial-type (two plus signs relative to four), thus again supporting the single-trial-type-dominance-effect.<sup>6</sup>

Turning now to the lower panel of Figure 5, we can see the way in which the DAARRE model explains what we will call here the dual-trial-type-dominance-effect (i.e., large and relatively even effects for the *Fork-Fork* and *Spoon-Spoon* trial-types). Although the Crel and RCI properties remain unchanged, relative to the Shapes-and-Colors DAARRE model, the Cfunc properties for the label and target stimuli are all now labelled with a plus and minus sign (i.e., +/-). The dual signing is used to denote that the orienting functions for those stimuli were not differentiated by the researchers. That is, spoons and forks were deliberately selected because they were deemed to be roughly equally bland, and occurring with approximately equal frequency in the natural environment. Indeed, the raw frequency for these words in the SUBTELX-NL Lexical Database was 227 for “Fork”, and 219 for “Spoon” (Keuleers, et al., 2010). On this basis the DAARRE model predicts that the Cfunc properties of the label and target stimuli will not be a determining factor in producing a difference in the size of the trial-type effects for the trial-types that share an RCI within blocks of trials. However, differences may still be observed between trial-types that do not share RCIs. In the current example (i.e., spoons and forks), the DAARRE model indicates that there is increased coherence for the *Fork-Fork* and *Spoon-Spoon* trial-types relative to the other two. Specifically, the coherence between a coherent Crel (+) and a coherent RCI (“Yes”) is more coherent than the coherence between an incoherent Crel (-) and an incoherent RCI (“No”).

---

<sup>6</sup> Although the DAARRE model highlights three variables that may interact to increase or decrease levels of coherence (i.e., the Crel, Cfunc, and RCI properties of the stimuli), we are not proposing three functionally distinct types or classes of coherence. Coherence, as a concept in RFT, may be interpreted as the extent to which a current pattern of AARRing is consistent (i.e. coherent) with the behavioral history that gave rise to that AARRing. Critically, the level of coherence involved in a particular pattern of AARRing may be attributable to multiple interactive variables, but this does not imply a different type of coherence for every interactive pattern that may be identified. Coherence thus remains a unitary concept within the DAARRE model as currently expressed.

More informally, this may be referred to as a positivity bias (i.e., the bias found in natural languages in which confirming tends to dominate over negating).

At this point, the DAARRE model appears to explain the single-trial-type-dominance-effect for the Shapes-and-Colors IRAP and the dual-trial-type-dominance-effect for the Forks-and-Spoons IRAP. But how might it explain the moderating influence of prior experience, and the read-aloud procedure reported here, and the instruction effects reported by Finn et al. (2016)? First, consider the impact of prior experience. Perhaps, the highly experienced participants were more influenced by the Cfunc properties of the label and target stimuli because many reaction-time tasks require participants to orient to particular features of the stimuli presented within the task (indeed, for some popular tasks, such as the Stroop, color itself is the relevant dimension) and thus they tended to show the strongest single-trial-type-dominance-effect.

Second, consider the moderating influence of the read-aloud procedure. Perhaps this “intervention” reduced the impact of experimental experience on the single-trial-type-dominance-effect because all participants were required to respond in a similar manner to the individual stimuli presented on each trial-type of the IRAP. That is, all participants were forced to read each label and target separately, and the response option they chose, on each trial. Perhaps, therefore, any differences between the high and low experienced participants in their relative sensitivities to coherence, between the response options and the label and target stimuli, may have been undermined. At the present time, exactly how the relative sensitivities were changed remains to be specified.

Finally, consider the instructional effects as reported by Finn et al. (2016). Perhaps, the different types of instructions provided at the beginning of each block of trials may have influenced the relative impact of the Cfunc versus Crel control within the IRAP. For example, when particular stimuli were specified in the instructions (e.g., “respond as if colors are colors

and shapes are shapes”), this may have increased the extent to which the orienting functions of the stimuli impacted upon the participants’ IRAP performances, thus generating stronger differential trial-type effects. Note also that in Experiment 2 of Finn et al., the differential trial-type effect was almost completely absent when specific stimuli were not instructed and the IRAP commenced with a history-inconsistent block. Perhaps commencing with such a block immediately “punished” responding based on the Cfunc (orienting) functions of the stimuli, and thus participants were more inclined to engage in Crel-based responding across all remaining blocks. Of course, all of the foregoing is entirely post-hoc, as it would have to be given the inductive nature of the research program, but it seems useful to offer potential explanations here because they may be highly instructive in terms of conducting future research.

We should emphasize once more that the current series of experiments constitute an example of a highly inductive approach to developing an increasingly sophisticated understanding of the behavioral processes involved in IRAP performances. The research started with the finding that different types of instructions appear to impact upon IRAP trial-type effects, as reported by Finn et al. (2016). In attempting to replicate this earlier work a specific trial-type effect again emerged that was difficult to explain in terms of the REC model (i.e., the single-trial-type-dominance-effect). In exploring this effect, and trying to predict and influence it experimentally, a number of variables were identified that allowed us to propose the DAARRE model. We have deliberately presented the model in a post-hoc fashion because that is exactly how it emerged. Of course, it may be used to guide future research but we anticipate that the model itself will need to be further developed and refined to accommodate additional complexities in IRAP performances. For example, we have not considered situations in which more than one Cfunc dimension is involved (e.g., separate orienting and evaluative dimensions). However, we believe that we are now at a point in our

research program where it would be useful to share our current findings and the conceptual developments that are now emerging from our empirical research.

In closing, it seems important to reflect, if only briefly, on the strategy of proposing the DAARRE model itself. The point at which a new model is proposed will always remain a “judgement call” on behalf of the researchers involved in proposing that model. In this instance, the basis for introducing a new model was driven, in part, by the fact that the REC model does not readily accommodate the single-trial-type-dominance effect. More importantly, however, in tackling this deficit in the REC model it became apparent to us that the model also failed to draw on a critically important distinction in RFT itself – the distinction between the Crel and Cfunc properties participating in a given relational network. Proposing a model that fully recognizes this distinction does not merely render the model more RFT-consistent, it should in principle serve to increase precision and scope (i.e., help to explain more with less). Thus, for example, as an increasing range of other variables, such as instructions, various types of pre-experimental experience, the potential impact of self-generated rules during exposure to the IRAP, and procedural variables in the IRAP itself, are shown to impact upon IRAP performances, new models should not be needed at each point. Instead, the impact of an increasing array of potentially important “independent variables” may be interpreted or explained in terms of the dynamic interactions between the Crel, Cfunc, and RCI properties of the contextual variables involved. Of course, in time the DAARRE model itself may fail to account for some important findings, but when and if this occurs it will highlight the need to adjust or modify RFT itself (i.e., because if additional contextual properties above and beyond Cfunc, Crel, and RCI are needed, they are currently not specified within RFT). Nevertheless, it is our hope that the DAARRE model will provide us with sufficient precision and scope as we continue to grapple with the dynamics of arbitrarily applicable relational responding in both basic and applied research settings.

### **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. Ciara McEnteggart declares that she 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., Hussey, I., & Luciano, C. (2016). Relational frame theory: Finding its historical roots and future development. In R. Zettle, S. C. Hayes, D. Barnes-Holmes, & T. Biglan (Eds), *Wiley Handbook of contextual behavioral science* (pp. 161-175). Cambridge: Wiley-Blackwell. DOI: 10.1002/9781118489857.ch8
- Barnes-Holmes, D., Barnes-Holmes, Y., Stewart, I., & Boles, S. (2010). A sketch of the Implicit Relational Assessment Procedure (IRAP) and the Relational Elaboration and Coherence (REC) model. *The Psychological Record, 60*, 527-542.
- Barnes-Holmes, D., Murphy, A., Barnes-Holmes, Y., & Stewart, I. (2010). The Implicit Relational Assessment Procedure: 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.
- Carpenter, K. M., Martinez, D., Vadhan, N. P., Barnes-Holmes, D., & Nunes, E. V. (2012). Measures of attentional bias and relational responding are associated with behavioral treatment outcome for cocaine dependence. *The American journal of drug and alcohol abuse, 38*, 146-154. DOI: 10.3109/00952990.2011.643986
- Dodds, P. S., Clark, E. M., Desu, S., Frank, M. R., Reagan, A. J., Williams, J. R., ... & Megerdooian, K. (2015). Human language reveals a universal positivity bias. *Proceedings of the National Academy of Sciences, 112*(8), 2389-2394. DOI: 10.1073/pnas.1411678112
- Dymond, S., & Roche, B. (2013). *Advances in relational frame theory: Research and application*. Oakland, CA: New Harbinger.
- Finn, M., Barnes-Holmes, D., Hussey, I., & Graddy, J. (2016). Exploring the behavioral dynamics of the implicit relational assessment procedure: The impact of three types of

- introductory rules. *The Psychological Record*, 66, 309-321. DOI:10.1007/s40732-016-0173-4
- 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. (2016a). Relational frame theory: The basic account. In R. Zettle, S. C. Hayes, D. Barnes-Holmes, & T. Biglan (Eds), *Wiley Handbook of contextual behavioral science* (pp. 176-232). Cambridge: Wiley-Blackwell. DOI 10.1002/9781118489857.ch9
- Hughes, S., & Barnes-Holmes, D. (2016b). Relational frame theory: Implications for the study of human language and cognition. In R. Zettle, S. C. Hayes, D. Barnes-Holmes, & T. Biglan (Eds), *Wiley Handbook of contextual behavioral science* (pp. 233-291). Cambridge: Wiley-Blackwell. DOI 10.1002/9781118489857.ch10
- Keuleers, E., Diependaele, K., & Brysbaert, M. (2010). Practice effects in large-scale visual word recognition studies: A lexical decision study on 14,000 Dutch mono- and disyllabic words and nonwords. *Frontiers in Psychology*, 1, 174. DOI 10.3389/fpsyg.2010.00174
- Leech, A., Barnes-Holmes, D., & Madden, L. (2016). The Implicit Relational Assessment Procedure (IRAP) as a measure of spider fear, avoidance, and approach. *The Psychological Record*, 66, 337-349. DOI 10.1007/s40732-016-0176-1
- Maloney, E., & Barnes-Holmes, D. (2016). Exploring the behavioral dynamics of the

Implicit Relational Assessment Procedure: The role of relational contextual cues versus relational coherence indicators as response options. *The Psychological Record*, 66, 1-9. DOI 10.1007/s40732-016-0180-5

Nicholson, E., & Barnes-Holmes, D. (2012). The Implicit Relational Assessment Procedure (IRAP) as a measure of spider fear. *The Psychological Record*, 62, 263-278.

Sidman, M. (1971). Reading and auditory-visual equivalences. *Journal of Speech, Language, and Hearing Research*, 14, 5-13.

Sidman, M. (1994). *Equivalence relations and behavior: A research story*. Boston, MA: Authors Cooperative.