Feel the Func: Interpreting IRAP Performances Based on Cfunc versus Crel Stimulus Properties

Feel the Func: Interpreting IRAP Performances Based on Cfunc versus Crel Stimulus Properties

Renato Bortoloti (1,2), Margarete Schmidt (1,2), Colin Harte (2,3,4), Dermot Barnes-Holmes (5)

(1) Universidade Federal de Minas Gerais, Brazil
(2) Instituto Nacional de Ciência e Tecnologia sobre Comportamento, Cognição e Ensino, Brazil
(3) Universidade Federal de São Carlos, Brazil
(4) Paradigma – Centro de Ciências e Tecnologia do Comportamento, Brazil
(5) Ulster University, Northern Ireland

Acknowledgements: This research was supported by the National Institute of Science and Technology on Behavior, Cognition, and Teaching (INCT-ECCE). The INCT-ECCE is financially supported by the Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP, Grant # 2014/50909-8), the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES, Grant #88887.136407/2017-00), and the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq, Grant # 465686/2014-1). Preparation of this manuscript was supported by a productivity research grant to the first author by the CNPq (Grant # 309041/2021-0) and by a fellowship awarded to the third author by FAPESP (Grant #2019/24210-0).

Correspondence concerning this article should be sent to renatobortoloti@gmail.com or colin.n.harte@gmail.com

Abstract

Two experiments tested a basic assumption of the differential arbitrarily applicable relational responding effects (DAARRE) model by examining the extent to which functional (Cfunc) and relational (Crel) properties of stimuli impact performances on the implicit relational assessment procedure (IRAP). Experiment 1 required participants to complete IRAPs in both their native and a foreign language. Twenty-one Brazilian participants (fluent in both Portuguese and English) completed two IRAPs, one in which Portuguese-language stimuli were predominantly used and a second in which English-language stimuli were predominantly used. The IRAP trial-type containing Portuguese-language (i) labels, (ii) positively valenced targets and (iii) response options produced the largest IRAP effect across all eight trial-types (four within each IRAP). Consistent with the DAARRE model, the Cfunc properties of the native language stimuli appeared to dominate over the Crel properties, relative to the foreign language stimuli. Experiment 2 sought to extend Experiment 1 using a known-groups design involving stimuli that were assumed to possess differential Cfunc properties across two groups of Brazilian soccer fans. Critically, responding to the IRAP trial-types did not require any explicit evaluative response in that the IRAP required participants to simply categorize team badges with team names. The largest IRAP effect observed for each group was for the trial-type that presented the name and badge for that group's supported team. Both experiments provide support for the DAARRE model analysis by indicating that the Cfunc properties of stimuli are critically important when interpreting IRAP performances. A number of caveats to this conclusion are discussed.

Keywords: IRAP, Cfunc, Crel, DAARRE model.

In the classic novel *The Magic Mountain* (Mann, 1927), the German Hans Carstop is enraptured in erotic temptation, lust and love by the francophone Clawdia Chauchat. After months of silent flirting, the very first time Carstop addresses bold words to Chauchat he prefers to do so in French instead of German, and the excerpt below explains why.

> *- Comment? C'était une phrase tout à fait indifferente, ce que j'ai dit lá. Moi, tu le remarques bien, je ne parle guère le français. Pourtant, avec toi je prefere cette langue à la mienne, car pour moi, parler français, c'est parler sans parler, em quelque manière: sans responsabilité, ou comme nous parlons en rêve. Tu comprends?*[1]

Carstop seems to claim that switching to French when he talks about sensitive personal issues helps to reduce the emotional impact that words in his native language would typically evoke. In effect, the excerpt suggests that such a strategy helps Carstop to feel less vulnerable and emotionally exposed in this interaction with Chauchat. This perspective is consistent with a large body of literature according to which a foreign language is perceived as less emotional than the native language (e.g., Caldwell-Harris, 2015; Circi et al., 2021; Pavlenko, 2017).

This widely recognized difference between the use of native and foreign languages may be relevant to recent research emerging from relational frame theory (RFT; Hayes et al., 2001), a well-established behavior-analytic account of human language and cognition. Specifically, this work has focused on an experimental tool known as the implicit relational assessment procedure (IRAP). The IRAP was originally designed to assess the strength or probability of natural verbal relations as defined within RFT (Barnes-Holmes et al., 2008). The IRAP requires participants to shift, in different blocks of trials, between responding in accordance with an established learning history and contrary to that history. For example, some blocks of trials might require responding to flowers as positive while others require responding to flowers as negative. Shorter latencies in blocks of trials that require responding in a manner consistent with their learning history, compared to latencies in history-inconsistent blocks, is defined as the IRAP effect. Typically, each IRAP trial involves the simultaneous presentation of a *label*, a *target*, and two response options, such as 'Similar' and 'Different' or 'True' and 'False', or 'Yes' and 'No'. In an IRAP focused on responses to flowers, for example, the label might consist of the name of a flower, the target a positively valenced word, and two response options, such as Rose-Love-True/False. In a history-consistent block responding "True" would be considered correct, whereas in a history-inconsistent block "False" would be correct.

[1] How? It was a completely indifferent sentence, what I said there. Me, you notice it well, I hardly speak French. However, with you I prefer this language to mine, because for me, to speak French is to speak without speaking, in some way: without responsibility, or as we speak in dreams. You understand?

An early model for interpreting IRAP effects, the relational elaboration and coherence (REC; Barnes-Holmes et al., 2010) model, focused on explaining response patterns produced on the measure by focusing primarily on the relation between the label and target stimuli; in RFT terms, the Crel property (e.g., coordination). However, this model could not readily account for differential trial-type effects that emerged between trials involving relatively innocuous stimuli that did not differ substantively in their emotional or valence properties (e.g., colors and shapes). For example, in a study by Finn et al. (2016), the IRAP effect was larger on trials that required confirming colors were colors (e.g., Color-Red-True) than on trials that required confirming shapes were shapes (e.g., Shape-Square-True). Such an effect, in which one trial-type differentially dominates over another, despite requiring the same response option within blocks of trials, was referred to as a single trial-type dominance effect (STTDE; Finn et al., 2018).

Based on a recent conceptual analysis of IRAP results, a new model has been proposed in an attempt to explain these differential trial-type effects. The model, referred to as the differential arbitrarily applicable relational responding effects (DAARRE; Finn et al., 2018) model, highlights the importance of not only the relational (Crel) properties of stimuli, but also the functional (Cfunc) properties of the stimuli and response options. In effect, the DAARRE model explicitly includes the general functional properties of the stimuli – including their orienting (attentional), evoking (emotional), and/or motivative properties. Specifically, the model predicts that individuals completing an IRAP are not just sensitive to the relations between label and target stimuli, but are also sensitive to the coherence (i.e., overlap) among the functional properties of the stimuli and response options. Indeed, recognizing the critical role played by the Cfunc and Crel properties of stimuli has helped to interpret differential patterns of IRAP effects across a range of recent studies (e.g., Bortoloti et al., 2019, 2020; Finn et al., 2019; Pidgeon et al., 2020; Pinto et al., 2020, Schmidt et al., 2021; see Barnes-Holmes & Harte, 2022a, 2022b).

One way in which to further test the DAARRE model may involve drawing on the important distinction between the functional properties of native and foreign languages discussed above in *The Magic Mountain* example. In terms of the DAARRE model, we might predict that the functional (Cfunc) properties of verbal stimuli in one's native language would play a larger role in producing differential trial-type effects on the IRAP than verbal stimuli from a foreign language. That is, words in our native language have a longer history of being related, in an RFT sense, to stimuli and events in the natural environment; in contrast, word-referent relations in a foreign language, even one in which the individual is relatively fluent, would typically have a relatively briefer history in this regard. Indeed, it may be the case that the symbolic function of a word in a foreign language may be derived in part from its relation to the semantically similar native word. In any case, the basic prediction is that the Cfunc

properties of words in a native language may dominate to a greater extent over their Crel properties relative to words in a foreign language.

To begin to test this basic idea, consider an experiment in which Brazilian participants (who speak both Portuguese and English) are exposed to two IRAPs, one in which Portuguese-language stimuli are predominantly used and a second in which English-language stimuli are predominantly used. Based on the foregoing conceptual analysis, we assumed that the IRAP trial-type containing Portuguese-language (i) labels, (ii) positively valenced targets and (iii) response options would produce the largest IRAP effect across all eight trial-types (four within each IRAP). According to the DAARRE model, this is the trial-type that would be considered maximally coherent because all of the Cfunc and Crel properties within that trial-type would overlap functionally for native Portuguese speakers. All remaining trial-types across the two IRAPs would involve some level of incoherence such as an English word or a negatively valenced target. We refrained from making further predictions concerning the remaining trial-types at this time because this was the first IRAP study that we know of to examine directly the impact of native versus foreign languages on IRAP performances.

A second experiment was also run which involved asking a broadly similar question but employing two groups that differed in the predicted Cfunc properties of the stimuli employed. For ease of communication, the details of this experiment will be provided after the presentation of Experiment 1.

Method

Participants

Twenty-five Brazilian airline workers (13 males) were invited to participate in the research (because they were likely to be fluent in both Portuguese and English). All of them were pilots or flight attendants of international routes. Their native language was Portuguese but they were also fluent English speakers (ICAO level 4 or higher). Written informed consent was obtained from all participants.

Setting, apparatus and procedure

Sessions were conducted individually in a quiet room of a Brazilian international airport. The IRAP task was presented on a laptop computer. Each participant completed two similar IRAPs, one presented in Portuguese (hereafter referred to as the Native-language IRAP) and the other in English (hereafter referred to as the Foreign-language IRAP). The IRAP trials were divided into two different types of alternating blocks, half of which required responding in a pro-Brazilian/anti-English manner (e.g., Brazilian name-Positive adjective-True; English name-Positive adjective-False) and half required responding in a pro-English manner (e.g., English name-Positive adjective-True; Brazilian name-Positive adjective-False). On each IRAP trial, one label stimulus appeared at the top of the screen, one target stimulus in the center, and

two response options at the bottom left- and right-hand side. Both the Native- and Foreign-language IRAPs comprised blocks of 24 trials – at least two practice blocks and six test blocks. The practice blocks were presented until participants achieved >/= 80% of correct choices within a median response time of 2 s in one pro-Brazilian and one pro-English block, consecutively. Rules for responding were displayed at the beginning of each block of trials (see below). For the six test blocks, participants were not required to maintain the accuracy or latency criteria to progress from one test block to the next. However, if a participant did not maintain the criteria (>/= 80% accuracy and < 2000 ms response latency across each block), their data were excluded from subsequent analyses.

Native-language IRAP

At the beginning of the pro-Brazilian blocks, the participants read: “Relacione os nomes brasileiros a adjetivos positivos e nomes em inglês a adjetivos negativos”; at the beginning of the pro-English blocks, the participants read: “Relacione os nomes brasileiros a adjetivos negativos e nomes em inglês a adjetivos positivos” (see below for English translations). Next, a label name – either a common Brazilian (João, Pedro, Mateus) or English (John, Michael, Smith) name – was presented at the top of the screen. A single word target – a positive (bonito, jovem, amável) or a negative (feio, velho, nervoso) adjective – was presented at the center. Finally, two response options – “sim” and “não” – were presented at the bottom corners of the screen (“sim” on the left and “não” on the right). These label-target combinations yielded four individual trial-types: Brazilian name-Positive; Brazilian name-Negative; English name-Positive; English name-Negative.

Foreign-language IRAP

At the beginning of the pro-Brazilian blocks, participants read: “Relate Brazilian nouns to positive adjectives and English nouns to negative adjectives”; at the beginning of the pro-English blocks, the participants read: “Relate Brazilian nouns to negative adjectives and English nouns to positive adjectives”. All remaining aspects were similar to the Native-language IRAP except that the target adjectives were presented in English (handsome, polite, nice, ugly, sad, naughty), as were the response options (“yes” on the left and “no” on the right).

In both IRAPs, the response options remained at the same position for the duration of the procedure. All stimuli remained visible until the participant pressed one of the response keys. The task consisted of choosing one of these options by pressing the “d” or the “k” key for “sim/yes” or “não/no” respectively. Choosing the response considered correct in a given block removed all the stimuli from the screen and, after 400 ms, the next trial was presented. The choice of the response considered incorrect produced a red X in the middle of the screen (immediately below the target adjective). The next trial was not presented until the participant pressed the correct key.

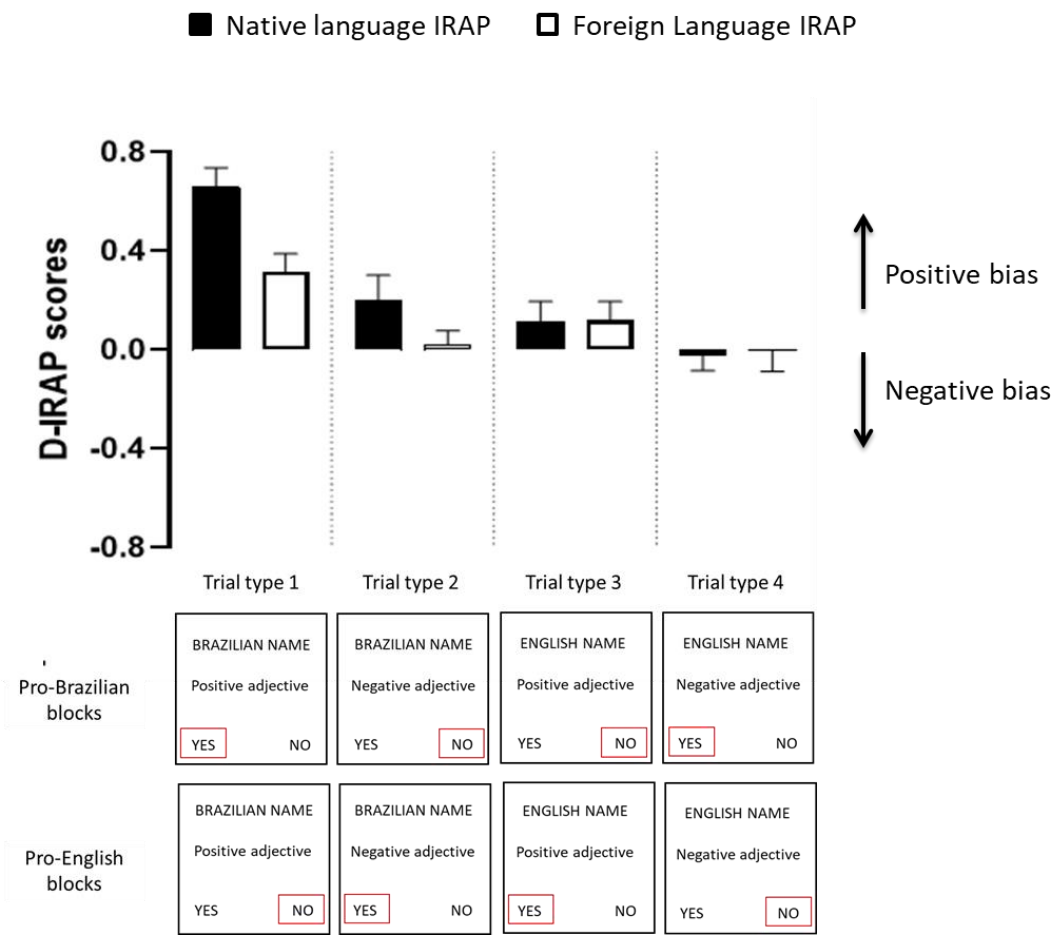
Response latency, defined as the time in milliseconds (ms) that elapsed between trial onset and the emission of a correct response, was the main datum recorded during the IRAP task. All latency data were processed by the D-IRAP algorithm, available in the IRAP software. In brief, the average latency scores for each of the trial-types within each test block pair were averaged and divided by the standard deviation calculated across that block pair (see Barnes-Holmes et al., 2010 for more detail). Following completion of all blocks, a written message indicated the end of the experiment. The participant was thanked, debriefed, and any questions they had about the experiment were answered.

Results and Discussion

Of the 25 participants who were exposed to the IRAPs, 21 successfully met the inclusion criteria in both the Portuguese and the English versions. The mean D-IRAP scores for these 21 participants for the four IRAP trial-types in each condition (i.e., Portuguese and English) are presented in Figure 1. Note that D-IRAP scores for trial-types 3 and 4 have been inverted (i.e., divided by -1) to provide a common axis along which to compare the 4 trial-types. Indeed, such inversions are generally necessary in order to make meaningful comparisons among all trial-types involving explicitly evaluative responses, such as those in the current experiment.

Fig 1

Mean D-IRAP scores extracted for the 21 participants who met the IRAP inclusion criteria.



Note. A higher D-IRAP score indicates a greater difference in response latencies between pro-Brazilian and pro-English blocks. Scores for trial-types 3 and 4 have been inverted and thus positive scores indicate a positive response bias (e.g., responding 'yes' more quickly than 'no' to an English name and positive adjective) and negative scores a negative response bias (e.g., responding 'yes' more quickly than 'no' to an English name and negative adjective).

Visual inspection of Figure 1 indicates that responding to Brazilian names and positive adjectives in the native language IRAP produced by far the largest D-IRAP effect. The effect for this trial-type on the foreign language IRAP was also quite large, but was approximately only half the size. The remaining trial-type effects were all relatively small, although the score for the Brazilian name-negative adjective trial-type, in the native language IRAP, appeared larger than its counterpart in the foreign language IRAP.

As mentioned in the Introduction, we assumed that the IRAP trial-type containing Portuguese-language labels and positively valenced targets (i.e., Portuguese-Language IRAP

trial-type 1) would produce the largest IRAP effect across all eight trial-types (four within each IRAP). In order to test this assumption, seven paired samples t-tests were conducted between the Portuguese-Language trial-type 1 and all other trial-types across both IRAPs. All proved to be statistically significant (all *p*'s < .001; see Table 1). We also conducted Bayesian paired samples t-tests, which suggested very strong evidence for the difference between the Portuguese-Language IRAP trial-types 1 and 2 (BF_{10} = 37.854), and extremely strong evidence for the difference between the Portuguese-Language trial-type 1 and all other trial-types (BF_{10} >/= 220.905).

Table 1

Paired samples t-tests between the Native Language IRAP trial-type 1 and all other trial types in Experiment 1

						95% CI for Cohen's d	
Measure 1	Measure 2	*t*	*df*	*p*	Cohen's *d*	Lower	Upper
Native Language TT1	Native Language TT2	3.872	20	< .001	.845	.337	1.338
Native Language TT1	Native Language TT3	5.126	20	< .001	1.118	.561	1.659
Native Language TT1	Native Language TT4	6.288	20	< .001	1.372	.762	1.964
Native Language TT1	Foreign Language TT1	4.729	20	< .001	1.032	.491	1.556
Native Language TT1	Foreign Language TT2	5.837	20	< .001	1.274	.685	1.845
Native Language TT1	Foreign Language TT3	4.644	20	< .001	1.013	.476	1.534
Native Language TT1	Foreign Language TT4	4.656	20	< .001	1.016	.478	1.537

Although we refrained from making further predictions concerning the remaining trial-types, we performed two post-hoc exploratory analyses to assess whether (i) the STTDE was also present in the foreign language IRAP alone and (ii) whether the visual difference between trial-type 2 across IRAPs was statistically significant. To first explore whether a STTDE was also present in the foreign language IRAP alone, three paired sample t-tests were conducted between Foreign-Language trial-type 1 and the remaining three trial-types in that IRAP. Results showed that trial-type 1 was significantly different from two of the other three (trial-type 2: *t*(20) = 2.69, *p* = .014, Cohen's *d* = .59 95% CI [.12, 1.05]; trial-type 4: t(20) = 2.35, p = .029, Cohen's *d* = .51, 95% CI [.05, .96]). The difference between trial-types 1 and 3 was not statistically significant (*p* > .08). Note, however, with Bonferroni corrections, only the difference between trial-types 1 and 2 was marginally significant (*p* = .042). Bayesian paired samples t-tests suggested moderate evidence for the difference between trial-type 1 and 2 (BF_{10} = 3.79), anecdotal evidence for the difference between trial-type 1 and 4 (BF_{10} = 2.078), and no

support for or against a difference between trial-types 1 and 3 (BF_{10} = .897). Finally, a paired samples t-test was conducted to explore the difference between trial-type 2 across both IRAPs, however this was not statistically significant (p = .09); a Bayesian paired samples t-test also provided no support for or against a difference with respect to this comparison (B10 = .853).

In summary, there was very strong evidence for a STTDE when participants completed the native language IRAP but not for the foreign language IRAP. Furthermore, the effect for trial-type 1 (Brazilian name-positive adjective) in the native language IRAP was significantly larger than all four trial-types from the foreign language IRAP. These findings are broadly consistent with the general argument that the Cfunc properties of words in a native language may dominate to a greater extent over their Crel properties relative to words in a foreign language, at least in the context of an IRAP. In drawing this conclusion, it should be acknowledged that the evidence is based on a homogenous group in terms of language history (i.e., all participants were native Portuguese speakers). A more robust test of the impact of Cfunc properties might involve comparing two separate groups that differ in terms of the assumed Cfunc properties of the stimuli presented within an IRAP. Experiment 2 adopted this general analytic strategy.

Experiment 2

Given that it would be very difficult to recruit participants who were native English speakers but fluent in Portuguese, at least in a Brazilian context, we employed different stimuli that we assumed would possess differential Cfunc properties across two groups. Specifically, we recruited two groups of participants who differed in terms of their support for specific Brazilian soccer teams. They were asked to complete a single IRAP which presented the names of the two teams along with their associated team badges. Critically, responding to the IRAP trial-types did not require any explicit evaluative response. That is, participants were simply required to categorize the team name and badge as true or false. If the Crel properties of the stimuli were dominating, little difference in patterns of IRAP effects would be expected between the two groups of fans (simply because they were just being asked to confirm which badge went with which team). If, however, the Cfunc properties of the stimuli play a significant role in the IRAP performances, differential patterns would be expected between the two groups. More specifically, opposing STTDEs may be observed that are consistent with the group's team allegiance. In other words, the largest IRAP effect should be observed for the trial-type that presents the name and badge for the group's supported team.

Method

Participants

Fifty-seven supporters from two of the main rival soccer teams in Minas Gerais state – Atletico and Cruzeiro –, who had at least a symbol of the team tattooed on some part of their body[2], were recruited by means of social network contacts. The Atletico group comprised 29 (23 males) and the Cruzeiro group comprised 28 (19 males) supporters. All participants were naïve in IRAP research and provided informed consent through an online form.

Setting, apparatus and procedure

Due to the pandemic scenario of 2021, the experimental sessions were conducted through remote access, by means of the Anydesk software (available at https://anydesk.com/pt/downloads/windows). Anydesk is a software that provides the remote connection between physically distant computers, as long as they are connected to the internet. The online monitoring application "Minha Conexão" ("my conection", available at https://www.minhaconexao.com.br) was used by both the researcher and the participant, and the procedure only started when the connection showed download and upload speeds of at least 50 Mbps. The IRAP software was installed on just one laptop to ensure a single keyboard delay and all remote access was directed to that machine. Data from each participant were collected individually in sessions that were monitored in real time by the researcher.

Each participant completed a single IRAP. On each trial, the name of one of two Brazilian soccer teams – either "Cruzeiro" or "Atletico" – was presented at the top of the screen as label stimuli, a single logo (i.e., the club badge) of either the Cruzeiro or Atletico team was presented at the center as target stimuli, and the response options – "True" and "False" (presented in Portuguese) – were presented at the two bottom corners of the screen ("True" on the left and "False" on the right). These label-target combinations yielded four individual trial types: Cruzeiro name-Cruzerio badge; Cruzeiro name-Atletico badge; Atletic name-Cruzeiro badge; Atletico name-Atletico badge.

At the beginning of blocks considered history consistent, the participants read: "Relate CRUZEIRO to a Cruzeiro's symbol and ATLETICO to an Atletico's symbol"; at the beginning of blocks considered history inconsistent, the participants read: "Relate ATLETICO to a Cruzeiro's symbol and CRUZEIRO to an Atletico's symbol" (also presented in Portuguese). These blocks were considered history consistent and inconsistent, respectively, in the sense that they required responding in accordance or opposition with presumed pre-experimentally established relations. For example, responding to the Atletico team name and the Cruzeiro team

[2] This participant selection criterion was based on the study by Haydu et al. (2015), which investigated the establishment of equivalent relations by "passionate" soccer supporters.

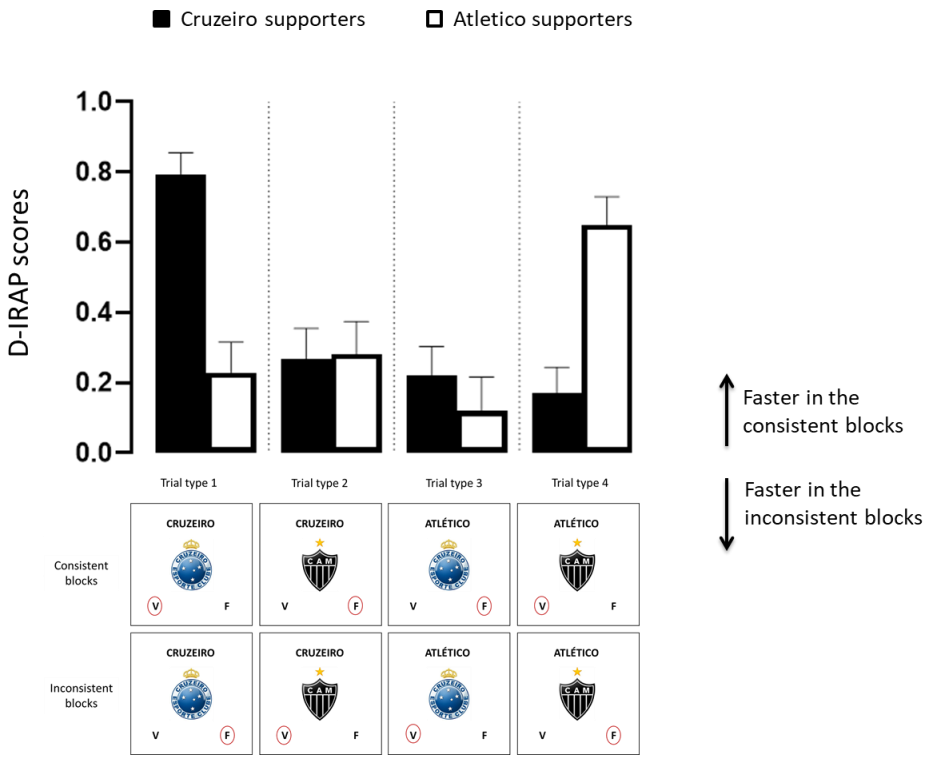
badge as “True” would generally be considered inconsistent with natural language contingencies in the verbal communities of these participants. All remaining aspects of the IRAP procedure and data processing were similar to Experiment 1.

Results and Discussion

Of the 57 participants who were exposed to the IRAPs, 39 successfully met the inclusion criteria, 18 in the Cruzeiro group and 21 in the Atletico group. The mean D-IRAP scores for these participants for the four IRAP trial-types in each group are presented in Figure 2. Note that, unlike Experiment 1, the D-IRAP scores for trial-types 3 and 4 were *not* inverted for Experiment 2 (i.e., divided by -1) because the IRAP was non-evaluative (i.e., both groups were asked to simply categorize the team names and badges).

Fig 2

Mean D-IRAP scores extracted for the 39 participants who met the IRAP inclusion criteria



Note. A higher D-IRAP score indicates a greater difference in response latencies between history consistent and history inconsistent blocks.

Visual inspection of Figure 2 indicates that responding on trial-types 1 and 4 differed dramatically depending on the soccer team participants supported. Specifically, participants who

supported the Cruzeiro team produced the largest D-IRAP score on the trial-type on which they were asked to categorize the word Cruzeiro and the badge for that team (trial-type 1). That is, these participants found it much easier to respond true rather than false to these stimuli appearing together. On balance, participants who supported the Atletico team produced the largest D-IRAP score on the trial-type on which they were asked to categorize the word Atletico and the badge for that team (trial-type 4). That is, these participants found it much easier to respond true rather than false to these stimuli appearing together. D-IRAP scores on the remaining trial-types did not appear to differ greatly from one another.

As mentioned previously, we assumed that the largest IRAP effect should be observed for the trial-type that presents the name and badge for the group's supported team. In order to test this assumption, and in line with the analyses conducted in Experiment 1, a series of paired-samples t-tests were conducted between trial-type 1 for the Cruzeiro group and all other trial-types across both groups, and between trial-type 4 for the Atletico group and all other trial-types across both groups. Both trial-types differed significantly from all other trial-types across groups (all p's < .01) apart from each other (p = .30; see Table 2). We also conducted Bayesian paired samples t-tests which provided no support for or against a difference with respect to the comparison between trial-type 1 from the Cruzeiro group and trial-type 4 from the Atletico group (BF_{10} = .399); however, there was very strong evidence in support of the difference between trial-type 1 from the Cruzeiro group and trial-type 2 from the Atletico group (BF_{10} = 42.895); very strong evidence for the difference between trial-type 4 from the Atletico group and trial-types 1 (BF_{10} = 72.433) and 2 (BF_{10} = 97.852) from the same group, and trial types 3 (BF_{10} = 43.955) and 4 (BF_{10} = 38.850) from the Cruzeiro group; moderate evidence for the difference between trial-type 4 from the Atletico group and trial-type 2 from the Cruzeiro group (BF_{10} = 8.414); and extremely strong evidence for all other comparisons (all BF_{10}s >/= 230.430).

Table 2

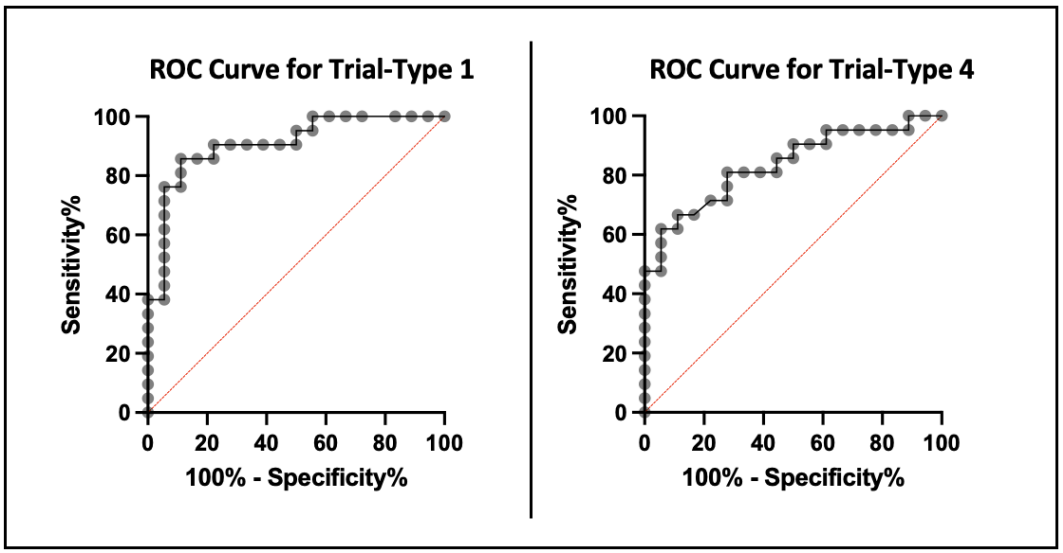
Paired samples t-tests between the two assumedly critical trial-types for each group and all other trial-types

						95% CI for Cohen's d	
Measure 1	Measure 2	*t*	*df*	*p*	Cohen's *d*	Lower	Upper
Cruzeiros TT1	Cruzeiros TT2	6.375	17	< .001	1.503	0.811	2.174
Cruzeiros TT1	Cruzeiros TT3	5.631	17	< .001	1.327	0.677	1.957
Cruzeiros TT1	Cruzeiros TT4	6.451	17	< .001	1.520	0.824	2.196
Cruzeiros TT1	Atletico TT1	5.553	17	< .001	1.309	0.663	1.934
Cruzeiros TT1	Atletico TT2	4.038	17	< .001	0.952	0.382	1.503
Cruzeiros TT1	Atletico TT3	4.935	17	< .001	1.163	0.550	1.756
Cruzeiros TT1	Atletico TT4	1.068	17	0.300	0.252	-0.221	0.718
Atletico TT4	Atletico TT3	5.155	20	< .001	1.125	0.566	1.667
Atletico TT4	Atletico TT2	4.334	20	< .001	0.946	0.420	1.455
Atletico TT4	Atletico TT1	4.189	20	< .001	0.914	0.394	1.418
Atletico TT4	Cruzeiros TT2	3.159	17	0.006	0.745	0.212	1.261
Atletico TT4	Cruzeiros TT3	4.051	17	< .001	0.955	0.384	1.506
Atletico TT4	Cruzeiros TT4	3.986	17	< .001	0.939	0.372	1.488

Given the very large differences between the two groups on trial-types 1 and 4, we sought to determine, post-hoc, the predictive ability of the IRAP performances in terms of indicating which team a participant supported. To do this, we employed a signal detection test which involved constructing a Receiver Operator Characteristic (ROC). A ROC is a graph in which the probability of a true positive is plotted against the probability of a false positive (Fawcett, 2006). An area under the curve (AUC) can then be calculated from this analysis; that is, the statistical likelihood that a randomly chosen member of one group (e.g., Cruzeiro fans in the current study) will have a higher score than a randomly chosen member of the other group (e.g., Atletico fans). As such, an IRAP performance with a perfect ability to predict group membership would have an AUC = 1.0, and a IRAP performance with no such ability would have an AUC = ~0.5. We conducted two ROC analyses; one for the supporter's groups for trial-type 1 and one for the two groups for trial-type 4. The ROC analysis for trial-type 1 showed a strong ability to identify team allegiance (AUC = .91, $p < .0001$; see Figure 3 left panel), and the analysis for trial-type 4 showed a slightly lower but still relatively strong predictive ability (AUC = .83; $p = .0004$; see Figure 3 right panel).

Fig 3

ROC curves for the two assumedly critical trial-types in Experiment 2



In summary, opposing STTDEs were observed consistent with the participants' team allegiance. Specifically, the largest IRAP effect observed for each group was for the trial-type that presented the name and badge for that group's supported team. Furthermore, these two trial-types reliably predicted group membership. These differential patterns of results indicate that the Cfunc properties of the stimuli played a substantial role in producing the differential patterns of IRAP performances observed. In other words, because participants were not asked to evaluate the teams but rather simply categorize the team name with their badge, the dramatically different IRAP effects for trial-types 1 and 4 cannot be explained by appealing to their Crel properties alone.

General Discussion

The current study sought to explore the extent to which the variables highlighted within the DAARRE model can be used to explain some of the effects produced on the IRAP; in particular, the relative dominance of the Cfunc properties of the stimuli over their Crel properties. The first experiment explored this suggestion by requiring participants to complete IRAPs in both their native and a foreign language. The assumption here was that the Cfunc properties of words in a native language may dominate to a greater extent over their Crel properties relative to words in the non-native language. The second experiment sought to extend Experiment 1 but in a known-groups context involving stimuli that were assumed to possess differential Cfunc properties across two groups. In this case, if the Cfunc properties of the stimuli played a significant role in producing IRAP performances, differential patterns would be expected between the two groups (little difference would be expected given Crel dominance).

The results of both experiments, therefore, seem to provide support for the DAARRE model analysis. Specifically, strong STTDEs were evident in both experiments and were predicted by learning history, through native language in Experiment 1 and more formally (using a ROC analysis) by group membership in Experiment 2. Overall, the current findings thus indicate that the Cfunc properties of stimuli seem increasingly important in interpreting patterns of effects produced on the IRAP.

It is worth emphasizing once again that the different patterns produced between the two groups of soccer supporters in Experiment 2 emerged in the context of an IRAP that simply asked the participants to categorize the names and badges of the two teams. That is, they were not asked to *relate* the team names or badges to either positively or negatively valenced stimuli. As such, the differential patterns cannot be interpreted in terms of such relational properties but instead seem to require an analysis based on the coherence among the Cfunc properties of the individual stimuli and the response options. Such a result highlights the fact that an IRAP performance may be highly sensitive to specific properties of the stimuli presented therein, even when the task itself (e.g., simply categorizing the names and badges of soccer teams) was not designed to target those properties. At the very least, this finding serves to highlight the potential pitfall in relying upon individual IRAP trial-type effects as a reliable proxy for psychological variables that a researcher aims to target in a specific study. In Experiment 2, for example, the IRAP task simply required participants to categorize names and badges of football teams. Psychologically, all that was being targeted here was categorizing the teams and badges correctly (versus incorrectly). However, dramatic differences emerged between the two groups, indicating that the IRAP effects were driven by more than simply categorizing. It follows, therefore, that any attempt to use the IRAP as a measure of a particular psychological construct should be approached with extreme caution, and at the very least informed by a relatively thorough functional-analytic understanding of the IRAP tool itself (see Barnes-Holmes & Harte, 2022).

On a related point, the current findings may be directly relevant to a reinterpretation of a previously published IRAP study that employed the tool as a measure of attitudes. Specifically, Power et al. (2017) compared IRAP performances using stimuli designed to assess racial biases among white Irish participants and black African residents who had recently immigrated to Ireland. The white participants produced large differences among the trial-types that were interpreted at the time as indicative of racial bias. In contrast, the black participants produced relatively small differences among the trial-types, suggesting the absence of racial bias. In light of the results of Experiment 1 of the current study, perhaps the lack of differential trial-type effects for the latter group of participants was driven, at least in part, by the fact that they were completing the task in a non-native language. At this point of course, we can only speculate, but

once again the current results highlight another potential pitfall in interpreting IRAP effects in the absence of a clear functional-analysis of the instrument itself.

Another IRAP effect that has been increasingly reported in recent years and is readily explained by the DAARRE model is referred to as the dissonant target trial-type effect (DTTTE; e.g., Finn et al., 2019; Kavanagh et al., 2019; Schmidt et al., 2021). This IRAP pattern is characterized by participants finding it easier to select the response option that is functionally similar to the target stimulus on a given trial than the response option that is functionally dissimilar. For example, participants may find it easier to choose 'No' on trials with a negative adjective as the target stimulus given that both are assumed to have negative Cfunc properties (relative to positive adjectives and the response option 'Yes'). Although we did not focus on making predictions about the DTTTE in the current set of experiments, it is interesting to note that no clear visual evidence for this effect emerged. One potential reason for this may be to do with the fact that the response options remained fixed for the duration of both experiments (e.g., "Yes" and "True" always appeared on the left-hand side of the screen). Given that the DTTTE may be explained by the functional overlap between the Cfunc properties of the target stimulus and response option, perhaps fixing the location of the response options undermined this effect. That is, as participants completed the IRAP, the two fixed response locations overrode the stimulus properties of the on-screen stimuli (e.g., "Yes" and "No"). Indeed, the impact of randomizing the response options in IRAP trials versus keeping their positions fixed has received little attention (Campbell et al., 2011). In any case, this point serves to highlight the importance of developing a thorough functional analysis of the variables involved in the IRAP.

A related issue is the fact that Experiment 1 of the current study employed the response options "Yes" and "No" while Experiment 2 employed "True" and "False." Indeed, previous research has demonstrated that the type of response options employed within an IRAP can differentially impact upon the patterns of effects observed (Maloney et al., 2016, 2020; Murphy et al., 2022). However, such research has typically made a distinction between the use of Crels as response options (e.g., "similar", "different") versus those employed currently (i.e., "True", "False", "Yes", "No"), referred to as relational coherence indicators (RCIs; see Maloney et al., 2016). Thus, given that both sets of response options employed here are RCIs, this may not present an issue in the current context. Nonetheless, given the subtlety of the behavioral dynamics involved in producing patterns of IRAP effects (i.e., relative degrees of coherence between Cfunc and Crel properties and response options), there may be currently unknown subtle differences involved between the use of these two sets of RCIs. Future research could explore these potential subtleties.

To conclude, the current study may be seen as part of a broader effort to develop an increasingly precise functional analysis of the controlling variables that are typically brought to bear by the IRAP as a behavior-analytic tool. Although there are limitations to the current study,

including the use of relatively small sample sizes and using fixed-location response options, the findings are broadly in line with a growing body of research that has identified some of the key variables that determine IRAP performances. Given that some of these variables may be interpreted as spurious (e.g., language proficiency) when using the IRAP as a simple proxy for psychological constructs, it seems important to alert other researchers to these issues. Perhaps the most important take-home message from the current study is the critical role played by the Cfunc properties of the stimuli employed within an IRAP, even when participants are not required to relate the stimuli based on these properties. Although this conclusion potentially undermines the IRAP as a simple measure of implicit cognition, it appears to provide a useful experimental context for analyzing the complex cluster of interacting variables involved in the behavior-analytic study of human language and cognition.

Data availability

The datasets involved in the current study are available from the corresponding authors upon reasonable request.

Declarations

Conflicts of Interest. The authors declare no known conflicts of interest.

Ethical Approval. All procedures performed involving human participants were in accordance with the ethical standards of the institutional and national research committee, as well as with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Consent to Participate. All participants provided consent prior to participation. The study was conducted in coherence with all necessary ethical standards and was approved by the appropriate ethics review boards. All the authors have approved the manuscript and agree with submission to the journal.

Compliance with Ethical Standards. This manuscript has not been published or presented elsewhere, in part or in its entirety, and is not under consideration by another journal.

References

Barnes-Holmes, D., Hayden, E., Barnes-Holmes, Y., & Stewart, I. (2008). The implicit relational assessment procedure (IRAP) as a response-time and event-related-potentials methodology for testing natural verbal relations: A preliminary study. *The Psychological Record, 58*, 497–516. https://doi.org/10.1007/BF033 95634

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. https://doi.org/10.1007/BF03395726

Barnes-Holmes, D. & Harte, C. (2022a). The IRAP as a Measure of Implicit Cognition: A Case of Frankenstein's Monster. *Perspectives on Behavior Science, 45*, 559-578. https://doi.org/10.1007/s40614-022-00352-z

Barnes-Holmes, D. & Harte, C. (2022b). Relational frame theory 20 years on: The Odysseus Voyage and Beyond. *Journal of the Experimental Analysis of Behavior, 17*(2), 240-256. https://doi.org/10.1002/jeab.733

Bortoloti, R., de Almeida, R. V., de Almeida, J. H., & de Rose, J. C. (2019). Emotional faces in symbolic relations: A happiness superiority effect involving the equivalence paradigm. *Frontiers in Psychology, 10*, 1–12. https:// doi. org/ 10. 3389/ fpsyg. 2019. 00954

Bortoloti, R., de Almeida, R. V., de Almeida, J. H., & de Rose, J. C. (2020). A commentary on the dynamics of arbitrarily applicable relational responding involving positively valenced stimuli and its implications for the IRAP research. *The Psychological Record, 71*, 481–486. https:// doi. org/ 10.1007/ s40732- 020- 00413-2

Caldwell-Harris, C. L. (2015). Emotionality differences between a native and foreign language: Implications for everyday life. *Current Directions in Psychological Science, 24*, 214–219. https://doi.org/10.1177/0963721414566268

Campbell, C., Barnes-Holmes, Y., Barnes-Holmes, D., & Stewart, I. (2011). Exploring screen presentations in the implicit relational assessment procedure (IRAP). International *Journal of Psychology and Psychological Therapy, 11*(3), 377-388.

Circi, R., Gatti, D., Russo, V., & Vecchi, T. (2021). The foreign language effect on decision-making: A meta-analysis. *Psychonomic Bulletin & Review*, *28*, 1131–1141. https://doi.org/10.3758/s13423-020-01871-z

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters, 27*, 861-874. https://doi.org/10.1016/j.patrec.2005.10.010

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*(2), 309–321. https://doi.org/10.1007/s40732- 016- 0173-4

Finn, M., Barnes-Holmes, D., & McEnteggart, C. (2018). Exploring the single-trial-type-dominance- effect on the IRAP: Developing a differential arbitrarily applicable relational responding effects (DAARRE) model. *The Psychological Record, 68*(1), 11–25. https:// doi. org/ 10. 1007/s40732- 017- 0262-z

Finn, M., Barnes-Holmes, D., McEnteggart, C., & Kavanagh, D. (2019). Predicting and influencing the single trial-type dominance effect. *The Psychological Record, 69*(3), 425–435. https://doi.org/10.1007/s40732-019-00347-4

Haydu et al. (2015). Effects of preexperimental history on the formation of stimulus equivalence classes: A study with supporters of Brazilian soccer clubs. *Psychology & Neuroscience, 8*, 385-396. http://dx.doi.org/10.1037/h0101276

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

Kavanagh, D., Matthyssen, N., Barnes-Holmes, Y., Barnes-Holmes, D., McEnteggart, C., & Vastano, R. (2019). Exploring the use of pictures of self and other in the IRAP: Reflecting upon the emergence of differential trial type effects. *International Journal of Psychology and Psychological Therapy, 19*(3), 323-336.

Maloney, E. & Barnes-Holmes, D. (2016). Exploring the behavioral dynamics of the implicit relational assessment procedure: The role of relational contextual cues and relational coherence indicators. *The Psychological Record, 66*, 395-403. https://doi.org/10.1007/s40732-016-0180-5

Maloney, E., Foody, M., & Murphy, C. (2020). Do response options in the implicit relational assessment procedure (IRAP) matter: A comparison of contextual relations versus relational coherent indicators. *The Psychological Record, 70*, 205- 214. https://doi.org/10.1007/s40732-019-00360-7

Mann, T. (1927). *The magic mountain*. New York: Alfred A. Knopf.

Murphy, C., Maloney, E., & Kelly, M. (2022). The role of relational contextual cues and relational coherence indicators on the implicit relational assessment procedure. *The Psychological Record, 72*, 665-673. https://doi.org/10.1007/s40732-022-00512-2

Pavlenko, A. (2017). Do you wish to waive your rights?: Affect and decision-making in multilingual speakers. *Current Opinion in Psychology, 17*. https://doi.org/10.1016/j.copsyc.2017.06.005

Pidgeon, A., McEnteggart, C., Harte, C., Barnes-Holmes, D., & Barnes-Holmes, Y. (2021). Four selfrelated IRAPs: Analyzing and interpreting effects in light of the DAARRE model. *The Psychological Record, 71*, 397–409. https://doi.org/10.1007/s4073202000428-9

Pinto, J. A. R., de Almeida, R. V., & Bortoloti, R. (2020). The stimulus' orienting function may play an important role in IRAP performance: Supportive evidence from an eye-tracking

study of brands. *The Psychological Record, 70*, 257–266. https://doi.org/10.1007/s40732- 020- 00378-2

Power, P.M., Harte, C., Barnes-Holmes, D., & Barnes-Holmes, Y. (2017). Exploring racial bias in a European country with a recent history of immigration of black Africans. *The Psychological Record, 67*, 365-375. https://doi.org/10.1007/s40732-017-0223-6

Schmidt, M., de Rose, J. C., & Bortoloti, R. (2021). Relating, orienting and evoking functions in an IRAP study involving emotional pictographs (emojis) used in electronic messages. *Journal of Contextual Behavioral Science, 21*, 80–87. https://doi.org/10.1016/j.jcbs.2021.06.005