Exploring Differential Trial-type Effects and the Impact of a Read-Aloud Procedure On Deictic Relational Responding on the IRAP

Deirdre Kavanagh, Yvonne Barnes-Holmes, Dermot Barnes-Holmes, Ciara McEnteggart, & Martin Finn

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

deirdre.kavanagh@ugent.be

yvonne.barnesholmes@ugent.be

dermot.barnes-holmes@ugent.be

ciara.mcenteggart@ugent.be

martin.finn@ugent.be

Author Note

The data for the current manuscript was collected, and prepared with the support of the FWO Type I Odysseus Programme at Ghent University, Belgium. We are very grateful to the reviewers for their thorough and constructive commentaries.

Correspondence should be addressed to Deirdre Kavanagh, Department of Experimental, Clinical and Health Psychology, Ghent University, 9000 Ghent, Belgium,

Email: deirdre.kavanagh@ugent.be

Abstract

Under the rubric of Relational Frame Theory (RFT), researchers have investigated the role of deictic relational responding in the analysis of self in relation to others, place, and time, primarily through the use of an extended developmental protocol (Barnes-Holmes, 2001). In a move towards extending methodologies for studying deictic relational responding, more recent research has employed the Implicit Relational Assessment Procedure (IRAP) to measure deictic relational responding regarding I versus OTHER (Barbero-Rubio et al., 2016). The initial purpose of the current study was to partially replicate and extend this research. This extension involved the inclusion of a control condition in which no responding to self was involved, only responding to others. The results from Experiment 1 yielded significant IRAP effects for two of the four trial-types in both the Deictic and Control IRAPs. A second experiment involved a novel method for collecting IRAP data (a read-aloud procedure), which had been shown to yield significant effects for all four trial-types, and four significant effects were indeed recorded for both Deictic and Control IRAPs. Based on the current findings, a model is presented that seeks to explain the differential trial-type effects that are observed across the different IRAPs and the impact of the read-aloud procedure.

Keywords: Relational Frame Theory, IRAP, deictic, DAARRE Model

Since the seminal work of Sidman in the 1970s, behavior-analytic researchers have developed increasingly complex accounts of human language and cognition in terms of derived stimulus relations. A particularly rich vein of research in this regard is known as relational frame theory (RFT), which led to the publication of a full book-length treatment of human language and cognition in 2001 (Hayes, Barnes-Holmes, & Roche). One of the domains targeted in the book was the role of derived relational responding in the analysis of self in relation to others, time, and place. Specifically, three core deictic relations were identified. The *interpersonal* relations involve responding to *I* and *you*, the *spatial* relations involve responding to *here* and *there*, and the *temporal* relations involve responding to *now* and *then*.

There have been many studies on the deictic relations, most employing the Barnes-Holmes (2001) written protocol for assessing and establishing these relations when they are found to be absent in young children. The results of this body of research may be summarized as follows. (1) The data support the distinctions among the three types of deictic relations (McHugh, Barnes-Holmes, & Barnes-Holmes, 2004). (2) The deictic relations vary on a continuum of complexity from simple relations (e.g., I versus you) to reversed relations (e.g., if I were you and you were me), and even double reversed relations (e.g., if I were you and you were me, and if here were there and there were here, see McHugh et al.). (3) Deictic relations can be established if they are found to be absent, and when trained, they generalize to natural language (Heagle & Rehfeldt, 2006; Weil, Hayes, & Capurro, 2011). (4) There appears to be a developmental trend in which the interpersonal and simple relations emerge first (McHugh et al.). (5) Competence in deictic relational responding correlates with cognitive abilities and IQ (Gore, Barnes-Holmes, & Murphy, 2010). (6) There is overlap between competence in deictic responding and traditional theory of mind (ToM) tests (Weil et al.).

Additional research has attempted to use the Barnes-Holmes (2001) protocol to explore the potential role of the deictic relations in adult psychological suffering. For example, Villatte, Monestés, McHugh, Baque, and Loas (2008) compared individuals with high versus low self-reported social anhedonia. While both groups showed strong overall accuracy, some superiority was observed for the low anhedonia group on the I-YOU reversals and the I-YOU/HERE-THERE double reversals. A replication of this study by Villatte, Monestés, McHugh, Freixa i Baqué, and Loas (2010) with individuals with a diagnosis of schizophrenia similarly reported some superiority for a control group on I-YOU and NOW-THEN reversals. The authors concluded that relative *deficits* in deictic relational responding may constitute a feature of these types of psychological suffering.

The use of the Barnes-Holmes (2001) deictic protocol to draw conclusions about clinical phenomena has been criticized on several grounds (Hussey et al., 2014). (1) The protocol was explicitly designed for developmental purposes (i.e., use with young children) to establish deictic relations when they were found to be absent or deficient. (2) McHugh et al. (2004) demonstrated that even typically-developing adults show deficits on specific deictic relations when these relations are not presented as they typically are in natural language (see also Vitale, Barnes-Holmes, Barnes-Holmes, & Campbell, 2008). (3) RFT does not necessarily predict that psychological suffering involves *deficits* (rather than excesses) in relational responding. (4). It is *possible* that deficits or unexpected patterns of deictic relational responding in general (e.g., which color chair you are sitting on relative to someone else) *might* be observed in psychological suffering, but more meaningful effects would likely be obtained if the deictic relations were specific to the domain of interest (e.g., your levels of anxiety relative to others). Based on some of these criticisms, a recent line of research has sought to adapt an RFT-based methodology to study deictic relations, with a particular focus on clinical domains.

Specifically, in parallel with much of the research on deictic relations, some RFT researchers have been working on a procedure that would allow them to catch relational framing "in flight" (i.e., as it occurs in vivo in the natural environment). This has resulted in the development of the implicit relational assessment procedure (IRAP), which allows the researcher to juxtapose alternative relational responses and thus obtain a measure of the relative strength or probability of specific patterns of relational responding. The IRAP typically presents label and target stimuli (e.g., "pleasant" with a picture of a flower) and requires participants to confirm or disconfirm the relational coherence between them (i.e., "true" on coherent trials and "false" on incoherent trials). Thus, IRAPs comprise four trial types (e.g., *Flower-Pleasant Flower-Unpleasant, Insect-Pleasant, and Insect-Unpleasant*) that are generally analyzed independently in terms of the difference in response latencies between responding that is deemed consistent (coherent) versus inconsistent (incoherent) with a participant's verbal history. The resultant IRAP effects are often normalized using a *D*-algorithm, although some studies have analyzed response latencies directly. The body of empirical support for the IRAP has now reached over 50 published studies, with an increasing focus on clinically relevant phenomena (Vahey, Nicholson, & Barnes-Holmes, 2015). Employing the IRAP as an instrument for assessing deictic relational responding, particularly in the clinical domain, would provide an alternative to the Barnes-Holmes (2001) protocol. A recent study in which the IRAP was used to target responding to self versus others seems particularly relevant (Barbero-Rubio, Lopez-Lopez, Luciano, & Eisenbeck, 2016).

The study presented participants with their own names and the name of the researcher as label stimuli, and statements pertaining to specific characteristics of the self versus other as targets (e.g., "is in front of the laptop"). There were two response options ("yes" and "no") on each trial. The four trial types in this study were referred to as: *I-I* (participant name-participant characteristics), *Other-Other* (researcher name-researcher characteristics), *I-Other* (participant name-researcher characteristics), and *Other-I* (researcher name-participant characteristics). In general, the pattern of IRAP effects reported by Barbero-Rubio et al. (2016) indicated that participants' response latencies showed significantly more rapid responding on the *I-I* trial type relative to the other three trial types during consistent blocks. In addition, the difference in response latencies between consistent and inconsistent blocks for each trial type was in the predicted direction (i.e., shorter on consistent relative to inconsistent trials), and these differences were significant in terms of the normalized D_{IRAP}-scores.

The initial purpose of the current study was to conduct a systematic replication[1] and extension of Barbero-Rubio et al. (2016). The extension of the work involved the inclusion of a control condition in which no responding to self was involved, only to others. That is, the control condition involved an IRAP in which none of the trial types required responding to self, but only responding to two other separate individuals (i.e., the experimenter and a picture of another unknown participant). If the comparison between self and other in a deictic IRAP is an important variable, one might expect a different pattern of results in a control condition in which there is no contrast between self and other.

The results in Experiment 1 yielded significant IRAP effects for two of the trial types in both the deictic and control IRAPs, whereas four significant effects were observed in Barbero-Rubio et al. in which there was no control condition. A second experiment was undertaken that involved a novel method for collecting IRAP data, which had been shown to yield significant effects for all four trial types in a separate line of research being conducted by our group (Finn, Barnes-Holmes, & McEnteggart, in press). Specifically, we had found that relatively extreme differential trialtype effects were reduced when participants were asked to read aloud the stimuli and response options that appeared on each IRAP trial. Given

[1] It is important to emphasize that the current research was a systematic, rather than direct, replication of Barbero-Rubio et al. (2016), in that there were numerous procedural differences between the earlier work and the present research (see Sidman, 1960).

that relatively extreme differential trial-type effects were observed in Experiment 1 of the current study, we introduced the read-aloud procedure in Experiment 2. The results of this second experiment appear relevant to future research that will attempt to use the IRAP to study deictic relational responding and perspective-taking more generally; we will return to this issue in the General Discussion.

One final way in which the current research extended Barbero-Rubio et al. (2016) was the inclusion of self-report measures of self-esteem and the presence of psychotic-like experiences. We also retained one measure of perspective-taking employed in the previous study. These were included on an exploratory basis, hence no specific predictions were made. In reflecting upon the results we obtained here and in our other studies, we have begun to develop a model of the differential trial-type effects that are observed across different IRAPs (see Barnes-Holmes, Finn, McEnteggart, & Barnes-Holmes in press; Finn et al. in press). We will outline this model in the General Discussion because it has emerged inductively as we have undertaken the very research reported here.

Experiment 1

Method

Participants. Forty participants were recruited for Experiment 1, 28 females and 12 males. They ranged from 18-36 years old (M = 23.34). All participants were recruited through random convenience sampling from the Ghent University participant pool and were paid an hourly rate of 10 euro.

Materials and apparatus. Experiment 1 comprised three computer-based tasks: a familiarization IRAP, a deictic IRAP, and a control IRAP. In all three, participants were required to respond to two others rather than to the self versus others. The study also included three questionnaires: the Community Assessment of Psychic Experiences (CAPE), the Interpersonal Reactivity Index (IRI; perspective-taking sub-scale only), and the Rosenberg

Self-esteem Scale (RSES). All materials were presented in Dutch (but they are translated into English when referred to in the text of the current paper).

Familiarization IRAP. The IRAP was presented on standard personal computers. The IRAP software was used to present the instructions, the stimuli, and to record responses. The familiarization IRAP did not contain stimuli relevant to deictic relations and was employed simply to familiarize participants with the procedure because no specific pre-block rules for responding were presented in any of the IRAPs. The familiarization IRAP presented two label words at the top of the screen: *Fruits* and *Vegetables* (see Table 1). Eight target words were individually presented in the center of the screen; four were fruits (e.g., “Pear”) and four were vegetables (e.g., “Broccoli”). The response options “Yes” and “No” were presented at the bottom left- and right-hand corners. The familiarization IRAP comprised four trial types: *Fruit-Fruit, Vegetable-Vegetable, Fruit-Vegetable, and Vegetable-Fruit* (see Figure 1).

INSERT TABLE 1 & FIGURE 1 HERE

Deictic IRAP. The deictic IRAP presented two label stimuli (participant name/researcher name) on the top of the screen (see Table 2). The target stimuli comprised 12 statements; six described characteristics of the participant at the present time (e.g., “has a yellow Post-it”), and six described characteristics of the researcher (e.g., “has an orange Post-it”). Again, the response options were “Yes” and “No”. The four trial types were denoted as: *I-I* (participant name-participant characteristics), *Other-Other* (researcher name-researcher characteristics), *I-Other* (participant name-researcher characteristics), and *Other-I* (researcher name-participant characteristics).

INSERT TABLE 2 HERE

Control IRAP. The control IRAP presented the label stimuli “Dee” (researcher’s name) and “Ciara” (name of an individual, whose picture was placed on the wall in front of participants, see Table 3). To match the deictic IRAP, the target stimuli comprised 12 statements; six described features of the researcher (e.g., "has brown hair"), and six described features of the person in the picture (e.g., "has blond hair"). Again, "Yes" and "No" were the response options. In denoting the four trial types, the term *Researcher* will be used to refer to the actual researcher, and the term *Other* will be used to refer to the person shown in the picture. Please note that all four trial types involved responding to another and not the self. The four trial types were thus denoted as: *Other-Other* (picture of other and characteristics of other), *Researcher-Researcher* (i.e., researcher and researcher characteristics), *Other-Researcher* (picture of other and researcher characteristics), and *Researcher-Other* (researcher and characteristics of other).

INSERT TABLE 3 HERE

The Community Assessment of Psychic Experiences (CAPE; Stefanis et al., 2002). The CAPE measures psychotic-like experiences in the general population. The scale consists of 42 symptom items rated along three subscales: positive symptoms (20 items, e.g., "Do you ever feel as if there is a conspiracy against you?"), negative symptoms (14 items, "Do you ever feel that you experience few or no emotions at important events?") or depressive symptoms (eight items, "Do you ever feel sad?"). Each item is rated on two 4-point Likert scales from 0 (*never*) to 3 (*nearly always*) to indicate (1) the frequency of symptoms and (2) the level of distress associated with each symptom. The CAPE provides overall frequency and distress scores of psychic experiences and total frequency and distress scores for each of the three subscales. In order to account for partial non-responses, all scores are weighted for the number of valid answers per subscale (i.e., sum score divided by number of items completed). Overall frequency and distress scores were also weighted. In all cases, higher scores indicate greater frequency or distress regarding symptoms, but there are no clinical cut-offs for this measure. The Dutch version was completed by participants. The scale has demonstrated adequate reliability: positive dimension alpha = 0.63, negative dimension alpha = 0.64, and depressive dimension alpha = 0.62 (Konings, Hanssen, Van Os, & Krabbendam, 2006).

The Rosenberg Self-Esteem Scale (RSES; Robins, Hendin, & Trzesniewski, 2001). The RSES is a 10 item measure of self-esteem. All items (e.g., "I take a positive attitude toward myself") are rated on a four-point scale from 0 (*strongly agree*) to 3 (*strongly disagree*). The RSES yields an overall score, with a maximum of 30 and a minimum of 0. The Dutch version has shown high internal consistency (Cronbach's alpha = .86) and high congruent validity (Franck, De Raedt, Barbez, & Rosseel, 2008).

Perspective-taking (PT) sub-scale of the Interpersonal Reactivity Index (IRI; Davis, 1980). The PT subscale of the IRI measures perspective-taking. The subscale consists of seven items (e.g., "I sometimes find it difficult to see things from the other guy's point of view") rated on a five-point scale from 1 (*does not describe me*) to 5 (*describes me very well*). The subscale yields an overall score, with a maximum of 35 and a minimum of 1. High scores indicate strong perspective-taking, and low scores indicate weak perspective-taking. There are no clinical cut-offs for this measure. The Dutch version has demonstrated good internal consistency (Cronbach's alpha = .73), and construct validity (De Corte et al., 2007).

Procedure. All procedures in the current study were in accordance with the ethical standards of the institutional research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards. Informed consent was obtained from all individual participants. All participants were exposed to the same experimental sequence, as follows: familiarization IRAP, deictic IRAP, control IRAP, RSES, PT scale, and the CAPE.

Familiarization IRAP. The familiarization IRAP was employed to establish competent performances on a simple word-based IRAP (*Fruits* vs. *Vegetables*) prior to completion of the deictic IRAP. Participants were simply instructed to figure out, based on individual trial feedback, what the task involved. Consider a trial with the label "Fruits" and the target “Pear”. 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.

Consistent trial blocks required responding that was in accordance with the pre-experimental verbal history of the participants: *Fruit-Fruit/Yes, Vegetable-Vegetable/Yes, Fruit-Vegetable/No,* and *Vegetable-Fruit/No*. *Inconsistent* trial blocks required responding that was *not* in accordance with pre-experimental verbal relations: *Fruit-Fruit/No, Vegetable-Vegetable/No, Fruit-Vegetable/Yes,* and *Vegetable-Fruit/Yes*. The familiarization IRAP always commenced with a consistent block of trials. When participants selected the response option that was deemed correct within that block, the label, target, and response-option stimuli were immediately removed from the screen, and the next trial was presented after an inter-trial interval of 400 ms (the label, target, and response-option stimuli then appeared simultaneously at the beginning of the next trial). When participants selected the response option that was deemed incorrect for that block, the stimuli remained on the screen and a red “X” appeared beneath the target stimulus. 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). Participants were required to achieve both accuracy (≥ 80% correct responding) and latency criteria (≤ 2,000 milliseconds) in every block. As is typical in IRAPs, performance feedback was presented at the end of each block. The program automatically recorded response accuracy (based on the first response emitted on each trial) and response latency (time in ms between trial onset and the emission of a correct response) on each trial.

The familiarization IRAP differed from a typical IRAP in that it contained only practice blocks (i.e., these were not followed by test blocks). Participants were exposed to a maximum of three pairs of blocks, with 24 trials per block (12 for each type of target stimulus, fruit or vegetable). If participants achieved both accuracy and latency criteria on the first, second, or third pair of blocks, they proceeded to the deictic IRAP. All participants completed the familiarization IRAP within three sets of blocks.

Deictic IRAP. The format of the deictic IRAP was identical to Barbero-Rubio et al. (2016), except that explicit rules were not provided (it was assumed that the necessary competence on the task had been established by the familiarization IRAP). The deictic IRAP comprised a maximum of four pairs of practice blocks, followed by three pairs of test blocks. On each trial, there was a label at the top of the screen (*participant's name* or *researcher's name*), a target in the center of the screen (e.g., "is the participant" or "is the researcher"), and two response options ("Yes" and "No") at the bottom left and right of the screen. 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.

When participants selected the response option that was deemed correct within that block, an inter-trial interval of 400 ms was presented, after which the next trial occurred. When participants selected the response option that was deemed incorrect for that block, the stimuli remained on the screen and a red "X" appeared beneath the target stimulus. Only when the correct response option was selected did the program proceed to the 400 ms inter-trial interval (followed by the next trial). This pattern of trial presentations, with corrective feedback, continued until the entire block of 24 trials was presented. Trials were presented 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 blocks required responding that was in accordance with pre-experimental verbal relations: *I-I/Yes, Other-Other/Yes, I-Other/No,* and *Other-I/No.* Inconsistent blocks required the opposite: *I-I/No, Other-Other/No, I-Other/Yes,* and *Other-I/Yes.* Again, all participants experienced a consistent block first.

When participants completed a block of trials, the IRAP program delivered feedback on their performance during that block. The feedback consisted of a message informing them how accurately and how quickly they had responded. The latter was calculated from stimulus onset to the first correct response across all 24 trials within the block. Participants were required to achieve a minimum accuracy of 80% correct and a maximum median latency of no more than 2000 ms on each block. If participants achieved both accuracy and latency criteria on the first, second, third, or fourth pair of practice blocks, they proceeded to the first pair of test blocks; if they failed on the fourth pair of practice blocks participation in the experiment was terminated.

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

Control IRAP. The format of the control IRAP was similar to the deictic IRAP. Consistent blocks required responding that was in accordance with pre-experimental verbal relations: *Other-Other/Yes*, *Researcher-Researcher/Yes*, *Other-Researcher/No*, and *Researcher-Other/No*. Inconsistent blocks required the reverse: *Other-Other/No, Researcher-Researcher/No*, *Other-Researcher/Yes*, and *Researcher-Other/Yes*. Again, all participants experienced a consistent block first.

Questionnaires. Participants completed the three questionnaires in the following sequence: the PT subscale, the RSES, and the CAPE.

Results

The data and analyses for each IRAP and questionnaire are presented separately. Given the absence of test blocks in the familiarization IRAP, no data are presented from this procedure. Barbero-Rubio et al. (2016) reported raw reaction times as well as D_{IRAP} scores. In the current study, we did not report raw reaction times because we included a control condition in which the stimuli were different across IRAPs. Thus, any comparison in raw reaction times across these IRAPs would be difficult to interpret. Furthermore, as will become clear later, the introduction of the novel procedure in Experiment 2 (outlined subsequently) further undermines raw reaction times as a meaningful metric in the context of the current study. In the absence of analyses of the raw reaction times, we conducted analyses on the D_{IRAP} scores that were not in Barbero-Rubio et al.

IRAP Data. Mean response latencies for consistent and inconsistent blocks, divided according to trial type, were calculated for each participant. Specifically, in order to pass the practice blocks and advance to test blocks, participants were required to maintain an accuracy level of ≥ 80% correct and a median latency of ≤ 2,000 ms. Based on these criteria, three participants failed to complete practice blocks in the deictic IRAP and did not proceed to the test blocks. Exclusion criteria also applied to the test blocks, such that participants were required to maintain an accuracy level of ≥ 70% correct and a median latency ≤ 2,200 ms on two of the three successive pairs of the six test blocks. No participants failed to maintain these criteria; hence all data were included in the analysis of the deictic IRAP (N = 37). Any participant who failed the practice blocks of the deictic IRAP did not complete the control IRAP (i.e., three participants). The same criteria were applied to the analysis of the control IRAP, and one participant failed to pass the practice blocks on this basis. Another participant failed to maintain criteria across two of the three successive pairs of six text blocks in the control IRAP. The final number of participants included in the analysis for the control IRAP was 35.

Deictic D_{IRAP}-scores. Consistent with many published IRAP studies, D_{IRAP}-scores were calculated for each of the four trial types (see Barnes-Holmes, Barnes-Holmes, Stewart, & Boles, 2010), such that positive D_{IRAP}-scores during consistent blocks indicated responding "Yes" more quickly than "No" on *I-I* and *Other-Other* trial types and responding "No" more quickly than "Yes" on *I-Other* and *Other-I* trial types. Negative D_{IRAP}-scores indicated the opposite pattern: responding "No" more quickly than "Yes" on *I-I* and *Other-Other* trial types and responding "Yes" more quickly than "No" on *I-Other* and *Other-I* trial types.

The mean D_{IRAP}-scores and standard errors for each trial type are illustrated in Figure 2. The *I-I* and *Other-Other* trial types produced relatively strong IRAP effects, but the *I-Other* and *Other-I* trial types did not. A one-way repeated measures analysis of variance (ANOVA) yielded a significant main effect for trial type, $F(3,36) = 20.84$, $p < .001$, $\eta p2 = .37$. Post-hoc comparisons (Fisher's PLSD tests) indicated that *I-I* ($M = .57$, $SE = 0.06$) differed significantly from the three other trial types: *Other-Other* ($M = 0.34$, $SE = 0.05$, $p < .01$), *I-Other* ($M = .02$, $SE = 0.07$, $p < .001$), and *Other-I* ($M = 0.08$, $SE = 0.05$, $p < .001$). *Other-Other* also differed significantly from *I-Other* ($p < .001$) and *Other-I* ($p < .01$). Four one-sample *t*-tests indicated that only *I-I*, $t(36) = 9.34$, $p < .001$, $d = 1.6$, and *Other-Other,* $t(36) = 6.4$, $p < .00$, $d= 1.1$, differed significantly from zero. In contrast to Barbero-Rubio et al. (2016), only two of our trial types were significantly different from zero, whereas all four of their IRAP effects were[2].

INSERT FIGURE 2 HERE

[2] We subjected the raw reaction time data from this IRAP to the same analyses as those conducted by Barbero-Rubio et al. (2016) and found broadly similar results for the trial types that produced significant D_{IRAP}-scores in our study, but not for the trial types that produced non-significant effects.

***Control D_{IRAP}*-scores.** The mean D_{IRAP}-scores per trial type are illustrated in Figure 3. Relatively strong IRAP effects were recorded on the *Other-Other* and *Researcher-Researcher* trial types, with weak effects on the two remaining trial types. A repeated measures ANOVA yielded a significant main effect for trial type, $F(3,34) = 12.49$, $p < .001$, $\eta p2 = .27$. Post-hoc comparisons indicated that *Other-Other* ($M = .44$, $SE = 0.06$) differed significantly from *Researcher-Other* ($M = .06$, $SE = 0.06$, $p < .001$) and *Other-Researcher* ($M = .04$, $SE = 0.06$, $p < .001$). *Researcher-Researcher* ($M = .32$, $SE = .05$) also differed significantly from *Other-Researcher* ($p < .001$) and *Researcher-Other* ($p < .01$). Four one-sample *t*-tests indicated that only *Other-Other*, $t(34) = 7.55$, $p < .001$, $d = 1.3$, and *Researcher-Researcher*, $t(34) = 6.05$, $p < .001$, $d= 1$, differed significantly from zero.

INSERT FIGURE 3 HERE

Correlations: deictic IRAP and questionnaires. A summary of the means and standard deviations of all questionnaires and questionnaire subscales is provided in Table 4. A correlation matrix was calculated to determine if any of the D_{IRAP}-scores from the deictic IRAP predicted self-reported psychotic experiences (on the CAPE), self-esteem (on the RSES) or perspective-taking (on the PT subscale). The only significant correlations involved the *Other-Other* trial type with the overall CAPE frequency, $r(28) = -.384$, $p = .035$, and the CAPE positive subscale, $r\ (28)= -.475$, $p < .01$. That is, increased response bias in responding to others as others predicted lower levels of psychotic-like experiences.

INSERT TABLE 4 HERE

Discussion

The current data replicated the findings of Barbero-Rubio et al. (2016) to some extent. Specifically, the D_{IRAP}-scores for the *I-I* and *Other-Other* trial types were relatively strong and significant. In contrast to their study, however, the remaining two trial types were both weak and non-significant. Interestingly, a similar pattern was observed with the control IRAP in that two of the IRAP effects were strong and significant, and two were not. A detailed discussion of why the current pattern of trial-type effects obtained for the D_{IRAP}-scores did not closely match the original results will be provided in the General Discussion. Finally, all of the significant correlational analyses between the deictic IRAP and the questionnaires were recorded for the *Other-Other* trial type. At this point, we opted to conduct a second experiment that replicated the first. In this experiment, we added a response requirement to the IRAP that research in our laboratory had shown to moderate the relatively extreme differential trial-type effects observed here.

Experiment 2

While the research reported in the current article was being conducted, an unrelated study in our research group had found that the extreme differential trial-type effect was moderated by the introduction of what we call a *read-aloud procedure.* Specifically, participants are required to read aloud the label, target, and chosen response option at the time of selection on each IRAP trial throughout the entire procedure. Hence, in Experiment 2, all participants completed all IRAPs using a read-aloud procedure. Given the differences between the results of Experiment 1 and those found by Barbero-Rubio et al. (2016), we investigated whether introduction of the read-aloud procedure would produce a pattern of effects similar to the original study?

Method

Participants. A total of 66 individuals participated in Experiment 2, 58 females, seven males, and one individual who did not wish to be categorized as either male or female. Ages ranged from 18 to 48 years old (*M* = 22.98). All participants were recruited through random convenience sampling from the Ghent University participant pool and were paid an hourly rate of 10 euro.

Materials and apparatus. All materials and apparatus for Experiment 2 were identical to Experiment 1.

Procedure. The procedure for Experiment 2 was largely identical to Experiment 1, except that a read-aloud procedure was added to all three IRAPs. This simply required participants to read aloud the label, target, and chosen response option at the time of selection on each trial throughout the entire procedure. Similar to Experiment 1, the experimenter remained in the room throughout the procedure. However, rather than remaining silent, if the participant failed to read aloud, the experimenter reminded them to keep reading aloud during all trials. This was only necessary for a small number of participants, and each required only one reminder across all three IRAPs. The second experiment also differed from the first in that the order in which the IRAP blocks (i.e., consistent followed by inconsistent versus inconsistent followed by consistent) was counterbalanced across participants. Preliminary analyses yielded no significant effects for this procedural variable and thus it is not included in subsequent analyses.

Results

The data and analyses for each IRAP and questionnaire are presented separately. Similar to Experiment 1, given the absence of test blocks in the familiarization IRAP, no data are presented from this procedure.

IRAP data. All aspects of data processing for the IRAPs were similar to those employed in Experiment 1. All participants reached the required performance criteria on the practice blocks of the deictic IRAP. All participants also maintained the performance criteria during the test blocks, hence all data were included in the analyses ($N = 66$). Four participants failed to reach the required performance criteria on the practice blocks of the control IRAP, although all remaining participants maintained the performance criteria during the test blocks. The final number of participants included in the analysis for the control IRAP was 62.

Deictic D_{IRAP}-scores. The mean D_{IRAP}-scores for each trial type are illustrated in Figure 4. Unlike Experiment 1, all trial types produced relatively strong effects, although *I-I* and *Other-Other* were again stronger than the remaining two. A repeated measures ANOVA indicated a significant main effect for trial type, $F(3,65) = 8.98$, $p < .001$, $\eta p2 = .12$. Post-hoc comparisons indicated that I-I ($M = 0.4$, $SE = 0.04$) differed significantly from *I-Other* ($M = .19$, $SE = .05$, $p < .001$) and *Other-I* ($M = 0.21$, $SE = 0.05$, $p < .001$). *Other-Other* also differed significantly from *I-Other* ($p < .001$) and *Other-I* ($p < .002$). Unlike Experiment 1, *I-I* and *Other-Other* did not differ significantly from each other ($p = .5$). Four one-sample *t*-tests indicated that all trial types differed significantly from zero: *I-I*, $t(65)= 9.43$, $p < .001$, $d=1.17$, *I-Other*, $t(65) = 4.15$, $p < .001$, $d= .51$, *Other-I*, $t(65)= 4.5$, $p < .001$, $d= .56$, and *Other-Other*, $t(65) = 9.06$, $p < .001$, $d= 1.12$.

INSERT FIGURE 4 HERE

Control D_{IRAP}-scores. The mean trial-type D_{IRAP}-scores are illustrated in Figure 5. All trial types produced relatively strong effects, with the strongest observed on the *Other-Other* and *Researcher-Researcher* trial types. A repeated measures ANOVA indicated that trial type was significant, $F(3,60) = 6.41$, $p < .001$, $\eta p2 = .01$. Post-hoc comparisons indicated that *Other-Other* ($M = 0.34$, $SE = 0.04$) differed significantly from *Other-Researcher* ($M = .13$, $SE = .05$, $p < .001$) and *Researcher-Other* ($M = 0.13$, $SE = 0.05$, $p < .001$). *Researcher-Researcher* ($M = 0.28$, $SE = 0.04$) differed significantly from *Other-Researcher* ($p < .05$) and *Researcher-Other* ($p < .05$). Four one-sample *t*-tests indicated that all trial types differed significantly from zero: *Other-Other*, $t(60)= 8.44$, $p < .01$, $d=1.09$, *Other-Researcher*, $t(60) = 2.45$, $p < .05$, $d= .32$, *Researcher-Other*, $t(60)= 2.83$, $p < .01$, $d=.37$, and *Researcher-Researcher*, $t(60) = 6.49$, $p < .01$, $d=.84$.

INSERT FIGURE 5 HERE

Deictic IRAP-questionnaire correlations. A summary of the means and standard deviations of all questionnaires and questionnaire subscales is provided in Table 5. A correlation matrix only yielded significant results for the *I-I* trial type. Specifically, increased response biases in responding to I as I predicted: higher overall frequency of psychotic experiences, $r(63) = .316$, $p = .01$, higher levels of overall distress, $r(63) = .267$, $p = .03$, greater frequency in positive symptoms, $r(63) = .27$, $p = .03$, and greater frequency of depressive symptoms, $r(63) = .25$, $p = .04$, as measured by the CAPE.

INSERT TABLE 5 HERE

Discussion

The current data once again showed relatively strong IRAP effects on the *I-I* and *Other-Other* trial types, replicating our findings from Experiment 1 and those from Barbero-Rubio et al. (2016), although we no longer found significant differences between these two trial types. On balance, Experiment 2 now replicated the effects on *I-Other* and *Other-I* reported in the original study (i.e., they were both significantly different from zero). The additional analyses we conducted here, however, indicated that they were both significantly weaker than the *I-I* and *Other-Other* trial types. The control IRAP also yielded significant effects for all four trial types, although two of the trial types (*Researcher-Researcher* and *Other-Other*) continued to be significantly stronger than the two remaining trial types. Similar to the other line of research noted previously, therefore, the read-aloud procedure appeared to attenuate the differential trial-type effect, such that all four trial types (for both IRAPs) were now significantly different from zero. Finally, in contrast to Experiment 1, all of the significant correlations between the deictic IRAP and the questionnaires were recorded for the *I-I* trial type (rather than *Other-Other*).

General Discussion

The initial purpose of the current study was to conduct a systematic replication and extension of Barbero-Rubio et al. (2016). The results in Experiment 1 yielded significant effects for two of the trial types in both the deictic and control IRAPs, whereas four significant effects were observed by Barbero-Rubio et al. In a second experiment, a read-aloud procedure was implemented, and the data showed relatively strong IRAP effects on two trial types, replicating our findings from Experiment 1 and those from the original study; however, we no longer found significant differences between these two trial types.

In comparing our current findings with those reported by Barbero-Rubio et al. (2016), it is interesting to note that we obtained a different pattern of results in Experiment 1. Specifically, they found significant effects for all four trial types, whereas we did not; furthermore, the effect for the *I-I* trial type in our study was significantly different from the effect for the *Other-Other* trial type (these trial types did not differ significantly in the original study). In attempting to explain this difference, it is important to note first that some of the procedures involved in running the IRAPs differed substantively between the studies. For example, in Barbero-Rubio et al., participants received explicit perspective-taking instructions at the beginning of each IRAP block (i.e., "For the next block of trials, you have to respond as if you were you and Adrian were Adrian" and "For the next block of trials, you have to respond as if you were Adrian and Adrian were you"). In addition, participants in the original study were required to complete a deictic relational task (DRT), consisting of 20 scenarios, 12 of which involved reversed deictic relations (e.g., “Mario is swimming in the pool, and Ramon is sailing in a boat. If Ramon were Mario, what would he be doing?”) and eight double reversed deictic relations (e.g., “Luis is in Teide analyzing sediments, and Maria is in Kilimanjaro searching for the source of a river. If Luis were Maria and if the Kilimanjaro were the Teide, where would Luis be?”). In stark contrast, participants in Experiment 1 of the current study were exposed to a basic familiarization IRAP that focused on fruit and vegetables, with no reference to perspective-taking. Furthermore, when participants were exposed to the deictic and control IRAPs in the current study, no specific instructions concerning perspective-taking were provided either at the beginning of the IRAPs or before each block.

Given the foregoing procedural differences between Barbero-Rubio et al. (2016) and the current study, it is difficult to draw firm conclusions concerning the variables that were responsible for the different patterns of results across the two studies. It is worth noting, however, that the type of instructions that are presented to participants before and during IRAP tasks may have quite dramatic effects on performance (see Finn et al., 2016). The exact manner in which instructions have these effects remains to be elucidated (see Finn et al., in press); thus, further speculation at this point would be premature. In any case, it seems important to address the difference observed between Experiments 1 and 2 in the current study.

Experiment 1 produced what may be described as a *single-trial-type-dominance effect* for the *I-I* trial type, but Experiment 2, when the read-aloud procedure was introduced, produced a *dual-trial-type-dominance effect* for the *I-I* and *Other-Other* trial types. These differential trial-type effects have been an important focus of our research activity, both conceptually and empirically, for the past 12 months, and we have begun to develop a model that might help to explain them (see Finn et al., in press for a detailed treatment of the model). The findings reported in the current study, and in particular the different trial-type effects observed across the two experiments, are directly relevant to this model, and thus we will present the model here and articulate how it may help to explain our findings. On the grounds of intellectual honesty, we must be clear that our research strategy has been thoroughly inductive, and the model we outline has arisen partly in a post-hoc fashion from the very data we collected here. In this sense, the data we present were not designed to test the model, but placing the data in the context of the model, we feel, will be particularly instructive in terms of conducting future research.

In attempting to explain the single-trial-type-dominance effect for the *I-I* trial type, we assume that self-related terms possess relatively strong orienting or recognition responses relative to other-related words (Alexopoulos, Muller, Ric, & Marendaz, 2012). We make this assumption based on the fact that, in general, most individuals engage far more frequently in self-related verbal behavior than in verbal behavior related to others. The complete model that aims to explain the single-trial-type-dominance effect is named the Differential Arbitrarily Applicable Relational Responding Effects (DAARRE) model (pronounced as “dare”). The basic model as it applies to the self and other stimuli employed in Experiments 1 and 2 of the current study is presented in Figure 6. The reader is encouraged to consult Figure 6 while reading the following text. Before proceeding, it is worth noting that the single-trial-type-dominance effect was not observed in the control condition in Experiment 1, which supports the assumption that the self-related stimuli possess some functions that other-related stimuli do not.

INSERT FIGURE 6 HERE

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 self-related terms likely possess stronger orienting functions relative to other-related terms, the Cfunc property for self is labeled as positive and the Cfunc property for Other is labeled as negative. The negative labeling for Other should not be taken to indicate a negative orienting function, but simply an orienting function that is weaker than that of self. The labeling of the relations between the label and target stimuli indicates the extent to which they cohere or do not cohere based on the participants' relevant verbal history. Thus, an *I-I* relation is labeled with a plus sign (i.e., coherence), whereas an I-Other relation is labeled with a minus sign (i.e., incoherence). Finally, the two response options are each labeled with a plus or minus sign to indicate their functions as either coherence or incoherence indicators (see Maloney & Barnes-Holmes, 2016). In the current example, "Yes" (+) would typically be used in natural language to indicate coherence, and "No" (-) would be used to indicate incoherence. Note, however, that these and all of the other functions labeled in Figure 6 are behaviorally determined, by the past and current verbal history of the participant, and should not be seen as absolute or inherent in the stimuli themselves.

As can be seen from Figure 6, 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 deictic IRAP. The single-trial-type-dominance effect for the *I-I* trial type may be explained, as noted above, by the DAARRE model based on the extent to which the Cfunc and Crel properties cohere with the relational coherence indicator (RCI) properties of the response options across blocks of trials. To appreciate this explanation, note that the Cfunc and Crel properties for the *I-I* trial type are all labeled with plus signs; in addition, the RCI that is deemed correct for history-consistent trials is also labeled with a plus sign (the only instance of four plus signs in the diagram). In this case, therefore, according to the model this trial type may be considered maximally coherent during history-consistent trials. In contrast, during history-inconsistent trials, there is no coherence between the required RCI (minus sign) and the properties of the Cfuncs and Crel (all plus signs). According to the DAARRE model, this stark contrast in levels of coherence across blocks of trials serves to produce a relatively large IRAP effect. Now consider the *Other-Other* trial type, which requires that participants choose the same RCI as the *I-I* trial type during history-consistent trials, but here the property of the RCI (plus sign) does *not* cohere with the Cfunc properties of the label and target stimuli (both minus signs). During history-inconsistent trials, the RCI (minus sign) *does* cohere with the Cfunc properties but not with the Crel property (plus sign). Thus, the differences in coherence between history-consistent and history-inconsistent trials across these two trial types is not equal (i.e., the difference is greater for the *I-I* trial type) and thus favors the single-trial-type-dominance effect (for *I-I*). Finally, as becomes apparent from inspecting the figure for the remaining two trial types (*I-Other* and *Other-I*), the differences in coherence across history-consistent and history-inconsistent blocks is reduced relative to the *I-I* trial type (two plus signs relative to four), thus again supporting the single-trial-type-dominance effect.

At this point, the DAARRE model appears to explain the single-trial-type-dominance effect for the deictic IRAP. But how might it explain the apparent moderating influence of the read-aloud procedure which appeared to undermine the single-trial-type-dominance effect? Although entirely speculative, it is possible that requiring participants to read aloud each label and target stimulus as they appeared on-screen reduced or eradicated the influence of the orienting functions of those stimuli. In other words, because every label and target was given a similar function by the read-aloud requirement, this overshadowed the differential orienting functions that were present when reading aloud was not required. Thus, the remaining controlling variable was the Crel property, which was the same across the *I-I* and *Other-Other* trial types (both plus signs). As an aside, perhaps the perspective-taking instructions and DRT training provided in Barbero-Rubio et al. (2016) had a functionally similar impact to the read-aloud requirement (i.e., it attenuated the Cfunc properties of the IRAP and thus a dual-trial-type-dominance effect was observed).

In presenting the foregoing model, we recognize that it is specific to the IRAP, but if the IRAP is to be developed as a method for analyzing deictic relational responding, and perspective-taking more generally, it is essential that we understand as fully as possible the functional processes involved in the behavioral patterns we observe with this methodology.

Imagine, for example, that a deictic IRAP was used to explore potential differences in deictic relational responding between groups with different clinical profiles. It may be important to determine if those groups differ in terms of the orienting functions for self and other, rather than any difference in their ability to relate self-to-self and other-to-other.

In considering the potential relevance of the DAARRE model to understanding IRAP performances in the clinical domain, it seems important to reflect upon the pattern of correlations we obtained between the IRAP performances and the self-report measures of psychological suffering in the current study. Specifically, in Experiment 1, only the *Other-Other* trial type correlated with the CAPE. That is, increased response bias in responding to others as others predicted lower levels of psychotic-like experiences. In Experiment 2, however, only the *I-I* trial type correlated with the CAPE. That is, increases in response bias in responding to self as self predicted *higher* levels of psychotic-like experiences. Although wildly speculative, the fact that the pattern of correlations differed in the presence versus the absence of the read-aloud procedure may indicate that manipulating the dominance of the orienting versus relational functions of the IRAP impacts upon specific features of its predictive validity. In the current case, for example, the relatively strong pattern of self-self relational responding, in the absence of orienting functions (i.e., in the presence of the read-aloud procedure), predicted higher levels of psychological suffering. When relative differences in orienting functions were present in the IRAP (i.e., in the absence of the read-aloud procedure), the *Other-Other* trial type predicted lower levels of psychological suffering. In any case, the extent to which different functional properties of the IRAP predict psychological suffering will be an important avenue for future research.

In closing, it seems important to acknowledge a critical limitation in the two experiments reported here. Specifically, the control IRAPs were always presented after the deictic IRAPs. Thus, any difference between the deictic and control IRAPs may be due simply to an order effect. On balance, a simple sequence effect does not account for the differences observed between Experiments 1 and 2 because both involved the same deictic-control IRAP sequence. In any case, the current findings call for greater attention to the conditions under which IRAPs are run, including pre-exposure procedures and the types of instructions that are used, and the impact that these and other variables (such as the read-aloud procedure) may have on the functional properties of the IRAP in exploring specific domains, such as deictic relational responding and perspective-taking more generally.

Compliance with Ethical Standards

Funding: This research was conducted with funding from the FWO Type I Odysseus Programme at Ghent University, Belgium awarded to Dermot Barnes-Holmes.

Conflict of Interest: Deirdre Kavanagh declares that she has no conflict of interest. Yvonne Barnes-Holmes declares that she has no conflict of interest. Dermot Barnes-Holmes declares he has no conflict of interest. Ciara McEnteggart declares that she has no conflict of interest. Martin Finn declares that he has no conflict of interest.

Ethical approval: All procedures in the current study were in accordance with the ethical standards of the institutional research committee, and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards. Informed consent was obtained from all individual participants.

References

Alexopoulos, T., Muller, D., Ric, F., & Marendaz, C. (2012). I, me, mine: Automatic attentional capture by self-related stimuli. *European Journal of Social Psychology, 42*, 770-779.

Barbero-Rubio, A., Lopez-Lopez, J., Luciano, C., & Eisenbeck, N. (2016). Perspective-taking measured by implicit relational assessment procedure (IRAP). *The Psychological Record*, *66*, 243-252.

Barnes-Holmes, Y. (2001). *Analysing relational frames: Studying language and cognition in young children* (Unpublished doctoral thesis). National University of Ireland Maynooth.

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., Finn, M., McEnteggart, C., & Barnes-Holmes, Y. (in press). Derived stimulus relations and their roles in a behavior-analytic account of human language and cognition. *The Behavior Analyst.*

Davis, M. H. (1980). A multidimensional approach to individual differences in empathy. *JSAS Catalog of Selected Documents in Psychology, 10*, 85.

De Corte, K., Buysee, A., Verhofstadt, L., Roeyers, H., Ponnet, K., & Davis, M., (2007). Measuring empathetic tendencies: Reliability and validity of the Dutch version of the Interpersonal Reactivity Index (IRI). *Psychologica Belgica*, *47*, 235-260.

Finn, M., Barnes-Holmes, & McEnteggart, C. (in press). Exploring the single-trial-type-dominance-effect in the IRAP: Developing a differential arbitrarily applicable relational responding effects (DAARRE) model. *The Psychological Record*.

Franck, E., De Raedt, R., Barbez, C., & Rosseel, Y. (2008). Psychometric properties of the Dutch Rosenberg Self-esteem Scale. *Psychologica Belgica, 48,* 25-35

Gore, N. G., Barnes-Holmes, Y., & Murphy, G. (2010). The relationship between intellectual functioning and relational perspective-taking. *International Journal of Psychology and Psychological Therapy*, *10*, 1-17.

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

Heagle, A.I., & Rehfeldt, R. A. (2006). Teaching perspective-taking skills to typically developing children through derived relational responding. *Journal of Early Intensive Behavioral Intervention*, *3*, 1-34.

Hussey, I., McEnteggart, C., Barnes-Holmes, Y., Kavanagh, D., Barnes-Holmes, D., Parling, T., & Lundgren, T. *Flexible perspective-taking: New concepts and a new behavioural measure.* ACT CBS Conference. Dublin (December, 2014).

Konings, M., Bak, M., Hanssen, M., Van Os, J., & Krabbendam, L. (2006). Validity and reliability of the CAPE: A self-report instrument for the measurement of psychotic experiences in the general population. *Acta Psychiatrica Scandinavica*, *114*, 55–61.

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*, 395-403.

McHugh, L., Barnes-Holmes, Y., & Barnes-Holmes, D. (2004). Perspective-taking as relational responding: A developmental profile. *The Psychological Record*, *54*, 115-144.

Robins, R. W., Hendin, H. M., & Trzesniewski, K. H. (2001). Measuring global self-esteem: Construct validation of a single-item measure and the Rosenberg self-esteem scale. *Personality and Social Psychology Bulletin*, 27, 151-161

Sidman, M. (1960). *Tactics of scientific research: Evaluating experimental in psychology*. New York, NY: Basic Books

Stefanis, N. C., Hanssen, M., Smirnis, N. K., Avramopoulos, D. A., Evdokimidis, I. K., Stefanis, C. N., Verdoux, H., & Van Os, J. (2002). Evidence that three dimensions of psychosis have a distribution in the general population. *Psychological Medicine, 32*, 347-358.

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

Villatte, M., Monestés, J.-L., McHugh, L., Freixa i Baqué, E. F. i, & Loas, G. (2008). Assessing deictic relational responding in social anhedonia: A functional approach to the development of theory of mind impairments. *Journal of Behavioral Consultation and Therapy*, *4*, 360–373.

Villatte, M., Monestés, J.-L., McHugh, L., Freixa i Baqué, E., & Loas, G. (2010). Assessing perspective taking in schizophrenia using relational frame theory. *The Psychological Record*, *60*, 413–436.

Vitale, A., Barnes-Holmes, Y., Barnes-Holmes, D., & Campbell, C. (2008). Facilitating responding in accordance with the relational frame of comparison: Systematic empirical analyses. *The Psychological Record*, *58*, 365–390.

Weil, T. M., Hayes, S. C., & Capurro, P. (2011). Establishing a deictic relational repertoire in young children. *The Psychological Record*, *61*, 371-390.

Table 1

Labels, Targets, and Response Options in the Familiarization IRAP

Labels	
Fruit (Fruit)	Groenten (Vegetables)
Targets	
Appel (apple)	Wortel (carrot)
Banaan (banana)	Ardappel (potato)
Sinaasappel (orange)	Broccoli (broccoli)
Peer (peer)	Spruit (sprout)
Reponses	
Ja (Yes)	Nee (No)

Note. English translation in brackets.

Table 2

Labels, Targets, and Response Options in the Deictic IRAP

Labels	
David (Participant's name)	Dee (Researcher's name)
Targets	
zit hier (seated)	Staat recht (standing up)
is de vrijwillger (is the participant)	is de onderzoeker (researcher)
zit an het toetsenbord (with keyboard)	heeft een pen (holding a pen)
Kijkt naar het scherm (looking at screen)	heeft een notebook (holding a notebook)
is hier (here)	is daar (there)
heeft een gele Post-it (yellow Post-it)	heeft een oranje Post-it (orange Post-it)
Reponses	
Ja (Yes)	Nee (No)

Note. English translation in brackets.

Table 3

Labels, Targets, and Response Options in the Control IRAP

LABELS	
Ciara (Picture)	Dee (Researcher)
TARGETS	
zit hier (seated)	Staat recht (standing up)
is de vrijwillger	is de onderzoeker (researcher)
zit an het toetsenbord (with keyboard)	heeft een pen (holding a pen)
Kijkt naar het scherm (looking at screen)	heeft een notebook (holding a notebook)
heeft blond haar (has blond hair)	heeft bruin haar (has brown hair)
heeft een gele Post-it (yellow Post-it)	heeft een oranje Post-it (orange Post-it)
RESPONSES	
Ja (Yes)	Nee (No)

Note. English translation in brackets.

Table 4

Experiment 1 Descriptive Statistics for Self-reports

	M	*SD*
CAPE (weighted scores)		
Overall Frequency	1.67	.24
Frequency of Positive Symptoms	1.38	.25
Frequency of Negative Symptoms	1.91	.46
Frequency of Depressive Symptoms	1.99	.41
Overall Distress	1.01	.32
Distress associated with Positive Symptoms	1.79	.50
Distress associated with Negative Symptoms	3.39	2.17
Distress associated with Depressive Symptoms	2.37	.63
Rosenberg Self-Esteem Scale		
Overall Score	22.27	8.12
Interpersonal Reactivity Index (IRI)		
PT-subscale	18.13	3.41

Note: The maximum weighted score for all CAPE subscales is 4.00. The maximum score is 30 for the Rosenberg Self-Esteem Scale and 35 for the PT-subscale of the IRI. None of the scales have formal clinical cut-offs.

Table 5

Experiment 2 Descriptive Statistics for Self-reports

	M	*SD*
CAPE (weighted scores)		
Overall Frequency	1.69	.35
Frequency of Positive Symptoms	1.40	.31
Frequency of Negative Symptoms	1.75	.48
Frequency of Depressive Symptoms	2.22	.62
Overall Distress	2.05	.44
Distress associated with Positive Symptoms	1.85	1.01
Distress associated with Negative Symptoms	2.10	.51
Distress associated with Depressive Symptoms	4.87	4.15
Rosenberg Self-Esteem Scale		
Overall Score	18.06	5.61
Interpersonal Reactivity Index (IRI)		
PT-subscale	18.12	3.52

Note: See note to Table 4 for maximum scores.

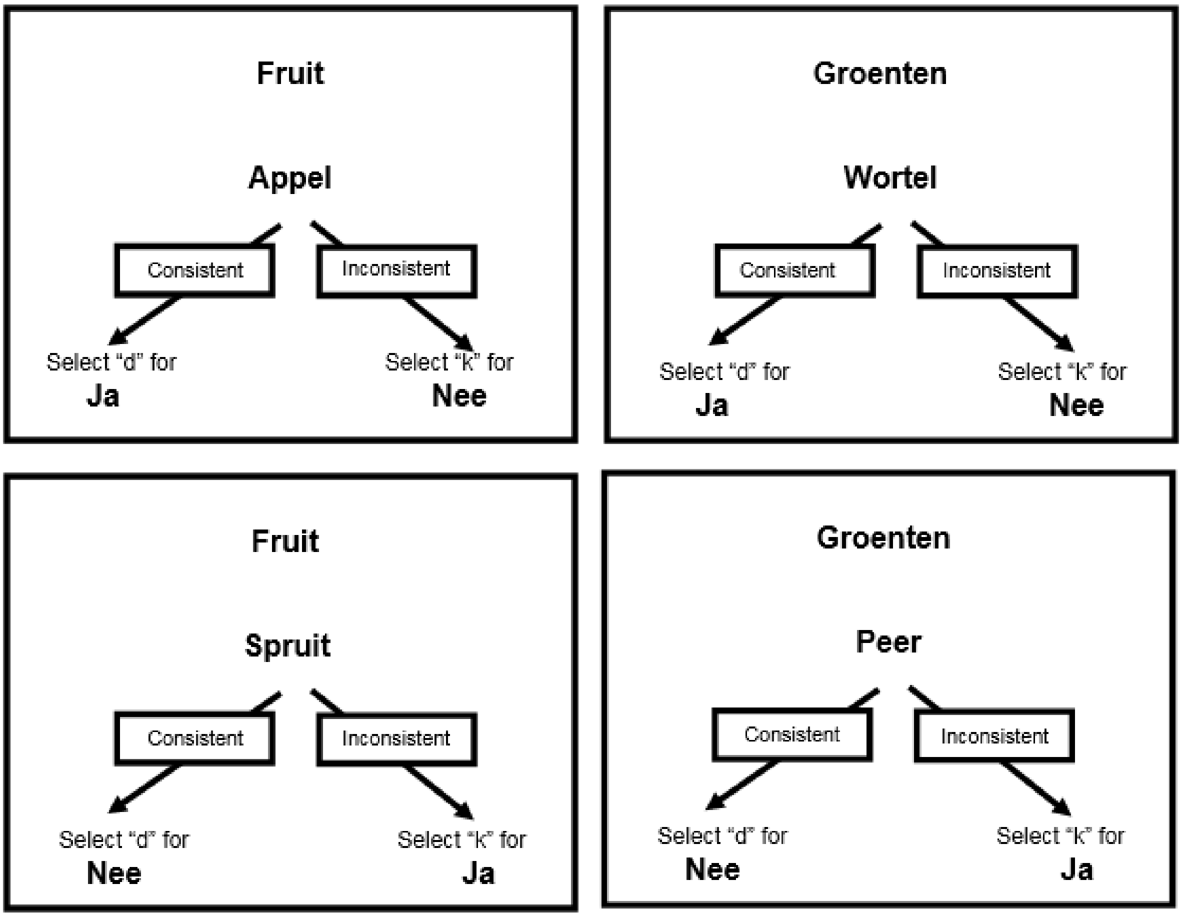

Figure 1. Examples of the four trial types in the familiarization IRAP: *Fruit-Fruit, Vegetable-Vegetable, Fruit-Vegetable,* and *Vegetable-Fruit.* The words *Consistent* and *Inconsistent* were not shown on-screen.

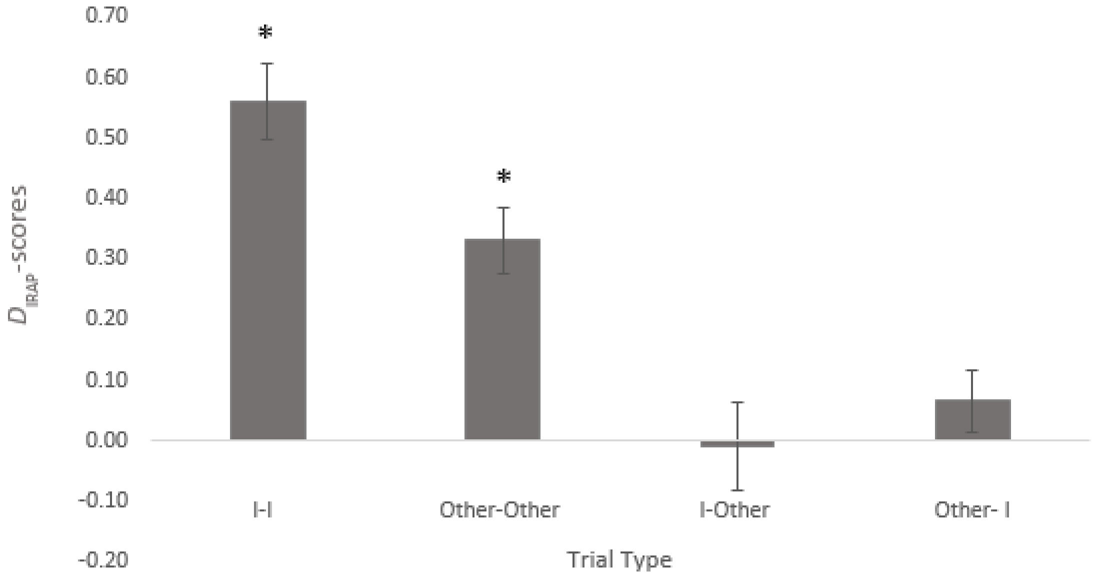

Figure 2. Mean D_{IRAP}-scores on the deictic IRAP trial types in Experiment 1. Positive D_{IRAP}-scores indicate history-consistent responding and negative D_{IRAP}-scores indicate history inconsistent responding. * Indicates D_{IRAP}-scores that are significantly different from zero.

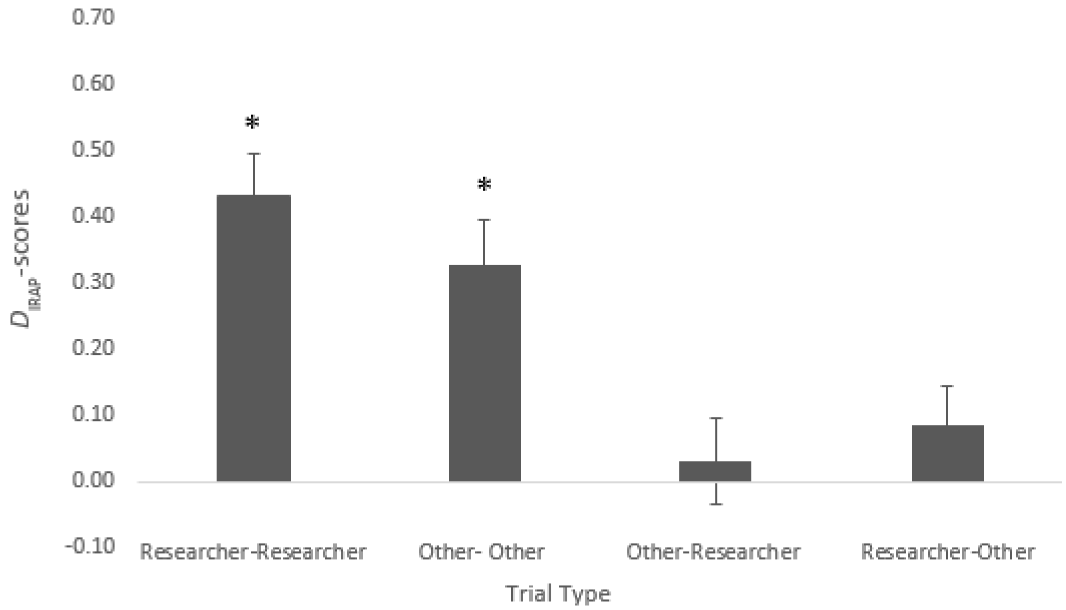

Figure 3. Mean D_{IRAP}-scores on the control IRAP trial types in Experiment 1. Positive D_{IRAP}-scores indicate history-consistent responding and negative D_{IRAP}-scores indicate history-inconsistent responding. * Indicates D_{IRAP}-scores that are significantly different from zero.

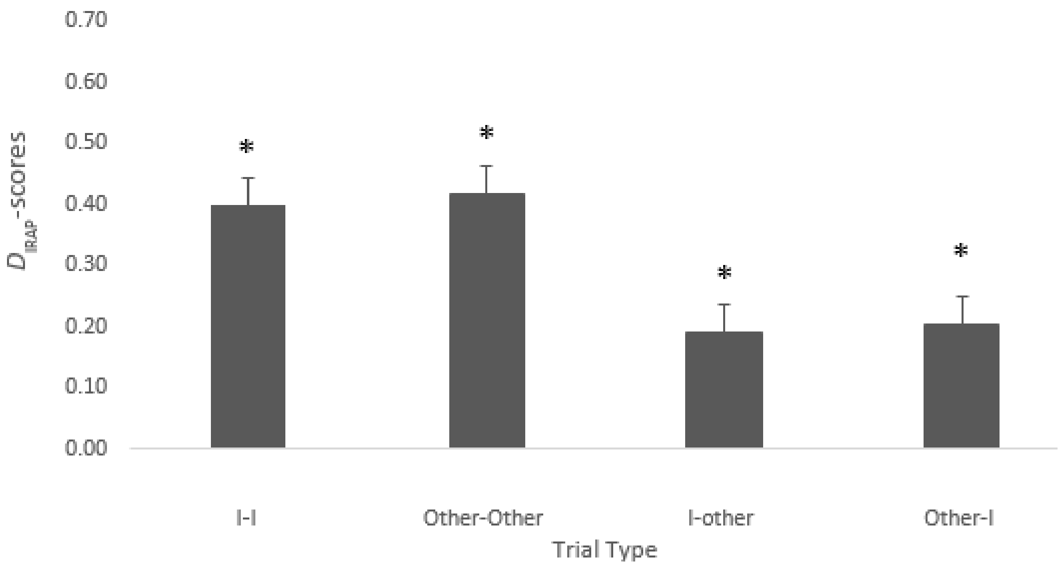

Figure 4. Mean D_{IRAP}-scores on the deictic IRAP trial types in Experiment 2. Positive D_{IRAP}-scores indicate history-consistent responding and negative D_{IRAP}-scores indicate history-inconsistent responding. * Indicates D_{IRAP}-scores that are significantly different from zero.

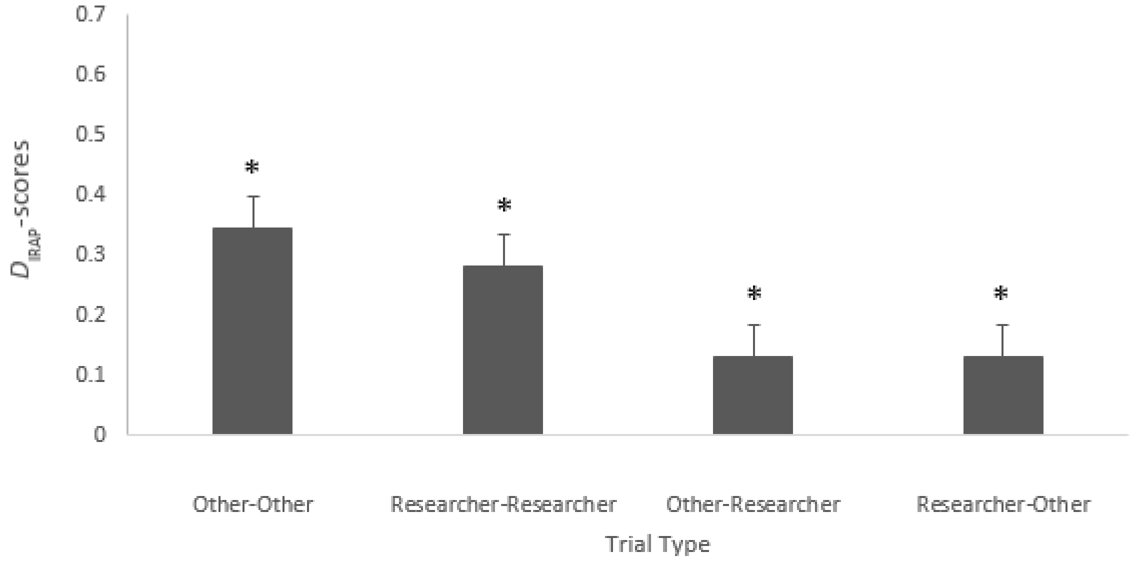

Figure 5. Mean D_{IRAP}-scores on the Control IRAP trial types in Experiment 2. Positive D_{IRAP}-scores indicate history-consistent responding and negative D_{IRAP}-scores indicate history-inconsistent responding. * Indicates D_{IRAP}-scores that are significantly different from zero.

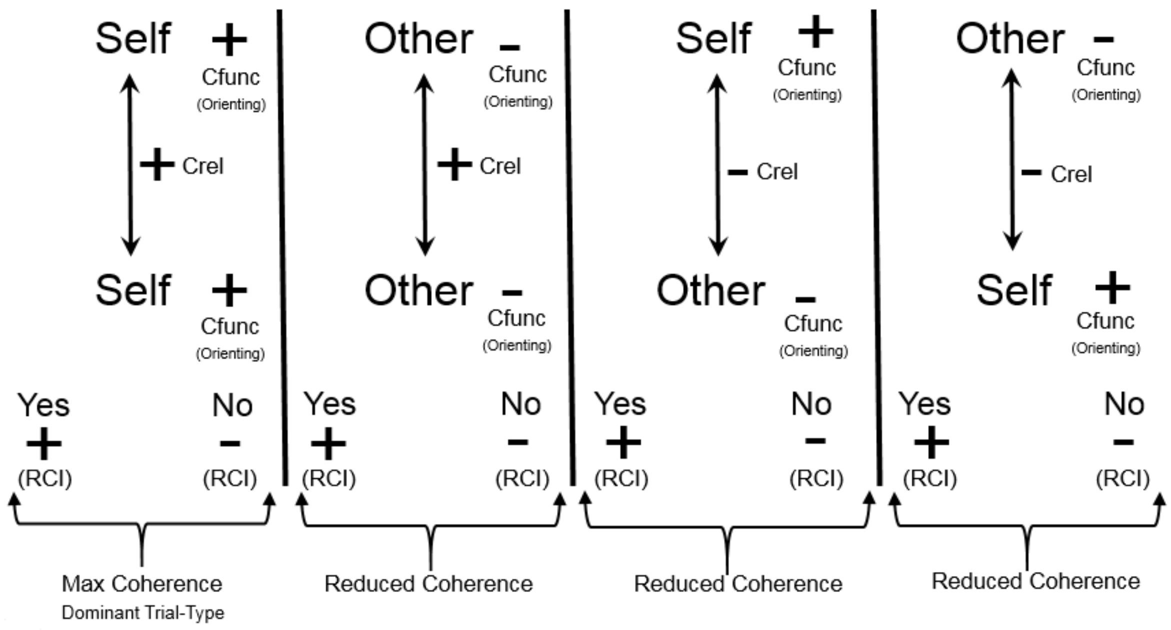

Figure 6. The DAARRE model as it applies to the deictic IRAP. The positive and negative labels refer to the relative positivity of the Cfuncs, for each label and target, the relative positivity of the Crels, and the relative positivity of the RCIs in the context of the other Cfuncs, Crels, and RCIs in that stimulus set.