View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Edge Hill University Research Information Repository

Psychonomic Bulletin & Review

Does interference between self and other perspectives in Theory of Mind Tasks reflect a common underlying process? Evidence from individual differences in theory of mind and inhibitory control.

Journal:	*Psychonomic Bulletin & Review*
Manuscript ID	PBR-BR-19-078.R1
Manuscript Type:	Brief Report
Date Submitted by the Author:	n/a
Complete List of Authors:	Qureshi, Adam; Edge Hill University Monk, Rebecca; Edge Hill College Samson, Dana; Universite catholique de Louvain Apperly, Ian ; University of Birmingham
Keywords:	Theory of mind, Executive Function, Self/other interference, Path analysis

SCHOLARONE™
Manuscripts

Dear Editors,

Please accept this revised paper '*Does interference between self and other perspectives in Theory of Mind Tasks reflect a common underlying process? Evidence from individual differences in theory of mind and inhibitory control*', Manuscript ID PBR-BR-19-078, for consideration at publication at Psychonomic Bulletin & Review, submitted as an original article.

Despite the vast literature on theory of mind (the ability to understand other agents have different beliefs, desires and knowledge to oneself), the nature and origin of interference effects seen between the 'self' and 'other' perspectives in ToM tasks have not been examined in terms of individual differences. The current study addresses two fundamental questions within the theory of mind literature: are there systematic individual differences in self-other interference, and are these differences due to individual differences in executive function? The study comprised a large-scale path analysis model examining the relationship between performance on inhibitory control and two theory of mind tasks that involve dealing with interference between self and other perspectives. Results suggest that self/other interference effects in the two ToM tasks were dissociable, and related to different tests of inhibitory control. We conclude that while self/other interference is a key part of ToM tasks, it may not be the only limiting step for healthy adults' performance on typical tasks.

Adam W. Qureshi[a*], Rebecca L. Monk[a], Dana Samson[b] & Ian A. Apperly[c]

[a]Edge Hill University, Ormskirk, UK, L39 4PY

[b]Université Catholique de Louvain, Louvain-la-Neuve, Belgium, L0.01.03 1348

[c]University of Birmingham, Edgbaston, UK, B15 2TT

For Review Only

Each author has contributed significantly to this work and agrees to the submission.

Correspondence concerning this article should be addressed to: Adam Qureshi, Department of Psychology, Edge Hill University, Ormskirk, L39 4QP, UK. E-mail: qureshia@edgehill.ac.uk Tel +44 (0) 1695 584 498; Fax: 44(0)1695 579 997

This paper contains original material and is not submitted for consideration, in press or published anywhere else and the authors declare no conflict of interest. If accepted, it will not be published elsewhere including electronically in the same form, in English or in any other language, without the written consent of the copyright-holder.

We state that none of the authors or their immediate family have a significant financial arrangement or affiliation with any product or services used or discussed in my paper, nor any potential bias against another product or service.

Kind regards.

A.W.Qureshi Ian A. Apperly Dana Samson Rebecca L.Monk

We hope we have addressed the concerns of both yourself and the reviewers with our revised submission. We detail our responses to all comments below, and thank you for the detailed, informative and useful points and suggestions made.

Editor comments

“your 'theory of mind' tasks may not actually require representing others’ mental states. If this is the case, your data may not allow us to draw conclusions about ToM in general, but just about these two tasks; and an alternative interpretation of your data would be that these two tasks are themselves measuring domain-general inhibition and not ToM. I'd like to see a robust rebuttal of this interpretation.”

***We have added a substantial new section to the discussion to address this concern.

Please expand on the implications of the finding that the general inhibition tasks don’t correlate particularly highly with each other; The reliability was poor for ‘ambiguous errors’ in the director task – please discuss this in your comments on reliability in the final sections of the manuscript; Please include a statement regarding ethical approval.

***Thanks for highlighting these. We have added a line re. low reliabilities to the Limitations in ToM measures section: **we note that while task reliabilities were generally high (< .7) the reliability for ambiguous errors in the Director task was .59.** The low correlations between the inhibitory control tasks are common in the literature (e.g. Friedman & Miyake, 2004), potentially due to task impurity issues. We have added a line to reflect this to the Data screening section. We have also added a line to the participants section regarding ethical approval: **Ethical approval was gained from the departmental ethics committee.**

Reviewer Comments to Author (if any):
Reviewer: 1

Comments to the Author
This paper involves two assessments of Theory of Mind abilities – visual perspective taking Level 1 and a referential communication task, the ‘Director’ task – along with four different measures of inhibitory control, aiming to establish (1) whether self/other interference effects are seen to the same extent across different tasks, (2) how inhibition predicts ToM task performance, and (3) whether the inhibition and ToM relationship varies across tasks. The study utilized a good sample size of 142 participants, and used path analysis to examine relationships between outcomes. Results showed a dissociable self/other interference effect across the two ToM measures; further, path analysis showed that each of the ToM task interference effects were predicted by different inhibitory control tasks. The authors conclude that self/other interference is not a unitary construct, varying across different task demands.

This paper is well-written and thorough, and was enjoyable to read, addressing a key research area in ToM literature (individual differences in ToM and executive control). The study involved a large subject number, and used a number of different tasks to assess both ToM and EF, allowing subtle differences between tasks and relationships across tasks to be assessed. I

For Review Only

have some minor clarification points as detailed below, but overall was impressed with the study.

Introduction:

Page 4, line 50-53: The authors review here prior literature that has suggested training participants to inhibit automatic imitation led to improved performance on a test of ToM, whereas generic inhibition training did not. This could do with a bit more explanation to make the prior findings clearer, including why decreased imitation would lead to ToM improvement and how this relates to the current study.

***We have added lines (in bold) to clarify why this effect may have occurred: "Santiesteban, White, Cook, Gilbert, Heyes and Bird (2012) found that training participants to inhibit automatic imitation improved performance on a test that required participants to use their ToM (the Director Task; Keysar, Lin & Barr, 2003), whereas training generic inhibition did not, suggesting that inhibition of automatic imitation and **ToM rely on control over representations related to self and other, and these control processes are not the same as those involved in conventional "inhibition" tasks.".**

With regard to the relationship to the current study, we go on to state that such findings (along with neuroscience findings, e.g. Bardi, Six & Brass, 2017; Wagner, Maril, Bjork & Schacter, 2001) suggest that self-other control is a more domain-specific process as "**generic inhibition training did not improve ToM performance; though see a recent meta-analysis by Darda and Ramsey (2019))**", and that this leads to our "prediction that there will be individual differencves in self-other control across different ToM tasks, and that these should not be related to individual differences in executive function.".

Page 4, line 29-33: this is a personal preference, but I would avoid starting a sentence with the word 'and' – this makes the paragraph a bit jumpy.

***So edited – thanks for spotting this.

Appendix A (Individual Task Results):

Shape-Matching Task (and others): the authors state that 'Response time values that were more than three standard deviations away from the condition mean for a participant were replaced by the cut-off value (of three standard deviations from the condition mean)'. Could the authors please comment on why this approach was taken? It is not clear why these values were not removed as outliers, rather than replaced. Additionally, please include information about how many data points were required to be changed/how many exceeded this cut off.

***We didn't remove the outliers as we followed the recommended data screening procedure of Friedman and Miyake (2004) and Oberauer (2005) in order to retain as many data points as possible, and also to reduce the effect of any extreme response times on the mean.

For those measures that used error rates (Go / No-Go (picture) and Go / No-Go tasks), no outliers were removed or replaced. For the Shape-matching task 1.41% of trials were replaced. For the Stop-Signal task, no initial data points were removed or replaced, but two participant SSRT values were replaced as per Tabachnick and Fidell (2001) due to an

For Review Only

extremely non-normal distribution. For the L1-VPT task, only matching trials were analysed. Of these 1.41% were incorrect responses, 2.04% had response times more than 2.5 SDs from the mean, and 0.47% were response omissions, totalling 3.92%. The Director Task replaced 0.48% of trials.

Lines covering the above have been added to section A3 of the new Appendix A (executive task details).

A minor point, but why for inhibitory control tasks were responses removed if over 3 standard deviations from the mean, but for the L1 VPT task 2.5 standard deviations was used as criteria?

***We followed the procedures most commonly employed in the published literature for the L1 VPT task (Qureshi et al., 2010) and for the various inhibitory control tasks, and have added a footnote to reflect this.

For Review Only

Discussion:

Page 21 – Lines 18-36: This paragraph needs to be more fully explained/contain further information to make argument clearer.

***We have edited this paragraph to improve the clarity of the argument, we hope that this is now clearer.

Reviewer: 2

Comments to the Author
The paper investigates individual differences in Theory of Mind performance and the relationship between self-other interference and inhibitory control. The research question is relevant and the manuscript has certainly the potential to provide an important contribution to the literature of the field. However, I list below some issues/questions that should be addressed to improve the quality of the manuscript.

1. A number of "executive control" tasks have been included in the study (some of which are not reported in the manuscript). I assume the authors meant to measure different aspects of inhibitory control but I have the impression that the theoretical motivation behind task selection is missing.

   It is also not clear what were the specific predictions with respect to correlations with individual tasks. In effect, results show that the two perspective taking tasks correlate with different measures of executive control (i.e., different tasks). The interpretation of this outcome seems to be challenging, but I think the authors should try to elaborate a bit more on that.

***We have added a short section on executive task selection on p9 of the main manuscript and in Appendix A included a more extensive discussion of the rationale for task selection, analysis, and predictions for relationships among tasks.

2. The sample size has been selected without an appropriate power analysis. A power analysis should be performed to determine the optimal sample for the study.

***This was an omission on our part, thank you for pointing it out. At the time of data collection, the rule of thumb sample size (Kline, 2005) was used which suggests a 10:1 ratio for free parameters as a realistic figure. This gives sample sizes of 180 for the initial model, 60 for the second model and 80 for the final model. He also suggests that a sample size of between 100 – 200 can be considered a medium sample size.

Preacher and Coffman (2006) provide a sample size and power analysis calculator based on RMSEA values. Taking α as .05, df = 20 (final model), and n = 142, and setting the null RMSE value as .08 (cut-off for a poor fitting model; Kenny, 2015) and the alternative RMSEA value as .00 (final model), this gives us a power of .67 to detect whether this is a poor fitting model.

The same calculator suggests with null RMSEA = .01 (excellent fit), alternative RMSEA as .08 (mediocre fit) and an observed power of .8, the sample size should be 169. Therefore we believe that considering the experimental nature of the study, the sample size is sufficient for the model, which is supported by the bootstrapping estimates. Indeed, changing the mediocre fit cut-off RMSEA value to .10 (as per Kline, 2005) changes the observed power to .92, and the recommended sample size to 107.

We have added the following after Table 4:

A power analysis was conducted on the final model fit values and sample size using Preacher and Coffman's (2006) power and sample size calculator for RMSEA. This suggests that the current sample size has a power of .92 to detect whether this was a poor fitting model[1]. Testing whether the sample size was sufficient for the final model showed that an ideal sample size should be 107 (with .10 as the RMSEA value for a poor fit; Kenny, 2014[2]). Kline (2005) suggests a 10:1 ratio for free parameters as a realistic figure, which gives an estimated sample size of 80 for the final model. The current sample size is appropriate for the current analyses.

Preacher, K. J., & Coffman, D. L. (2006, May). Computing power and minimum sample size for RMSEA [Computer software]. Available from http://quantpsy.org/.

Kenny, D. A. (2014, January 12). Structural Equation Modelling. Retrieved June 11, 2019, from http://davidakenny.net/cm/causalm.htm

3. It is not clear to me why the two perspective-taking tasks where chosen over other measures of self-other control. Moreover, aren't those two tasks typically considered to measure perspective taking, and not ToM tasks?

***We thank the reviewer for encouraging us to make our reasoning apparent here, but respectfully disagree with the suggestion that visual perspective-taking is distinct from ToM.

[1] Observed power = .8, df = 20, n =142, null RMSEA = .01 (excellent fit), alternative RMSEA = .00 (final model value).
[2] Observed power = .8, df = 20, null RMSEA = .01, alternative RMSEA = .10

For Review Only

In support of our position, we note that visual perspective-taking is widely seen as a component of the development of ToM in children, and features in much contemporary discussion of the cognitive basis of ToM in adults. Our choice of these tasks was further influenced by the fact that they are relatively unusual in having been analysed into component processes to which inhibitory control may make distinctive contributions. This contrasts with the majority of ToM "measures" which are often quite opaque with respect to their underlying functional processes. On p10 we now include a paragraph explaining this rationale.

4. Two different versions of a go/no-go task were included. How the distinction between letter and image-based task would be relevant to the goal of the study?

***The Go / No-Go (picture) based task required an additional step of processing (the identification of the image as being a bird or mammal), followed by the response inhibition/selection required to choose whether to respond or not, based on the phonological or orthographical information (did the name begin with a vowel or consonant). This represented another variation of task that requires inhibition, which also required an additional step of response selection, resulting in more individual variability in inhibition being captured. We have detailed this in section A2 of Appendix A (executive task details).

5. The interference effect for the shape-matching task was calculated by using only two of the four task conditions: the no-match distractor and the match no-distractor condition. This choice seems arbitrary to me (and I find the explanation provided in the appendix not completely convincing). I think the authors should rather consider the difference between the two distractor present conditions and the two distractor absent conditions (distractor present match + distractor present no-match)/2 – (distractor absent match + distractor absent no-match)/2.

***Again, thank you for this suggestion. We agree, and have reanalysed the data accordingly with the suggested difference scores. The resulting model shows the same relationships with the ToM tasks, but no correlation between the shape-matching and stop-signal tasks.

6. About the shape-matching task, it was reported in the appendix: "there was an interaction between distraction and matching as can be noted in Figure A3". However, perhaps because the scale of the figure, it is not clear what the interaction entails. Please adapt the graph accordingly.

***We have adjusted the y-axis on 480ms rather than 0ms, so the interaction caused by the longer response times to Target MisMatch compared to TargetMatch when the distracter was present is more visible.

7. The shape-matching task was an adapted version of a task developed by DeSchepper and Trieseman, 1996. How was the task modified with respect to the original version and why?

***We reduced the number of trials to 112 across 4 blocks rather than 200. The task was also simplified to measure resistance to distracters and resistance to inappropriate responses,

For Review Only

rather than negative priming as in DeSchepper and Treisman (1996). This was done by presenting the distracter shapes at the same time as target shapes in the experimental trials, requiring participants to actively keep the distracters from interfering with processing of the target shapes (Endo, Saiki & Sato, 2001).

We have added the following to section A2 of Appendix A (executive task details):

The task was simplified to measure resistance to distracters and resistance to inappropriate responses, rather than negative priming as in DeSchepper and Treisman (1996). This was done by presenting the distracter shapes at the same time as target shapes in the experimental trials, requiring participants to actively keep the distracters from interfering with processing of the target shapes (Endo, Saiki & Sato, 2001).

Endo, N., Saiki, J. & Saito, H. (2001). Determinants of occurrence of negative priming for novel shapes with matching paradigm. *Japanese Journal of Psychology*, 72 (3), 204 – 212

8. The method section lacks important details about the visual stimuli (stimulus size, duration…), responses recording, number of trials….etc…

***We have added a section for this (A2) in Appendix A (executive task details).

9. The path analysis, as well as the results, need to be described in more details in the main text, especially for the reader who is not familiar with this kind of approach, and not only in the appendix.

***Thank you for this advice. We have added a brief summary of the aim of path analysis to the start of the Models sub-section (page 14; **“**Path analysis aims to arrive at the most parsimonious model that explains the underlying data and does not significantly differ from it. Increasing the number of parameters in a model tends to improve fit, but necessarily decreases parsimony. The best model optimises fit and parsimony.**”** We have also given more detail on each model in the main results (while keeping the full tables/figures in Appendix B for brevity), including information on model comparisons (pages 14 – 15).

10. I am wondering whether having a fixed task order was the right choice. As tasks are supposed to measure similar constructs, it is plausible that performance in one task has an influence on the performance of subsequent tasks. Counterbalancing the task order should have controlled this.

***This is a very good point, and one we did not fully explain. We decided on a fixed order because this is widely-acknowledged to be the most powerful design in individual differences research (e.g. Carlson & Moses, 2001; Carlson et al., 2004; Friedman & Miyake, 2004; Miyake et al., 2000; Miyake, Friedman, Rettinger, Shah & Hegarty, 2001). While the patterns of correlation between tasks should be the same for any task order, variation in order within a sample introduces irrelevant variance that can obscure correlations. To illustrate, consider tasks A and B that are highly correlated, such that the true rank order of participants' performance is the same on A and B. Now suppose A and B are presented in a battery on

For Review Only

which participant tiredness naturally leads to decreasing performance over time. In a fixed order design in which task A always occurs before B, performance on B will be underestimated compared with A, but the true rank order of participants on each task will be preserved. In contrast, a design that varies task order makes it possible that a participants' rank on any one task may shift above or below that of other participants on that task just because they happened to complete the tasks in different orders (and so varied in levels of tiredness). This reduces the power for detecting the true association that exists between the tasks.

Fixing the task order avoids this problem, but we note that a given task order may result such a degree of tiredness that task performance can no longer be assessed reliably. To check whether the performance on tasks influenced one another, perhaps due to there being a finite resource for executive components (Muraven & Baumeister, 2000), participants completed the Go / No-Go task at the start and end of each session, and results compared. These showed no difference, suggesting that fatigue and/or influence of similar tasks on performance did not occur.

We have added the following to the Task order section: **This was done because a fixed task order is optimal for detecting correlations** (e.g. Carlson & Moses, 2001; Carlson et al., 2004; Friedman & Miyake, 2004; Miyake et al., 2000; Miyake, Friedman, Rettinger, Shah & Hegarty, 2001). **To check whether the performance on tasks influenced one another, perhaps due to there being a finite resource for executive components (Muraven & Baumeister, 2000), participants completed the Go / No-Go task at the start and end of each session, and results compared. These showed no difference, suggesting that fatigue and/or influence of similar tasks on performance did not occur.**

11. The Director task is a typical explicit task, because subjects are always instructed to take the perspective of the avatar. The L1 VPT task is, on the other hand, at least partially implicit (the effect of the avatar's perspective is automatic in "self" trials). A possible dissociation between implicit and explicit ToM is a matter of intense debate. I wonder whether one should not consider the interference effect in "self" and "other" trials separately. Possible implicit/explicit dissociation might come to help in the interpretation of the path analysis outcome.

***Thank you for this suggestion. We certainly recognise this interesting issue. Our analysis strategy derives from the challenges in deriving maximally informative dependent variables from tasks that obviously involve multiple processes. To illustrate, the magnitude of the altercentric interference effect in a Self trial is likely to depend on BOTH the propensity for a participant to process the avatar's perspective (or the weight or salience they accord that perspective), AND the participants' capacity to give a controlled response in the face of interference from the avatar's perspective. With appropriate changes made, the same analysis applies to egocentric interference on Other trials. [Note also, an analogous analysis would apply to alternative interpretations of this effect in terms of spatial cueing.] Following Bukowski & Samson, our strategy is to orthogonalise these factors, to derive indices of focus (weighting of self vs other) and conflict (ability to resolve self-other interference). We think this is maximally informative, but we agree that, since much of the published literature has looked separately at interference on Self and Other trials, it is also informative to look at these trials separately. Therefore in Appendix C we now present models using interference measures (from self perspective; egocentric interference, and other perspective; altercentric

For Review Only

interference). Results show that interference from the self perspective is related to the SST, but that interference from the other perspective is not related to any of the executive tasks. We have made a note to this effect in the result section.

12. Page 5, first line: "….implicated in analogous non-social executive control(". Please also see:
Brass, M., Derrfuss, J., & von Cramon, D. Y. (2005). The inhibition of imitative and overlearned responses: a functional double dissociation. Neuropsychologia, 43, 89–98.
Brass, M., Ruby, P., & Spengler, S. (2009). Inhibition of imitative behaviour and social cognition. Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 364(1528), 2359–67.
Darda & Ramsey. (2019). The inhibition of automatic imitation: a meta-analysis and synthesis of previous studies. Biorxiv.
Moreover, this work seems to be relevant to the discussion and should be at least mentioned: Conway, Catmur and Bird (2019). Understanding individual differences in theory of mind via representation of minds, not mental states. Psychon. Bull. Rev. In press.

***Thank you for bringing this literature to our attention. We have added lines and references to the section on non-social executive control (page 5), and also a line referencing the 'Mind-space' framework as a way to move forward with individual differences research in theory of mind to the discussion.

For Review Only

Does interference between self and other perspectives in Theory of Mind Tasks reflect a common underlying process? Evidence from individual differences in theory of mind and inhibitory control.

For Review Only

Adam W. Qureshi[a*], Rebecca L. Monk[a], Dana Samson[b] & Ian A. Apperly[c]

[a]Edge Hill University, Ormskirk, UK, L39 4PY

[b]Université Catholique de Louvain, Louvain-la-Neuve, Belgium, L0.01.03 1348

[c]University of Birmingham, Edgbaston, UK, B15 2TT

*Corresponding author: qureshia@edgehill.ac.uk

Declarations of interest: none

Abstract

Theory of mind (ToM), the ability to understand other agents have different beliefs, desires and knowledge to oneself, has been extensively researched. Theory of mind tasks involve participants dealing with interference between their self-perspective and another agents perspective, and this interference has been related to executive function, particularly inhibitory control. The current study assessed whether there are individual differences in self/other interference, and whether these effects are due to individual differences in executive function. 142 participants completed two theory of mind tasks (the Director Task and a level one visual perspective-taking task), which both involve self-other interference, and a battery of inhibitory control tasks. The relationships between the tasks were examined using path analysis. Results showed that the self/other interference effects of the two ToM tasks were dissociable, with individual differences in performance on the ToM tasks being unrelated and performance in each predicted by different inhibitory control tasks. We suggest that self/other differences are part of the nature of theory of mind tasks, but self/other interference is not a unitary construct. Instead, self/other differences result in interference effects in a variety of ways and at different stages of processing, and these effects may not be a major limiting step for adults' performance on typical ToM tasks. Further work is needed to assess other factors that may limit adults' ToM performance and hence explain individual differences in social ability.

For Review Only

Key words: Theory of mind; Executive Function; Self/other interference; Path analysis

Introduction

Theory of mind (ToM) has been defined as the ability to understand that other agents have different beliefs, desires and knowledge to oneself (Premack & Woodruff, 1978). While neurotypical adults clearly understand the principle that other people may have a different perspective to themselves (perhaps unlike children e.g. Perner & Ruffman, 2005; though see Onishi & Bailargeon, 2005), they nonetheless are prone to systematic difficulties arising from conflict between their own perspective (self-perspective) and another's. Most notably, adult participants show "egocentrism" across a very wide range of tasks, whereby their judgment of what someone else sees, thinks, or wants is slower or more error-prone when this differs from what participants themselves see, think, or want (Royzman, Cassidy & Baron, 2003). Understanding the nature and origin of such interference effects is critical for theories about the functional bases of ToM, and understanding variability in these effects between individuals holds out the promise of helping explain why some people are more socially able than others (e.g., Apperly, 2012). Despite the recent proliferation of tasks available for studying ToM in adults, little is yet known about individual differences in ToM performance. Therefore the present study addressed two fundamental questions: Are there systematic individual differences in self-other interference; and are these effects due to individual differences in executive functions?

For Review Only

Studies of ToM in adults have focussed on both typical and atypical participants. Specifically, research has examined healthy young adults (e.g., Bradford, Jentzsch & Gomes, 2015; Qureshi & Monk, 2018), and older adults (Bernstein, Thornton & Sommerville, 2011), patients with brain injury (e.g., Apperly, Samson & Humphreys, 2005; Stone, Baron-Cohen & Knight, 1998), those with Autism Spectrum Disorder (e.g., Baron-Cohen, Jolliffe, Mortimore & Robertson, 1997), psychopathy (e.g., Lockwood, Bird, Bridge & Viding, 2013) and dementia (e.g., Le Bouc, Lenfant, Delbeuck, Ravasi, Lebert, Semah & Pasquier, 2012).

Research in this area has also utilised a variety of different methodologies including dual-tasking (e.g., Bull, Phillips & Conway, 2008; Qureshi, Apperly & Samson, 2010), brain stimulation (e.g., Kalbe et al., 2010), neuro-imaging (e.g., Reiniers, Völlm, Elliott & Corcoran, 2014) and individual differences (Ryskin, Benjamin, Tullis & Brown-Schmidt, 2015). Nevertheless, a unifying feature of the tasks is that they generate a difference between the perspective of the participant (self-perspective) and that of the target (other-perspective), with the resulting interference between the self and other perspectives requiring resolution in order to judge what the other thinks, sees, or feels. This gives rise to the widely-reported phenomenon of "egocentric bias" towards the participant's own perspective, which is almost universally observed in studies of ToM (Royzman, Cassidy & Baron, 2003; Wellman, 2014). However, it is currently unclear whether self-other interference results from the same underlying functional process on different tasks. While everyday experience lends credibility to the hypothesis that some individuals are consistently more egocentric than others, few studies have examined whether this is the case.

For Review Only

One leading hypothesis is that interference between self and other perspectives in ToM tasks forms part of a broader domain of phenomena in which representations relating to self and other must be controlled (e.g., Cook, 2014). For example, observing another's action generates a tendency for "automatic imitation" of the action by one's self, which must be controlled if it is not the action required by the task or ongoing activity (Brass, Bekkering, Wohlschläger & Prinz, 2000). Santiesteban, White, Cook, Gilbert, Heyes and Bird (2012) found that training participants to inhibit automatic imitation improved performance on a test that required participants to use their ToM (the Director Task; Keysar, Lin & Barr, 2003), whereas training generic inhibition did not, suggesting that inhibition of automatic imitation and ToM rely on control over representations related to self and other, and these control processes are not the same as those involved in conventional "inhibition" tasks. Moreover, self-other control appears to depend upon regions of medial prefrontal cortex and temporo-parietal cortex that are distinct from more lateral prefrontal brain regions frequently implicated in analogous non-social executive control (Bardi, Six & Brass, 2017; Brass, Ruby & Spengler, 2009; Brass, Derrfuss & von Cramon, 2004; Wagner, Maril, Bjork & Schacter, 2001). Such findings suggest that the same self-other control process may underlie both imitation inhibition and perspective-taking, and that self-other control processes are in an important sense "domain-specific", and distinct from domain-general executive functions (as generic inhibition training did not improve ToM performance; though see a recent meta-analysis by Darda and Ramsey (2019)). This clearly leads to the prediction that there will be reliable individual differences in self-other control across different ToM tasks, and that these should not be related to individual differences in generic executive function.

For Review Only

However, there is also substantial evidence suggesting that domain-general executive function is involved in ToM, and in particular in self-other control. Research on both children (Carlson & Moses, 2001; Carlson, Moses & Breton, 2002, Carlson, Moses & Claxton, 2004) and adults (German & Hehman, 2006; Bradford, Jentzsch & Gomez, 2015) has shown a relationship between EF and ToM, and this has been suggested to explain deficits and impairments in ToM performance (Bull et al., 2008; McKinnon & Moscovitch, 2007). While there is some evidence of a role for working memory (e.g., Qureshi & Monk, 2018; Ryskin et al., 2015), inhibitory control has been frequently linked with enabling the participant to control interference between self and other perspectives in order to select the one required by the task. For example Qureshi, Apperly and Samson (2010) found that a secondary task taxing inhibitory control disproportionately impaired perspective-taking on trials involving self-other conflict compared with trials without conflict (see also Todd, Cameron & Simpson, 2017). Moreover, there is converging evidence from research using EEG, fMRI, neurospsychological studies of patients with brain injury, as well as TMS, that the inferior frontal gyrus – a brain region frequently implicated in generic inhibitory control – is involved in managing self-other conflict when it arises during visual perspective-taking (e.g., McCleery et al., 2011; Ramsey et al., 20013) and reasoning about false beliefs (e.g., Hartwright et al., 2012, 2016; Samson et al., 2005, 2016). This evidence leads to the prediction that individual differences in self-other control on ToM tasks could be related to individual differences in generic executive control (in particular inhibitory control). This evidence would also be consistent with reliable correlations in self-other control across different ToM tasks, but such relationships should be mediated by individual differences in domain-general executive control, rather than by any specific process for self-other control. Importantly, this way of thinking about ToM tasks also allows for the possibility that though self-other control is typically assumed to be a coherent process common to all ToM tasks, egocentrism and other self/other interference effects are instead a "family" of phenomena that can arise in different ways across different tasks or situations. On such an account self-other interference effects may not correlate across different ToM tasks but would correlate with executive function tasks. The pattern of those correlations would then depend upon when and how self-other interference arises during the ToM task, and the particular executive demands that arose as a result.

In the largest individual differences study of ToM to date, the current study therefore examines the relationship between performance on inhibitory control tasks and two ToM tasks that both involve dealing with interference between self and other perspectives. The two ToM tasks used were a Level 1 Visual Perspective Taking (L1-VPT) Task (Samson et al.,2010) and the Director Task (based on the experiment of Keysar, Lin & Barr, 2003).

The L1-VPT Task requires participants to judge how many dots appear on the walls of a room either from their own perspective or from the perspective of an avatar in the room.

For Review Only

This involves control processes relating to the ability to regulate the choice of responses (e.g., perspective selection) as well as more automatic processes which are associated with the calculation of an 'other' perspective (see Qureshi, Apperly & Samson, 2010). In terms of individual differences, this task has recently been used to create separate measures of "conflict" (relating to the interference between self and other perspectives) and "focus" (relative ease of judgements relating to self versus other; Bukowski & Samson, 2017). These measures allow the separation of two components that may contribute to successful perspective-taking, but which are confounded in standard measures of speed or accuracy of judging the perspectives of others.

The Director task involves the participant following the instructions of a director who has a spatially opposite perspective to their own, and moving objects that are mutually visible (target objects) while avoiding objects visible only to the participant (competitor objects; Apperly et al., 2010). This task therefore also requires participants to resolve interference between self and other perspectives, though errors appear to arise not from difficulty with taking the director's perspective, but with integrating this perspective with the director's message in order to constrain reference (Barr, 2008).

Path analyses were used to model the relationship between four inhibitory control tasks and measures of ToM tasks. Given the near ubiquity of egocentric effects and other self/other interference effects, it is plausible that these effects reflect a common underlying process, with shared variance such that someone who shows (say) high self/other interference on one task will also show high interference on another. Prediction 1: This observation predicts a significant path between measures of self/other interference in the different ToM tasks; an important prediction that has received little empirical attention to date (see Ryskin et al., 2015 for an exception in the context of perspective-taking during communication).

For Review Only

Prediction 2: If inhibitory control is required for dealing with self/other perspective interference and egocentrism (failures to inhibit self-perspective), paths would be expected between inhibitory control and ToM. Prediction 3: If predictions 1 and 2 are both supported then the domain-general account of self-other control predicts that the relationships between self-other interference on different ToM tasks will be fully mediated by inhibitory control, whereas the domain-specific account of self-other control predicts that it will not be mediated. Prediction 4: If only Prediction 2 is supported and if self-other interference arises in different ways for different ToM tasks, then self-other interference on different ToM tasks might correlate with different inhibition tasks.

Method

Participants

One hundred and fifty-four university student participants took part in the study for course credits or cash payment. Participant ages ranged from 18 to 44 (M = 21.8, SD = 4.4), with 31 males and seven who were left-handed (two of whom used their right hand normally). Ethical approval was gained from the departmental ethics committee.

Materials

All the experiments were designed and presented on a 15-inch Samsung SyncMaster 793s monitor connected to a 3.00 GHz Pentium based desktop PC using EPrime 1.1 (Schneider, Eschmann & Zuccolotto, 2002), apart from the visual perspective task which was presented using DMDX (Forster & Forster, 2003). A standard 102 keyboard was used for responses.

Inhibitory Control tasks

For Review Only

Based on Friedman and Miyake (2004), we chose tasks that required inhibition at different stages of processing, including resistance to distracting stimuli, resistance to distracting information held in working memory, and resistance to selecting a prepotent but incorrect response. Further details on the choice of tasks (A1), task methodology (A2) and all individual task results (A3) are shown in Appendix A, where two further IC tasks are described, a Simon task and a Cued Recall task[1]. These were not included in the final analyses due to having no relationship with any of the other tasks in the model and low reliability respectively.

Shape-Matching

Participants were shown a green target shape and a white shape (see. DeSchepper & Trieseman, 1996). On 50% of trials there was a red distracter shape placed over the target. Participants were asked to decide if the target matched the white shape. The dependent variable for the path analysis model was the difference in response times (to correct trials) between the two distracter present conditions and the two distracter absent conditions ((distracter present match + distracter present no-match)/2 – (distracter absent match + distracter absent no-match)/2)).

Stop-signal

Participants were presented with stimuli ("O" or "X") and instructed to respond by pressing the same key on the keyboard (Go trials: 75% of all trials), apart from Stop trials (25%) - trials in which a tone was heard at a set delay (stop-signal delay; SSD) after presentation of the stimulus. The SSD followed a dynamic tracking procedure starting at 250ms, and increased by 50ms on a correct (non-)response, and decreased by 50ms on an incorrect

[1] The original design selected two sets of 3 inhibitory control tasks as indicators for two latent variables for cognitive and response inhibition. However, the covariance structure did not support this theoretically motivated distinction between latent variables, and so tasks with adequate measurement properties were instead utilised as individual predictors.

For Review Only

response (to stop trials). Stop-signal reaction time (SSRT) was calculated as mean Go RT – mean Stop RT.

Go/No-Go (letter-based)

Participants were shown serially presented letters on-screen that they were required to respond to by pressing the spacebar (go trials) (Wager et al., 2005). This was the requirement for all letters except for the letter K (no go trial; 13.85% of all trials). The dependent variable was the False Alarm Rate (FAR) for No-Go trials.

Go/No-Go (image-based)

This task was based on the study of Schmitt et al. (2000) and required a two-step process to respond correctly: the initial step was based on semantic information (image of either a bird or a mammal) and the go / no go step on phonological or orthographical information (Go trial = initial letter was a consonant, No Go trial = initial letter was a vowel; van Turennout et al., 1997). The measure taken was the FAR to no go trials.

For Review Only

ToM tasks

Visual perspective-taking is widely regarded as an important component of children's developing ToM (e.g., Wellman, 2014) and of the mature ToM abilities of adults (e.g., Apperly, 2010). We selected two tasks that have been widely-used in laboratory investigations of ToM in adults, and for which there were task analyses of component processes to which inhibition might contribute: the L1-VPT task requires calculation of another person's visual perspective; the Director Task requires calculation of another person's perspective, plus the use of this information to interpret their instructions. In the discussion we evaluate the implications for the present work of alternative interpretations of these tasks.

L1-VPT task

The visual perspective-taking task used the stimuli and procedure of Experiment 1 in Samson et al. (2010). The stimuli consisted of a picture showing a lateral view into a room with the left, back and right walls visible and with red dots displayed on one or two walls (stimuli were created using the 3D animation program Poser 6, © Curious Lab). A centrally positioned human avatar faced either the left or right wall (Table 2). On 50% of trials the avatar's position meant that s/he saw the same dots as the participants (Consistent condition). On 50% of trials the avatar's position meant that s/he could not see some of the dots that were visible to the participants (Inconsistent condition). Indices of Conflict (inconsistent – consistent perspectives) and Focus (self-perspective – other perspective) were calculated as per Bukowski and Samson (2017). Inverse Efficiency Scores (IES; response time / 1 – error rate) were calculated for these conditions[2]. A higher value in the Conflict index indicated greater difficulty in handling conflicting perspectives, while positive values in the Focus index indicated better performance in taking the other person's perspective than the self-perspective (more altercentric rather than egocentric).

Director task

In a 4 x 4 experimental grid, a critical instruction was given by the director that could refer to a target object (mutually visible to the participant and instructor), or to a competitor object (located in one of the covered slots, and so visible only to the participant). In order to choose

---

[2] Use of IES is not recommended if speed-accuracy trade-offs are shown (indicated by no positive correlation between response times and error rates) and if the average error rate is above .10 (see Bukowski & Samson, 2017). The correlation between response time and error rate was positive (r (142) = .20, $p < .01$), and the mean error rate across conditions was 0.06 (standard deviation = 0.07), meaning both recommendations for using IES were met.

For Review Only

the target object, the participant needed to consider the instructor's perspective. The remaining objects in the grid were unrelated. In the equivalent control grids, the competitor object was replaced by another unrelated object.

The measure taken was the number of errors made to ambiguous trials (where the item in the instruction could refer to a competitor and a target object such as 'mouse' referring to either a computer mouse or a small mammal, for instance) and relational trials (where the best fitting object to the instruction was not mutually visible. For example, referring to a 'small ball', where the smallest ball was a golf ball only visible to the participant, but the smallest mutually visible ball was a tennis ball).

Task order

The order of the tasks was fixed so that participants were exposed to identical stimulus contexts (stimuli and order;see Carlson & Moses, 2001; Carlson et al., 2004; Friedman & Miyake, 2004; Miyake et al., 2000; Miyake, Friedman, Rettinger, Shah & Hegarty, 2001). This was done because a fixed task order is optimal for detecting correlations (e.g. Carlson & Moses, 2001; Carlson et al., 2004; Friedman & Miyake, 2004; Miyake et al., 2000; Miyake, Friedman, Rettinger, Shah & Hegarty, 2001). To check whether the performance on tasks influenced one another, perhaps due to there being a finite resource for executive components (Muraven & Baumeister, 2000), participants completed the Go / No-Go task at the start and end of each session, and results compared. These showed no difference, suggesting that fatigue and/or influence of similar tasks on performance did not occur.

The overall task order was as follows:

For Review Only

Go / No Go Task - L1 VPT (ToM) - Shape-Matching Task - Go / No Go Task (response inhibition)[3] - Stop-signal Task - Director Task (ToM) - Go / No Go (picture) Task. The tasks were split into two sessions. The first session lasted for approximately 70 minutes and consisted of the Go / No Go, Visual Perspective and Shape-Matching tasks, as well as the repeat of the Go / No Go task[4]. The second session also lasted for approximately 90 minutes and consisted of the Stop-Signal, Director and Go / No Go (picture)[5]. There were breaks within each task and between each task, so that participants were not tested continuously for the period of the sessions. Participants did the sessions on the same day (with a break between them) or on separate days (maximum gap between sessions was three weeks).

Analytical procedure

Data screening

The data set was checked for univariate outliers, and pairwise plots examined for any heteroscedasticity. Multivariate outliers were checked as per Tabachnick and Fidell (2001), resulting in 12 participants being excluded. Variances for the dependent variables were adjusted to all be within a 10:1 ratio (Kline, 2005). The final sample for analysis consisted of 142 participants (*M* age = 21.76, *SD* age = 4.37), with 28 males. This resulted in a total of 1132 data points.

After variance adjustment and transformations, the final descriptive statistics for the dependent and independent variables were as shown in Table 1.

Table 1. Final task descriptives

Measure	M (SD)	Variance	Reliability[6]

[3] This was a repeat of the first task in the session to compare performance at the start to the end of the session to see the effect of a.) fatigue on performance, and b.) the possibility of the reduction in the resource of self-control affecting performance by the end of the session (Muraven & Baumeister, 2001).
[4] And the Simon Task which was not included in the final analyses
[5] And Cued Recall task which was also not included in the final analyses
[6] Split-half reliabilities

For Review Only

Shape-Matching (ms) (square root transformation)	21.56 (4.88)	23.77	.92
Go / No Go (FAR)	4.28 (2.12)	4.51	.93
Go / No Go (picture) (FAR; Bird/Mammal)[7]	6.17 (3.05)	9.28	.74
Stop-Signal (SSRT) (log 10 transformation)	16.33 (6.74)	45.52	.99
Director Task (errors) (log 10 transformation)			
Ambiguous	2.34 (2.41)	5.79	.59
Relational	2.75 (2.66)	7.07	.83
Visual Perspective Indices (IES)			
Conflict	3.62 (2.25)	5.06	.85
Focus	-.07 (2.13)	4.55	.88

An initial correlation matrix between the variables is shown in Table 2. For the inhibitory control tasks, positive correlations were shown between the Shape-Matching and Stop-Signal tasks, the Go/No-Go and Go/No-Go (picture) tasks and the Go/No-Go and Stop-Signal tasks. While these are relatively low, this is common for tasks measuring inhibition (e.g. Friedman & Miyake, 2004) and may reflect task impurity. A positive correlation was shown between the Director task variables, and a negative one between the L1 VPT variables. The Shape-Matching task had positive correlations with both the Director (relational) variable, while the Go/No-Go (picture) task had positive correlations with both measures of the Director task. The Stop-Signal task also had a positive correlation with the L1 VPT (Conflict) variable.

For Review Only

Table 2. Correlation matrix between variables

	Shape Matching	Go NoGo	Go NoGo (picture)	Stop-Signal	Director Task (ambiguous)	Director Task (relational)	Visual Perspective (Conflict)
Go NoGo	-.04	-					
Go NoGo (picture)	.00	.23**	-				
Stop-Signal	.15	.18*	.08	-			
Director (ambiguous)	.09	.00	.34**	.02	-		
Director (relational)	.29**	.01	.25**	.05	.65**	-	
Visual Perspective (Conflict)	.06	.14	.00	.24**	-.03	-.04	-
Visual Perspective (Focus)	-.02	.08	-.04	.02	.02	-.03	-.18*

* $p < .05$, ** $p < .01$

[7] As FAR for the bird and mammal trials were significantly correlated and followed the same pattern, they were collapsed to form a single FAR for the go / no go (picture) task

Models

Analyses were carried out using AMOS 25. Path analysis aims to arrive at the most parsimonious model that explains the underlying data and does not significantly differ from it. Increasing the number of parameters in a model tends to improve fit, but necessarily decreases parsimony. The best model optimises fit and parsimony.

While the correlation matrix suggests that for the L1 VPT task, there was only a relationship between inhibitory control and the Conflict Index, in order to test the hypothesis that inhibitory control accounts for self/other perspective interference and egocentrism, paths from all inhibitory control tasks to all ToM measures were included in Model 1 (Figure 1**,** Appendix B). In the second model, all non-significant paths were removed (Figure 2, Appendix B). In model 3, covariances between the inhibitory control variables that were suggested by modification indices were added. This resulted in the final model (Figure 1) which outlines paths from differing inhibitory control variables to both Director task variables and the Conflict index of the L1 VPT task. This model also evidenced relationships between the inhibitory control variables. Due to the sample size and number of parameters, Bollen-Stine bootstrap analyses were conducted to more robustly assess model fit.

For Review Only

Results

Full descriptions, figures and tables for models 1 and 2 are shown in Appendix B and described in brief below.

Initial Model

No relationships were found between the Go / No-Go or Stop-Signal tasks and either of the Director task variables. The Go / No-Go (picture) task predicted performance on both ambiguous and relational variables of the Director task, while the Shape-Matching task only predicted performance on the relational variable. The Stop-Signal task predicted performance on the L1 VPT Conflict Index, and no tasks predicted performance on the L1 VPT Focus Index.

Model 2

Non-significant paths were removed, resulting in only the paths described in the initial model. Model comparisons suggested that removing these paths did not significantly affect the overall model fit (X^2 (12) = 6.29, p = .90).

For Review

Final model (model 3)

For model 3, covariances between the Go / No-Go and Go / No-Go (picture) tasks and between the Go / No-Go and the SST task were added (Figure). This significantly improved the model fit (X^2 (2) = 11.38 p < .01).

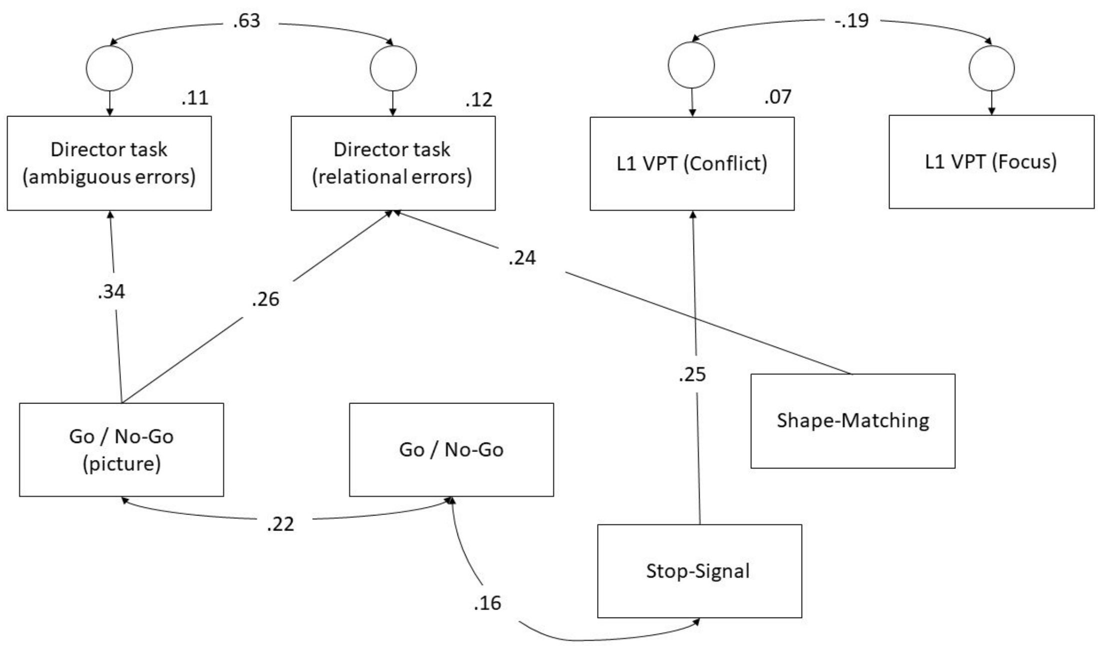


Figure 1. Model 3 (standardised coefficients)

The parameter estimates in Table 3 show that approximately 6% of the variance in the Conflict index was explained by the model, with around 11-12% of the variance of the director task (for both ambiguous and relational DVs) explained.

Table 3. Parameter estimates for final model (model 3)

Path	Unstandardised Estimate (Standard Error)	Standardised Estimate
Shape-matching →Director (relational)	.13 (.03)	.24***
Go/No-Go → Director (relational)		
Go/No-Go (picture) → Director (relational)	.22 (.07)	.26***
SST → Director (relational)		
Shape-matching →Director (ambiguous)		
Go/No-Go → Director (ambiguous)		
Go/No-Go (picture) → Director (ambiguous)	.26 (.06)	.34***
SST → Director (ambiguous)		
SST → L1 VPT (Conflict)	.08 (.03)	.25***

Covariances	Unstandardised Estimate (Standard Error)	Standardised Estimate
Director (relational) ↔ Director (ambiguous)	3.47 (.55)	.63***
L1 VPT (Conflict) ↔L1 VPT (Focus)	-.89 (.40)	-.19***
Go/No-Go ↔ SST	2.29 (1.18)	.16*
Go/No-Go ↔ Go/No-Go (picture)	1.38 (.55)	.22*

Variances	Estimate (Standard Error)
Shape-Matching	23.60 (2.81)***
Go/No-Go	4.46 (.53)***
Go/No-Go (picture)	9.22 (1.10)***
SST	45.20 (5.38)***
e1 (Director (relational))	5.99 (.72)***
e2 (Director (ambiguous))	5.10 (.61)***
e3 (L1 VPT (Conflict))	4.51 (.54)***
e4 (L1 VPT (Focus))	4.73 (.56)***

Squared Multiple Correlations	
Director (relational)	.12
Director (ambiguous)	.11
L1 VPT (Conflict)	.06
L1 VPT (Focus)	.00

* $p < .05$, ** $p < .01$, *** $p < .001$

For Review Only

The model parameters for the final model (Model 3; Table 4) show that the fit was excellent.

Table 4. Model fit parameters

	CMIN	df	*p*	NFI	CFI	AIC	RMSEA	Low	High	Bollen-Stine Bootstrap *p*
Model 1	17.35	10	.07	.88	.94	69.35	.07	.00	.13	.14
Model 2	23.64	22	.37	.84	.99	51.64	.02	.00	.08	.36
Model 3	12.26	20	.91	.92	1.00	44.26	.00	.00	.03	.87

A power analysis was conducted on the final model fit values and sample size using Preacher and Coffman's (2006) power and sample size calculator for RMSEA. This suggests that the current sample size has a power of .92 to detect whether this was a poor fitting model[8]. Testing whether the sample size was sufficient for the final model showed that an ideal sample size should be 107 (with .10 as the RMSEA value for a poor fit; Kenny, 2014[9]). Kline (2005) suggests a 10:1 ratio for free parameters as a realistic figure, which gives an estimated sample size of 80 for the final model. The current sample size is therefore appropriate for the current analyses.

Discussion

In the largest study to date, this research examines individual differences in adults' performance on two ToM tasks, with a particular focus on their ability to manage interference between self and other perspectives. Since this individual difference approach is relatively novel in relation to ToM in adult participants, it is noteworthy that two widely-used tasks – the Director Task (Keysar, Lin & Barr, 2003) and the Level-1 Visual Perspective-Taking Task (Samson et al., 2010) each showed reliable measurement characteristics. Moreover, performance on the two minor trial variants within the Director Task was strongly correlated.

[8] Observed power = .8, df = 20, n =142, null RMSEA = .01 (excellent fit), alternative RMSEA = .00 (final model value).
[9] Observed power = .8, df = 20, null RMSEA = .01, alternative RMSEA = .10

For Review Only

For the L1 VPT task, the relationship between Focus and Conflict indices was negative, suggesting that those better at handling conflicting perspectives were more likely to be better at taking the 'other' perspective[10].

Against this background of successful measurement there was a striking lack of any relationship between indices of these effects in the Director Task and L1-VPT. There was no evidence that the need to control interference between self and other perspectives drew upon a common process in these two tasks (no support for Prediction 1), whether that is a common process specific to self/other control (Bardi, Six & Brass, 2017) or a common process that relies on domain general executive control (Bull et al., 2008; German & Hehman, 2006; McKinnon & Moskovitch, 2007). Self/other interference in the two ToM tasks was correlated with measures of domain-general executive control (support for Prediction 2). However, each ToM task correlated with different executive control tasks. Since no relationship was observed between measures of self/other control, we could not examine whether such a relationship would be mediated by domain-general executive control (Prediction 3).

These results are surprising in light of the convergent evidence presented in the introduction suggesting that self/other interference and egocentrism are very commonly reported phenomena across multiple ToM tasks and other social tasks. Debate has mainly concerned whether the process for handling such interference is specific to self/other control, or domain-general (e.g., Cook, 2014; Happe et al., 2017; Hartwright et al., 2016). The underlying assumption that we are studying one phenomenon is usually taken for granted. Of course we must be cautious about rejecting this assumption, particularly on the basis of the absence of predicted effects. In what follows, we first consider whether our tasks were suitable for use as *measures* of individual differences in self/other interference. We next

For Review Only

[10] Lower Conflict index indicates better at dealing with conflicting perspectives; Higher Focus index suggests relatively better performance when taking 'other' perspective compared to self-perspective.

consider whether resolution of self/other interference is a limiting step for healthy adults. Finally, we consider whether self/other interference might in fact be a family of phenomena. Given the currently limited evidence on individual differences in ToM in adults, we hope this discussion will both help interpretation of our findings and guide badly-needed further work on this topic.

Limitations in ToM measures?

Both the Level-1 Visual Perspective-taking Task and the Director Task generated the anticipated self/other interference effects and measured individual differences in performance reliably. However, this does not entail that they reliably measured individual variability in self/other interference (we also note that while task reliabilities were generally high (< .7) the reliability for ambiguous errors in the Director task was .59). It is important to acknowledge that both tasks were devised as laboratory platforms to investigate egocentrism and self/other interference, primarily through manipulation of task conditions that might affect performance. As recently discussed in the context of individual differences in executive control (Friedman & Miyake, 2017; Hedge et al., 2018), such laboratory tasks are typically developed to minimise individual differences between the performance of participants, in order to maximise the chances of observing between-condition effects.

In response to such limitations, a small number of other tasks have been specifically designed to quantify individual differences in ToM (see Apperly, 2012). However, these tasks generate variance by requiring more subtle or complicated judgements, and in doing so become more opaque with respect to the underlying processes contributing to variance (e.g., Apperly, 2010). A significant objective for future work is therefore to devise new tasks that maximise individual differences, not just in overall "scores", but in theoretically motivated processes

For Review Only

related to ToM, such as self-other interference (or perhaps other parameters related to minds, rather than representing mental states, Conway, Catmur and Bird, 2019). With this important caveat in mind, we nevertheless note that both the L1 VPT and the Director Task have previously shown evidence of construct validity, through correlations with well-established measures of individual differences relevant to social functioning (Focus scores for L1 VPT correlate with the Interpersonal Reactivity Index, Bukowski & Samson, 2017; Errors on the Director Rask correlate with traits related to both autism and schizophrenia, Abu-Akel et al., 2015). We consequently believe the present findings are interpretable, given appropriate caution.

A further question is whether these tasks actually measure ToM at all, since recent evidence suggests that the processing they entail may not be domain-specific to ToM (e.g., Santiesteban, 2014, 2015, 2017). There has been longstanding discussion about the extent to which ToM relies upon distinctive cognitive processes. One position is that reasoning about mental states makes unique cognitive demands (e.g., Leslie., 1987) that are met by a domain-specific module (e.g., Baron-Cohen, 1995; Leslie et al., 2004). A different, and widely-held position is that ToM is more domain-general, making similar cognitive demands to a much broader category of reasoning problems, including reasoning about signs, words and pictures, as well as counterfactual thinking (Perner & Leahy, 2016; Riggs et al., 1998; Roessler & Perner, 2013). Both positions are compatible with a significant role for executive processes in the performance of ToM tasks, but the domain-general account also allows that executive processes contribute directly to representing or reasoning about mental states. Researchers have long sought evidence of domain-specificity by comparing behavioural performance when participants reason about mental states versus pictures, signs, or other non-mental perspectives, with relatively few results supporting domain-specificity. For example, while children with autism perform worse on false belief tasks than comparable false photograph

For Review Only

tasks (Leslie & Thiass, 1992) they perform at similar levels on false belief and false sign tasks (Iao et al., 2014). Typically-developing children's performance on false belief tasks is closely correlated with performance on false sign tasks (Leekam et al., 2008) as well as synonym judgement tasks (Doherty & Perner, 1998), all of which require reasoning about metarepresentations but not mental states (Perner & Leahy, 2016). In response to such results it has been argued that false belief tasks (and ToM more generally) make demands that are exacting, and meaningful in terms of understanding patterns of performance in children and adults, but these demands are not domain-specific to ToM (e.g., Apperly 2010). To our knowledge, nobody has used such results to argue that false belief tasks do not entail representing mental states. One reason for this might be that participants in these tasks are often directly requested to reason about mental states.

Specifically in relation to the L1VPT and Director tasks used in the current study, recent work that has tested the domain-specificity of these effects and found evidence that similar effects are obtained when the avatar is replaced with an arrow and when the director is replaced with a camera the view of which conditionalises their responses in the same way as the director (e.g., Santiesteban, 2014, 2015, 2017). These results fail to find support for the domain-specificity of the effects in the L1VPT and Director tasks, but as illustrated above, there is no necessary link between ToM and domain-specificity in accounts of ToM, and plenty of reasons for thinking that domain-general executive functions play central roles in enabling ToM. Moreover, while it cannot be ruled out that participants ignored instructions to think about the perspectives of the avatar and the director, we see no evidence for this (e.g., Apperly et al., 2010), and no reason why participants would have behaved in this way.

The concern that participants may not have been representing mental states at all has most force in the "self" condition of the L1VPT task in which participants were not explicitly asked to judge the avatar's perspective. Instead putatively "implicit perspective-taking" is

For Review Only

inferred from interference effects on participants' judgements of their own perspective. In this case it is informative to view results from the Self and Other trials separately (see Appendix C). These analyses find that the "altercentric interference" effect for Self trials is not related to any of the executive tasks in the present study, but it is related to "egocentric interference" on Other trials of the L1VPT task. Of course this evidence does not demonstrate that perspective-taking *is* occurring, and different patterns may have been obtained with different inhibition tasks. Nor would a significant correlation have demonstrated that perspective-taking was *not* occurring; it would not be surprising if the magnitude of interference from the avatar's perspective varied as a function of an individual's inhibitory capacity. However, as it stands, the present study provides no evidence that altercentric interference reflects domain general inhibition and not perspective-taking, which is the alternative hypothesis that has sometimes been proposed for this task (e.g. Heyes, 2014).

Is self/other control a limiting step for healthy adults?

Existing evidence suggests that self/other interference effects are particularly large in children (who have limited executive control; Steinbeis, 2016), in individuals with brain injury that has affected their capacity for executive control (e.g., Samson et al., 2005, 2016), and in healthy adults under cognitive load (e.g., Qureshi et al., 2010). Healthy adults without cognitive load also show self/other interference effects, and associated neural correlates (Cook, 2015; Steinbeis, 2016). However, the fact that healthy adults must resolve self/other interference in a given task does not mean that doing so is a limiting factor on their performance such that individuals might vary in their capacity to resolve this interference successfully. On this interpretation, there may be a common underlying process involved in self/other control across different tasks, but since all healthy adults could meet this need reliably it would not contribute to individual differences in performance that would correlate

For Review Only

between tasks. Correlations might only emerge if the ToM tasks placed sufficiently large demands on self/other control that healthy adults did vary in their capacity to meet these demands. Alternatively, the correlations that we did observe between self/other interference effects and measures of domain-general executive function could suggest that our tasks were indeed sufficiently taxing to generate at least some individual differences. However, these correlations accounted for only a small proportion of the variance in performance, and the fact that different executive tasks correlated with each ToM task calls into further question the proposition of a common self/other factor between the ToM tasks.

Self/other control and egocentrism may be a "family" of phenomena that place different executive demands.

It is a parsimonious hypothesis that all self/other interference effects arise from a common underlying process, but this may not be the best explanation for these phenomena. The need to distinguish mental states related to self and other surely is a core feature of ToM tasks, but it has been suggested that this feature can give rise to interference effects in different ways, and at different stages of processing (e.g., Apperly, 2010). This account would predict a "family" of self/other interference phenomena that resemble each other at a general level of description, but are not necessarily correlated because different demands are made at different stages of processing. For example, in the L1 VPT task participants must both *calculate* the perspectives of self and other and *select* the correct response based on one or other of these perspectives (Qureshi et al., 2010). Evidence from event related potentials suggests that perspective calculation occurs first, and involves posterior cortex (putatively temporoparietal junction), while selection occurs later and involves the frontal cortex (putatively the right inferior frontal gyrus; McCleery et al., 2011). Importantly, *both* calculation and selection ERP effects are influenced by self/other consistency, but

For Review Only

participants' response times and errors on this task appear to reflect the effect of self/other consistency on selection of a response according to the perspective required for that trial. The correlation observed in the present study between the conflict index from the L1 VPT task and the Stop Signal Task suggests that this selection may occur very late in processing, perhaps after responses according to both self and other perspectives have already been initiated.

In contrast, while the Director task requires level-1 visual perspective-taking, it additionally requires that information about the director's perspective is integrated with his instructions about which objects to move. Studies of eye movements suggest that participants may process the director's perspective in anticipation of his instruction, and that egocentric eye movements and errors originate in difficulty with the subsequent process of integrating this information with his instruction (Barr, 2008). The proposal that egocentric effects in the director task have a different functional origin from self/other interference effects in the L1 VPT task may explain why these effects did not, themselves, correlate, and why different domain-general executive tasks were correlated with each effect[11].

For Review Only

Conclusion

From the largest exploration of individual differences in in ToM and inhibitory control to date, we found that self/other interference effects in L1 VPT and Director Task were dissociable. Individual differences in performance on the two tasks were unrelated, and differences in performance on each task were related to different tests of executive function. Both findings converge with recent evidence from individual differences in perspective-taking during psycholinguistic tasks (Ryskin et al., 2015). We suggest that self/other

[11] A more specific discussion of the relationship between the individual executive tasks, and their relationships with the ToM tasks is in section A4 of Appendix A.

differences are part of the essence of ToM tasks, but that they give rise to interference effects in a variety of ways, at different stages of processing, resulting in a variety of requirements for executive control processes. We also suggest that the requirement to resolve self/other interference may not be the only, or even the most significant, limiting step for healthy adults' performance on typical tasks. Future work is needed to understand whether self/other interference sometimes places a higher burden for adults, and also to understand other factors that may limit adults' ToM performance and so may explain individual differences in social ability.

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

The data and materials for all experiments are available on request from the lead author

For Review Only

References

Apperly, I. A. (2012). What is “theory of mind”? Concepts, cognitive processes and individual differences. *Quarterly Journal of Experimental Psychology*, 65(5): 825 – 839

Apperly, I. A. (2010). *Mindreaders: the cognitive basis of “theory of mind”*. Taylor & Francis: Psychology Press: London.

Apperly, I.A., Carroll, D.J., Samson, D., Qureshi, A., Humphreys, G.W. & Moffatt, G. (2010). Why are there limits on theory of mind use? Evidence from adults’ ability to follow instructions from an ignorant speaker. *Quarterly Journal of Experimental Psychology*, *63*(6): 1201 – 1217

Apperly, I.A., Samson, D. & Humphreys, G. W. (2005). Domain-specificity and theory of mind: evaluating neuropsychological evidence. *Trends Cogn Sci*, 9(12): 572 - 7

Bardi, L., Six, P. & Brass, M. (2017). Repetitive TMS of the temporo-parietal junction disrupts participant’s expectations in a spontaneous Theory of Mind task. *Social Cognitive and Affective Neuroscience,* 12(11): 1775 – 1782

Baron-Cohen, S, (1995) *Mindblindness: an essay on autism and theory of mind.* MIT Press: Bradford Books.

Baron-Cohen, S., Jolliffe, T., Mortimore, C., & Robertson, M. (1997). Another advanced test of theory of mind: Evidence from very high-functioning adults with autism or Asperger Syndrome. *Journal of Child Psychology and Psychiatry*, 38, 813-822.

Barr, D. (2008). Pragmatic expectations and linguistic evidence: listeners anticipate but do not integrate common ground. *Cognition,* 109(1): 18 – 40

For Review Only

Bernstein, D. M., Thornton, W. L. & Sommerville, J. A. (2011). Theory of mind through the ages: older and middle-aged adults exhibit more errors than do younger adults on a continuous false belief task. *Exp Aging Res*, 37(5): 481 – 502

Bradford, E.E.F., Jentzsch, I. & Gomez, J-C. (2015). From self to social cognition: Theory of Mind mechanisms and their relation to Executive Functioning. *Cognition*, 138: 21 – 34.

Brass, M., Bekkering, H., Wohlschläger, A., Prinz, W (2000). Compatibility between observed and executed finger movements: comparing symbolic, spatial, and imitative cues *Brain and cognition*, 44 (2), 124-143

Brass, M., Derrfuss, J., & von Cramon, D. Y. (2005). The inhibition of imitative and overlearned responses: a functional double dissociation. Neuropsychologia, 43, 89–98

Brass, M., Ruby, P., & Spengler, S. (2009). Inhibition of imitative behaviour and social cognition. Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 364(1528), 2359–67.

Bukowski, H. & Samson, D. (2017). New Insights into the Inter-Individual Variability in Perspective Taking. *Vision.* 1(1): 8 https://doi.org/10.3390/vision1010008

Bull, R., Phillips, L. H. & Conway, C. A. (2008). The role of control functions in mentalizing: Dual-task studies of Theory of Mind and executive function. *Cognition*, 107: 663 – 672

Carlson, S. M. & Moses, L. J. (2001). Individual Differences in Inhibitory Control and Children's Theory of Mind. *Child Development*, *72*(4): 1032 – 1053

For Review Only

Carlson, S. M., Moses, L. J. & Breton, C. (2002). How specific is the relation between executive function and theory of mind? Contributions of inhibitory control and working memory. *Infant and Child Development*, *11*(2), 73 – 92

Carlson, S.M., Moses, L. J. & Claxton, L. J. (2004). Individual differences in executive functioning and theory of mind: An investigation of inhibitory control and planning ability. *Journal of Experimental Child Psychology*, *87*, 299 – 319

Conway, Catmur and Bird (2019). Understanding individual differences in theory of mind via representation of minds, not mental states. *Psychonomic Bulletin & Review*. https://doi.org/10.3758/s13423-018-1559-x

Cook, J. L. (2014). Task-relevance dependent gradients in medial prefrontal and temporoparietal cortices suggest solutions to paradoxes concerning self/other control. *Neuroscience and Biobehavioral Reviews.* 42: 298 – 302

Darda, K. M. & Ramsey, R. (2019). The inhibition of automatic imitation: a meta-analysis and synthesis of previous studies. Biorxiv.

DeSchepper, B. & Treisman, A. (1996). Visual Memory for Novel Shapes: Implicit Coding Without Attention. *Journal of Experimental Psychology: Learning, Memory & Cognition, 22* (1): 27 – 47.

Doherty, M.J., and J. Perner. 1998. Metalinguistic awareness and theory of mind: Just two words for the same thing? Cognitive Development 13: 279–305.

Forster, K. I., & Forster, J. C. (2003). DMDX: A Windows display program with millisecond accuracy. Behavior Research Methods, Instruments & Computers, *35*: 116-124

For Review Only

Friedman, N. P. & Miyake, A. (2004). The Relations Among Inhibition and Interference Control Functions: A Latent-Variable Analysis. *Journal of Experimental Psychology:General, 133* (1): 101 – 135

Friedman, N. P. & Miyake, A. (2017). Unity and diversity of executive functions: Individual differences as a window on cognitive structure. *Cortex*, 86: 186 – 204

German, T. P., & Hehman, J. A. (2006). Representational and executive selection resources in 'theory of mind': Evidence from compromised belief-desire reasoning in old age. *Cognition, 101*(1), 129-152.

Happé, F., Cook, J. L., & Bird, G. (2017). The structure of social cognition: In (ter) dependence of sociocognitive processes. *Annual Review of Psychology*, *68*, 243-267.

Hartwright, C.E., Hansen, P.C. & Apperly, I.A. (2016). Current knowledge on the role ofthe Inferior Frontal Gyrus in Theory of Mind – a Commentary on Schurz and Tholen (2016), *Cortex, 85,* 133 – 136. doi: 10.1016/j.cortex.2016.10.005

Hedge, C., Powell, G., & Sumner, P. (2018). The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences. *Behavior Research Methods*, *50*(3), 1166-1186.

Heyes, C. (2014). Submentalizing I am not really reading your mind. *Perspectives on Psychological Science, 9(2),* 131-143.

Iao, L.-S., & Leekam, S. (2014). Non-Specificity and theory of mind: New evidence from a non-verbal false sign task and children with Autism Spectrum Disorders. *Journal of Experimental Child Psychology, 122,* 1-20.

For Review Only

Iao, L.-S., Leekam, S., Perner, J., & McConachie, H. (2011). Further evidence for non-specificity of theory of mind in preschoolers: Training and transferability in the understanding of false beliefs and false signs. *Journal of Cognition and Development, 12,* 56-79.

Kalbe, E., Schlegel, M., Sack, A.T., Nowak, D.A., Dafotakis, M., Bangard, C., Brand, M., Shamay-Tsoory, S., Onur, O.A. & Kessler, J. (2010). Disassociating cognitive from affective theory of mind: a TMS study. *Cortex*, 46(6): 769 – 780

Kenny, D. A. (2014, January 12). Structural Equation Modelling. Retrieved June 11, 2019, from http://davidakenny.net/cm/causalm.htm

Keysar, B., Lin, S. & Barr, D. J. (2003). Limits on theory of mind use in adults, *Cognition, 89*(1): 25 – 41

Kline, R. B. (2005). *Principles and practice of structural equation modelling (2nd ed.)*. New York: Guilford Press

Le Bouc, R., Lenfant, P., Delbeuck, X., Ravasi, L., Lebert, F., Semah, F. & Pasquier, F. (2012). *Brain*, 135(10): 3026 – 3038

Leekam , S. , Perner , J. , Healey , L. , & Sewell , C. (2008). False signs and the nonspecificity of theory of mind: Evidence that preschoolers have general difficulties in understanding representations . *British Journal of Developmental Psychology , 26* , 485 – 497 .

Leslie, A. M. (1987). Pretense and representation: The origins of "theory of mind". *Psychological Review, 94,* 412–426.

For Review Only

Leslie, A.M., Friedman, O. & German, T.P. (2004) Core mechanisms in “theory of mind”. *Trends in Cognitive Sciences 8(12),* 528-533.

Lockwood, P.L., Bird, G., Bridge, M. & Viding, E. (2013). Dissecting empathy: high levels of psychopathic and autistic traits are characterized by difficulties in different social processing domains. *Frontiers in Human Neuroscience,* 7: 760

McKinnon M. C. & Moscovitch, M. (2007). Domain-general contributions to social reasoning: Theory of mind and deontic reasoning re-explored. *Cognition*, 102: 179 – 218

McCleery, J.P., Surtees, A.D.R., Graham, K.A., Richards, J.E. & Apperly, I.A. (2011). The Neural and Cognitive Time Course of Theory of Mind. *The Journal of Neuroscience,* 31(36): 12849 - 12854

Meurrens, Sabine ; George, Amandine ; Vandermeeren, Yves ; Samson, Dana. (2016) *Exploring the role of the right inferior frontal gyrus in visual perspective taking using transcranial direct current stimulation.*.British Neuropsychological Society (BNS) (London, Great Britain, 26 – 27 October, 2016).

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A. & Wager, T. D. (2000). The Unity and Diversity of Executive Functions and Their Contributions to Complex “Frontal Lobe” Tasks: A Latent Variable Analysis. *Cognitive Psychology*, *41*, 49 – 100

Miyake. A., Friedman, N. P., Rettinger, D. A., Shah, P. & Hegarty, M. (2001). How Are Visuospatial Working Memory, Executive Functioning, and Spatial Abilities Related? A Latent-Variable Analysis. *Journal of Experimental Psychology: General*, *130* (4): 621 – 640

For Review Only

Muraven, M. & Baumeister, R. F. (2000). Self-Regulation and Depletion of Limited Resources: Does Self-Control Resemble a Muscle? *Psychological Bulletin*, *126* (2): 247 – 259

Onishi, K. H., & Bailargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science*, 308, 255–258.

Perner, J., & Leahy, B. (2016). Mental files in development: Dual naming, false belief, identity and intensionality. *Review of Philosophy and Psychology, 7(2),* 491-508.

Perner, J., & Ruffman, T. (2005). Infants' insight into the mind: How deep? *Science*, 308, 214–216.

Preacher, K. J., & Coffman, D. L. (2006, May). Computing power and minimum sample size for RMSEA [Computer software]. Available from http://quantpsy.org/.

Premack, D. & Woodruff, G. (1978). Does the Chimpanzee Have a Theory of Mind. *Behavioural and Brain Sciences*, *4*: 515 – 526

Qureshi, A. W., Apperly, I. A. & Samson, D. (2010). Executive function is necessary for perspective selection, not Level-1 visual perspective calculation: Evidence from a dual-task study of adults. *Cognition*, *117*, 230 – 236

Qureshi, A. W. & Monk, R. L. (2018). Executive function underlies both perspective selection and calculation in Level-1 visual perspective taking. *Psychonomic Bulletin & Review*. https://doi.org/10.3758/s13423-018-1496-8

Reiniers, R.L., Völlm, B.A., Elliott, R. & Corcoran, R. (2014). Empathy, ToM and self-other differentiation: an fMRI study of internal states. *Social Neuroscience,* 9(1): 50 – 62.

For Review Only

Riggs, K. J., Peterson, D. M., Robinson, E. J., & Mitchell, P. (1998). Are errors in false belief tasks symptomatic of a broader difficulty with counterfactuality? Cognitive Development, 13, 73–90.

Roessler, J., & Perner, J. (2013). *Teleology: Belief as perspective.* In Baron-Cohen, S., Tager-Flusberg, H. & Lombardo, M. (Eds.) *Understanding other minds: Perspectives from developmental social neuroscience*, OUP, 35-51.

Royzman, E. B., Cassidy, K. W. & Baron, J. (2003). "I know, you know": Epistemic egocentrism in children and adults. *Review of General Psychology*, 7 (1): 38

Ryskin, R.A., Benjamin, A.S., Tullis, J. & Brown-Schmidt, S. (2015). Perspective-Taking in Comprehension, Production, and Memory: An Individual Differences Approach. *Journal of Experimental Psychology: General, 144* (5): 898 - 915

Samson, D., Apperly, I.A., Braithwaite, J. & Andrews, B. (2010). Seeing it their way: Evidence for rapid and involuntary computation of what other people see. *Journal of Experimental Psychology: Human Perception and Performance, 36*(5): 1255-1266.

Samson, D., Apperly, I.A., Kathirgamanathan, U. & Humphreys, G.W. (2005). Seeing it my way: a case of a selective deficit in inhibiting self-perspective. *Brain,* 128, 1102 – 1111

Santiesteban, I., Catmur, C., Coughlan Hopkins, S., Bird, G., & Heyes, C. (2014). Avatars and arrows: Implicit mentalizing or domain-general processing? *Journal of Experimental Psychology: Human Perception and Performance, 40(3)*:929-37.

Santiesteban, I., Kaur, S., Bird, G., & Catmur, C. (2017). Attentional processes, not implicit mentalizing, mediate performance in a perspective-taking task: Evidence from stimulation of the temporoparietal junction. *Neuroimage, 155,* 305-311.

For Review Only

Santiesteban, I., Shah, P., White, S., Bird, G., & Heyes, C. (2015). *Mentalizing or submentalizing in a communication task? Evidence from autism and a camera control. Psychonomic Bulletin & Review, 22(3),* 844-849.

Santiesteban, I., White, S., Cook, J., Gilbert, S.J., Heyes, C. & Bird, G. (2012). Training social cognition: From imitation to Theory of Mind. *Cognition*, *122*, 228 - 235

Schneider, W., Eschmann, A. & Zuccolotto, A. (2002). *E-Prime user's guide*. Pittsburgh, PA: Psychology Software Tools

Schmitt, B. M., Münte, T. F. & Kutas, M. (2000). Electrophysiological estimates of the time course of semantic and phonological encoding during implicit picture naming. *Psychophysiology, 37*: 473–484

Steinbeis, N. (2016). The role of self-other distinction in understanding others' mental and emotional states: neurocognitive mechanisms in children and adults. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, *371*(1686), 20150074.

Stone, V.E., Baron-Cohen, S. & Knight, R.T. (1998). Frontal lobe contributions to theory of mind. *Journal of Cognitive Neuroscience*, 10, 640-656

Tabachnick, B. G. & Fidell, L. S. (2001). *Using multivariate statistics (4th ed.)*. MA: Allyn & Bacon

Todd, A. R., Daryl Cameron, C. & Simpson, A. J. (2017), Dissociating processes underlying level-1 visual perspective taking in adults. *Cognition*. *159*, 97 – 101

van Turennout, M., Hagoort, P. & Brown, C. M. (1997). Electrophysiological Evidence on the Time Course of Semantic and Phonological Processes in Speech Production.

For Review Only

Journal of Experimental Psychology: Learning, Memory & Cognition, *23* (4): 787 – 806

Wager, T. D., Sylvester, C-Y. C., Lacey, S. C., Nee, D. E., Franklin M. & Jonides, J. (2005). Common and unique components of response inhibition revealed by fMRI. *NeuroImage*, *27*: 323 – 340

Wagner, A. D., Maril, A., Bjork, R. A. & Schacter, D. L. (2001). Prefrontal Contributions to Executive Control: fMRI Evidence for Functional Distinctions within Lateral Prefrontal Cortex. *NeuroImage*, 14: 1337 – 1347

Wellman, H. M. (2014). *Making minds: How theory of mind develops*. Oxford University Press.

For Review Only

Appendix A1

IC task selection and predictions

The classification and choice of inhibitory control tasks was based on Friedman and Miyake (2004). Their initial classifications were designed to measure the different stages of the informational processing of inhibition. This starts with resistance to distracter interference (selecting relevant information and ignoring irrelevant information), followed by resistance to proactive interference (once information has entered working memory) and, finally, prepotent response inhibition (select relevant responses and resist incorrect ones). They found that resistance to distracter interference and prepotent response inhibition loaded on the same latent variable (defined as Response-Distracter Inhibition). As such, it was concluded that Response-Distracter Inhibition and Resistance to Proactive Interference are distinct components of inhibition (Friedman & Miyake, 2004). Our tasks we used were therefore originally chosen with the objective of identifying latent variables for these two components and then assessing their relationship to ToM.

In Friedman & Miyake (2004) Response-Distracter Inhibition tasks included the Go / No-Go task (though these also include some requirement for response selection), Stop-signal task and anti-saccade tasks. Due to laboratory infrastructure, the anti-saccade task could not be used, so we selected a variant Go/No-Go task where the Go and No-Go judgements were based on semantic and orthographic information, as opposed to perceptual information. This follows the principle that latent-variables are best estimated using tasks that differ in as many respects as possible other than the hypothesised underlying common process. From the tasks that Friedman and Miyake (2004) found to load on on the latent variable for Resistance to Proactive Interference we selected the Cued Recall task, Simon task and Shape-Matching tasks.

For Review Only

However, initial data screening and confirmatory factor analyses showed that two of the chosen tasks (Simon and Cued Recall tasks) showed no relationship between any of the other tasks and low reliability respectively. We therefore had to exclude these tasks from further analyses, and follow an alternative analytical approach of using the remaining tasks as individual predictors.

Specific predictions about potential relationships between the individual inhibitory predictors and the ToM tasks are as follows:

1.) The Director task involves the participant avoiding interference from their own perspective when integrating information about the director's perspective with his instructions, and also the need to avoid selecting a distracting incorrect response (Barr, 2008). This leads to the expectation of a role for both Response-Distracter Inhibition and Resistance to Proactive Interference, and therefore potential relationships with any of the selected inhibition tasks.

2.) The L1-VPT Conflict index relates to interference between self/other perspectives, and literature suggests that participant errors and response times reflect the effect of self/other consistency on selection of a response (Apperly, 2010, Qureshi et al., 2010, McCleery et al., 2011). This suggests a functional role for Response-Distractor Inhibition, and so we predicted a relationship between L1-VPT Conflict index and inhibition tasks involving Response-Distractor Inhibition. The Focus index represents relative performance on judgments that relate to the self (versus other). As such, we did not expect to see a relationship with the Focus Index and inhibitory control tasks.

For Review Only

Appendix A2

Task methodology (further details)

Shape-matching task

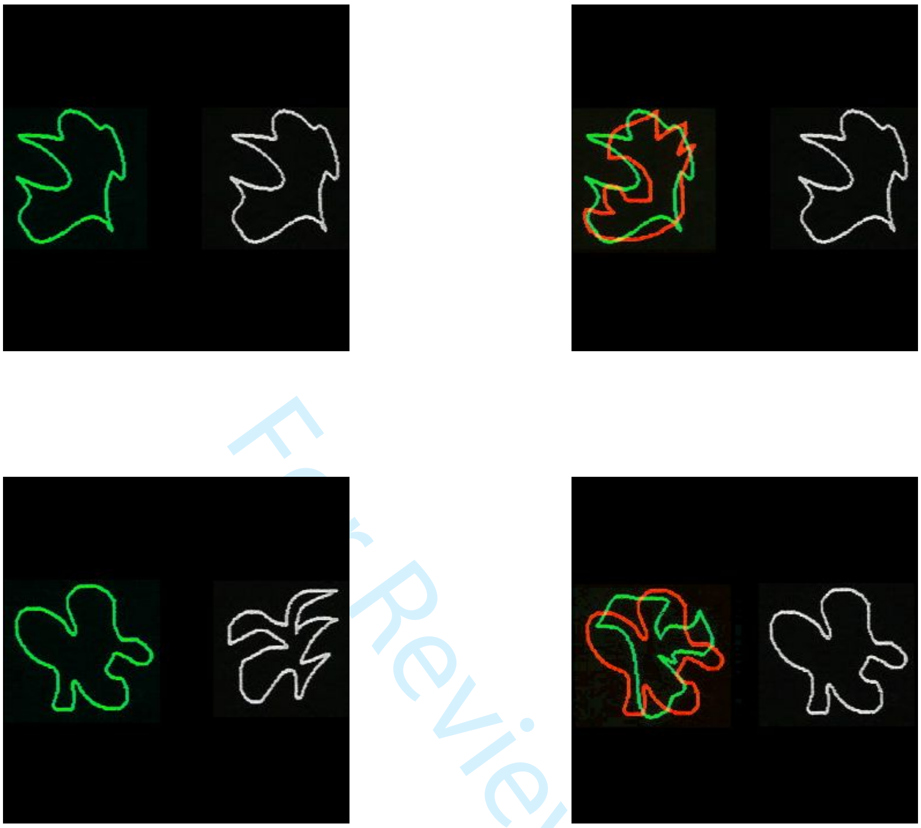
The shapes used were taken from DeSchepper and Treisman (1996); eight each of green (target), red (distracter) and white (matching) shapes. The task was simplified to measure resistance to distracters and resistance to inappropriate responses, rather than negative priming as in DeSchepper and Treisman (1996). This was done by presenting the distracter shapes at the same time as target shapes in the experimental trials, which requires participants to actively keep the distracters from interfering with processing of the target shapes (Endo, Saiki & Sato, 2001).

The shapes measured 135 x 124 pixels and were presented on the left (green/red shapes) and right (white shape) of a 300 x 300 pixel black square. The number of times each type of shape occurred was balanced across the task. Across the total of 112 trials (56 distracter trials and 56 no-distracter trials), the number of times each shape occurred was balanced in four separate blocks of 28 trials. Within these, the number of matching (green = white) and non-matching (red = white) stimuli were also balanced.

Participants were instructed to decide whether the green (target) shape matched the white shape. They were told to ignore the red (distracter) shape when it was present. Each trial had the following procedure: The participants were presented with a READY prompt, to which they had to respond to in order to proceed. A blank screen was then shown for 1100ms, followed by a fixation point for another 500ms. The shapes were then presented until the participant responded. Then another blank screen was shown for 100ms, upon which the

For Review Only

READY prompt appeared for the start of the next trial. Images of the trials are shown in Figure 1:



For Review Only

Figure 1. Trial images: Top left = no distracter, match; Top right = distracter, match; Bottom left = no distracter, no match; Bottom right = distracter, no match.

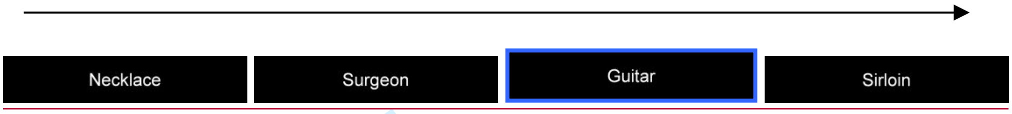
Cued Recall task

The task was designed to measure pro-active interference and was based on the study of Tolan and Tehan (1999). Participants were presented with either one block (filler trials) or two blocks (control and experimental trials) of four serially presented words (one per 1000ms). In all trials this was followed by a numerical magnitude judgment task as a distraction[1]. Participants were then presented with a cue word to vocally recall a word from the only block for filler trials or from the second block (control and experimental trials). The

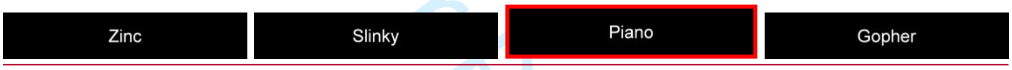
[1] Participants were presented with eight serially presented random numbers of two digits (one per 1000ms), and had to judge aloud if they were above or below 50.

cue word was a category word related to the target word in the second block, but also to a foil word in the first block of experimental trials (there was no foil word in control trials). The foil words were always a more common instance of the category given by the cue word (than the target words), so increasing pro-active interference. An example experimental trial is shown in Figure 2.

Block 1 (read aloud):



Block 2 (read silently):



Distracter task

Cue:

Figure 2. Cued Recall experimental trial (for the purposes of illustration, the blue border indicates foil word, and the red border indicates target word; these borders were not present in experimental stimuli)

For Review Only

The experiment consisted of 10 one block filler trials and 40 two block trials, the latter consisting of 20 control and 20 experimental trials. To increase interference, the ALOUD block always came before the SILENT block. This was followed by the distracter task and then the cue word. The dependent measure was the number of correct control trials minus the number of correct interference trials. This was chosen because the number of correct control trials was taken to be the measure of the participants' general recall ability (dealing with the interference caused by the mixed modality (reading aloud v silently), and the distracter task). The interference trials additionally had pro-active interference from the foil word in the first block. The measure therefore assessed participants' ability to deal with pro-active interference while taking into account their general recall ability.

Simon task

The task was based on Peterson et al. (2002). Participants were shown horizontally-orientated arrows (left or right), presented on the left or right of a central fixation point. Participants were instructed to respond according to the orientation of the arrows, not the position. As the arrow position is visually more salient than the arrow orientation, responding to the orientation requires the inhibition of any response to the position. The majority of trials were congruent for orientation and position (328 congruent, 82 incongruent)., with left and right positions and orientation balanced equally across four blocks of 102 trials for a total of 408 trials.

Each trial began with a fixation point for 200ms, followed by an arrow for a maximum of 1300ms or until a response was made. The difference between response times to accurate incongruent and congruent trials was taken to be a measure of any inhibitory processing cost present only in the incongruent trials (any processing cost common to the incongruent and congruent trials would be accounted for).

For Review Only

Go / No-Go task

The experiment consisted a total of 520 trials across eight blocks of 65 experimental trials (each block also had an additional five trials at the start (these were eliminated in analyses as practice trials). There were either eight or ten no go (K) trials (and a corresponding 57 or 55 go trials) in each block Each letter was presented centrally on a black screen.

Participants viewed a series of letters presented on a computer screen to which they had to respond by pressing the spacebar, except for the letter 'K'. If the participant responded incorrectly to the letter K, a tone would be heard for 250ms. Each trial consisted of an initial fixation cross for 250ms, followed by a letter. Each letter remained on-screen for 500ms, or until the participant responded.

Go / No-Go (picture task)

This task was based on the study of Schmitt et al. (2000). Participants were presented with images of either birds or mammals (there was a familiarisation procedure prior to the task[2]). Participants responded by pressing the Q key if a bird was shown, and with a P if the image was of a mammal. If the name of either began with a vowel, they were instructed not to respond. This task creates a Go/No-Go situation because an initial Go decision is based on the rapidly-available semantic information (bird/mammal category) which must then be cancelled or allowed based on later-available phonological or orthographic information, (van Turrenout et al., 1997). The majority of images were go trials (36 each for birds and mammals, compared to 9 no-go trials for each), repeated twice and balanced across eight blocks for a total of 180 trials. The first three trials were removed from further analyses as practice trials.

During the experiment, after an initial fixation cross for 250ms, the images were presented for 750ms, or until the participant responded. If the participant responded incorrectly, a tone was heard for 250ms. The images were scaled to 500 x 500 pixels.

For Review Only

[2] The participants were shown (serially) all the images of birds and mammals, together with their names, at the start of the experiment (each image remained on-screen until the participant pressed a key). They were then shown the images again, in a different order, for 3000ms, in which time they were instructed to name each image. Pilot work had already shown people were familiar with the mammals and birds used.

Stop-signal task

A total of 256 trials across eight blocks were presented to participants, consisting of either a "X" or "O". After an initial fixation cross for 500ms, the stimuli were shown on-screen for 1000ms or until the participant responded, followed by a blank screen for 2000ms (or again until the participant responded). 64 of the trials were stop trials, where a tone was heard after a variable delay, with the remaining being go trials, all balanced across blocks. Participants were instructed not to respond if they heard the tone, and the stop signal delay started at 250ms after stimuli presentation and then followed a dynamic tracking procedure (+50ms if they responded correctly, and -50ms if incorrectly). Letters were presented centrally on a 400 x 300 pixel black screen.

For Review Only

L1-VPT task

There were a total of 96 matching trials: 48 trials in which participants were asked to verify their own perspective (with 24 consistent perspective trials and 24 inconsistent perspective trials) and 48 trials in which participants were asked verify the avatar's perspective (with 24 consistent perspective trials and 24 inconsistent perspective trials). In addition to these 96 test trials, 96 mismatching filler trials were included (with the same distribution as for test trials) as well as 16 additional anti-strategy filler trials (eight matching and eight mismatching in which no circles were pinned on the wall). The 208 items were split into four blocks of 52 items. Each trial image measured 640 x 480 pixels, with the avatar presented centrally. Between zero and three dots were presented at the avatar's eye level on neither, both or either wall.

Before being shown the image, participants were cued with a verbal perspective description that informed them which perspective to take ("you" versus "he"/"she") and which perspective content to verify (i.e., a number of dots visible, between zero and three). They were then asked to decide whether the cue information matched the relevant content shown in the image. In half of the trials, the cue matched the image and on the other half the cue did not match the image. An initial fixation cross was on-screen for 750ms, followed by the cues, which were also shown for 750ms. The gap in between fixation and cues was 500ms. The stimuli picture remained on-screen until the participant responded. If no response was received within 2000ms of the picture being on-screen, the trial was recorded as no response. Trial conditions are shown in Figure 24:

For Review Only

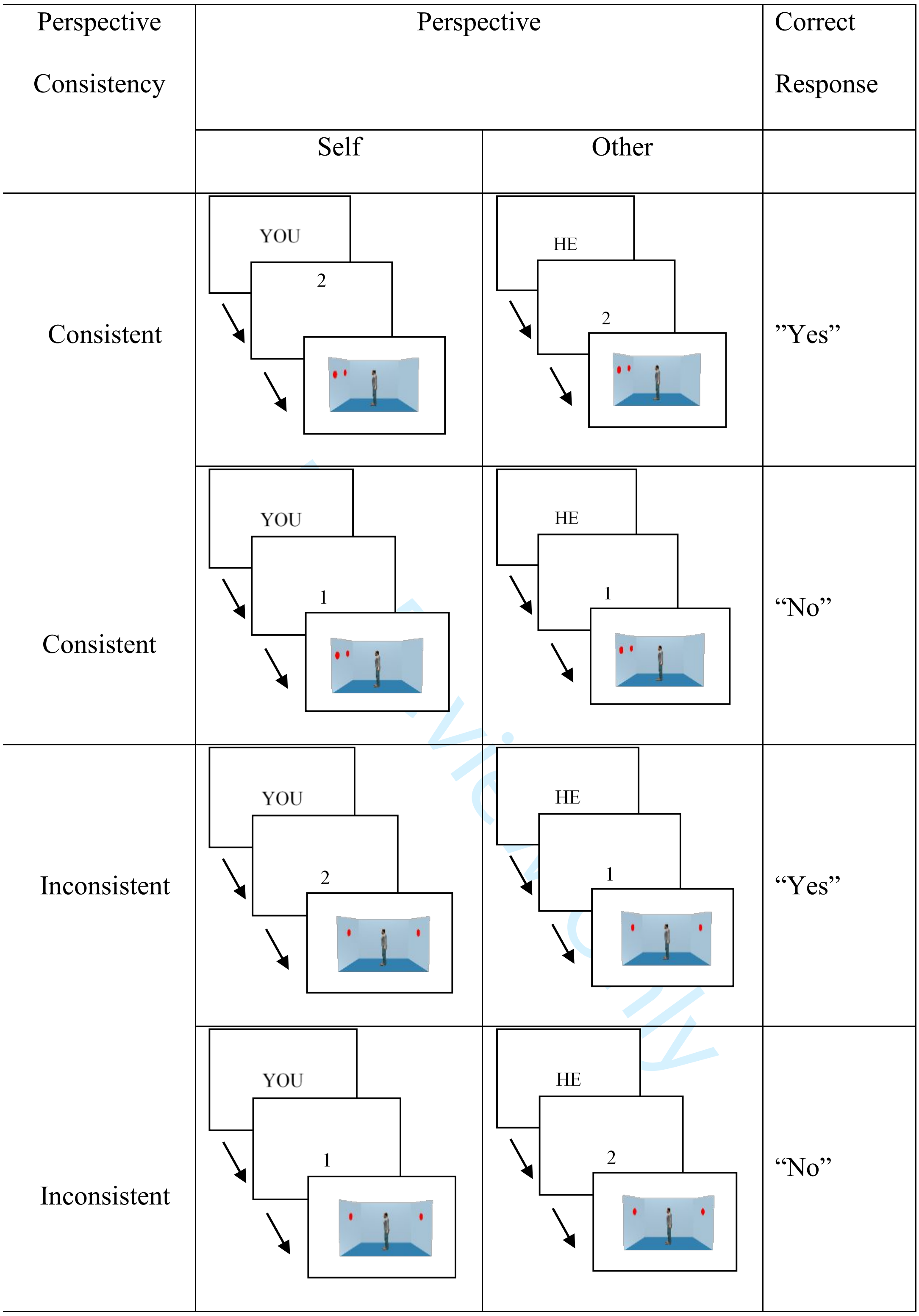


Figure ~~2~~3. L1 VPT trial conditions

Director task

Stimuli replicated those in Apperly et al. (2010). A total of 32 different grid arrays were used, each with 16 slots (4 x 4 grid). In each grid, 5 slots were occluded from the view of the director (male figure with a male voice), with the remaining 11 visible to the participant and the director. These were arranged in one of four different patterns, sized at 720 x 540 pixels. Each grid contained eight items, created from simple cartoon images. There were128 instructions used across the experiment, with between three and five for each grid. Two practice grids were also used with the same layout.

Sixteen of the grids were 'experimental' grids, with one critical instruction that could refer to an object in an occluded slot (if only the participant's perspective was taken into account), or to one in a mutually visible slot (if the director's perspective was accounted for). The position of this critical instruction varied between first and fourth in the series for each grid. Half of the experimental grids were 'relational' trials, and half were 'ambiguous' trials. Relational instructions followed the pattern "Move the [adjective noun] one slot left/right/up/down", and the noun occurred on average 1,047ms after the start of the instructions. For ambiguous trials, the instructions followed the formula "Move the [noun] one slot left/right/up/down", and the noun occurred on average 742ms after the instruction. All other instructions were filler items that followed the same pattern of the critical instruction, and always referred to mutually visible objects.

The remaining 16 grids were matched control grids (also half relational and ambiguous), where the item referred to in the critical trial was replaced with an item that could not be a potential referent (e.g. aeroplane for small ball). Every grid was presented for 5000ms before instructions began, and these were then given at 5000ms intervals. An example grid is shown in Figure ~~3~~4.

For Review Only

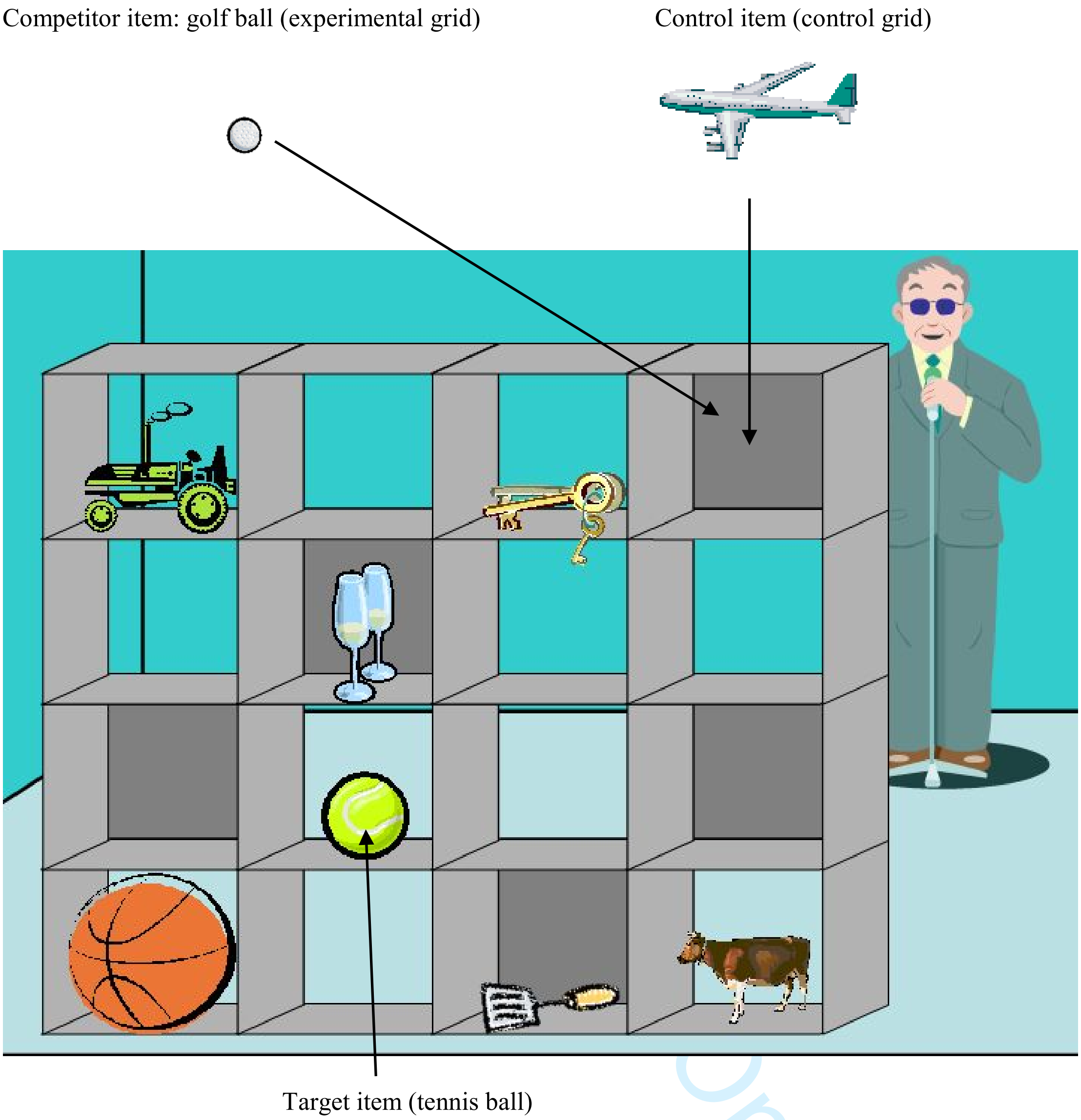


Figure 4. Example grid for a relational grid of the Keysar task

Critical instruction: Move the small ball down one slot

In the experimental grid, the smallest ball from the participant's viewpoint is the competitor item (golf ball), but the target item (the smallest ball from the director's viewpoint) is the tennis ball.

In the control grid, the smallest ball from the participant viewpoint is the target item (tennis ball), which is the same as for the director.

Only

Appendix A3

Individual task results

The data screening process consisted of several steps. The following steps were adapted from those used by Friedman and Miyake (2004) and Oberauer (2005) in order to reduce the effect of any extreme response time values on the mean.

The critical dependent variables of the various tasks were calculated, and any values more than three standard deviations above or below the task mean were replaced by the cut-off value (of three standard deviations above or below the mean)[3]. All pairwise plots were then examined for nonlinearity and heteroscedasticity. The distributions of the critical dependent variables were then checked, as were univariate outliers, and any transformations were carried out. Checks for multivariate outliers were then done, and finally the variance of each variable was examined and modified as per Kline (2005). Each task required different calculations for its critical dependent variable, which are all detailed below. The reliability of all the critical measures was calculated by adjusting split-half correlations using the Spearman-Brown prophecy formula (Salthouse et al., 2006). Each task was first analysed individually to establish that the expected effects were observed.

Go/No-Go

The data were analysed using a 2-way (position; start v end) repeated measures ANOVA. The dependent variable was the False Alarm Rate (FAR) on No-Go trials.

[3] For the L1 VPT we followed the procedure of Samson et al. (2010) and used 2.5 SD rather than 3SD).

For Review Only

There was no difference in the performance of participants in the task carried out at the start of the session and the performance in the task carried out at the end of the session (see Figure A1; $F_{(1, 141)} = .35$, $p = .55$, $\eta_p^2 = 0.02$), so FAR was collapsed over position. FAR was significantly different from zero (t (141) = 23.53, $p < .01$).

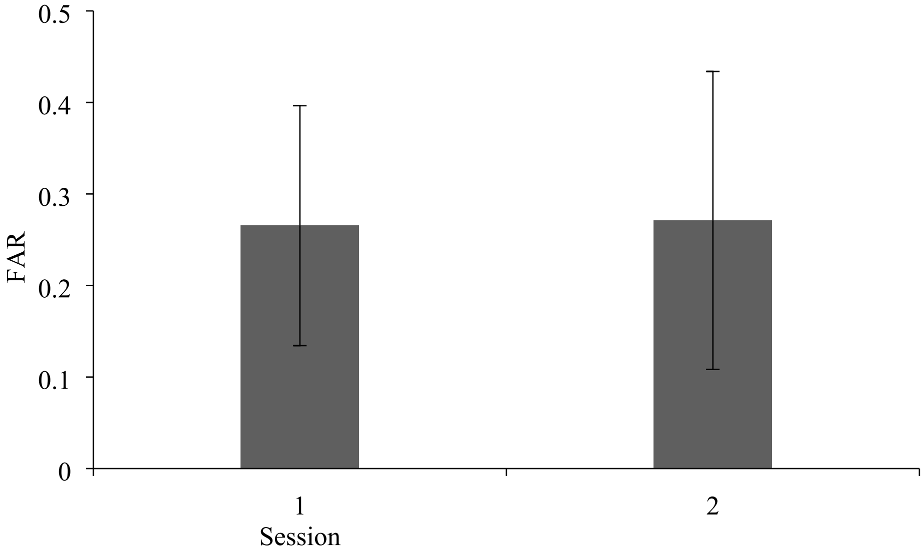


Figure A1. FAR by session (bars = standard deviation)

iew Only

As the distribution was skewed, a square-root transformation was carried out. The reliability of the measure was excellent (.93).

Go / No-Go (picture)

A repeated measures ANOVA was conducted, with image type (bird v mammal) as the condition, and FAR as the dependent variable. There was a main effect of image, with higher FAR for bird images than for mammals ($F_{(1, 141)} = 9.67$, $p < .01$, $\eta_p^2 = .07$; see Figure A2). However, due to the correlation between the FARs for birds and mammals suggesting similar processes were involved, they were collapsed to form a single FAR measure. The reliability of the measure was satisfactory (.74).

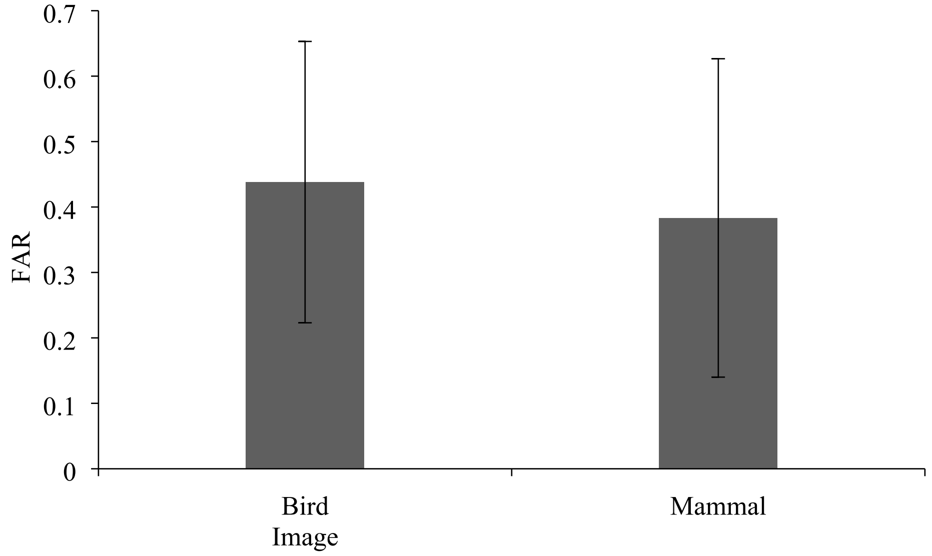


Figure A2. FAR by image (bars = standard deviations)

For Review Only

Stop-signal task

The critical dependent variable was the mean response time to the go trials (correct response only) minus the mean response time to the stop trials. The mean stop signal response time (SSRT) values of the participants was similar to that found in previous literature (232ms (SD = 237.09) compared with 189ms in Aron et al. (2003)).

The distribution of SSRT was extremely non-normal. In order to normalise the distribution the times above three standard deviations were replaced by the highest value below three standard deviations plus one (for the lowest above three standard deviations, plus two for the second lowest, etc.). This was done as described above in three waves – twice for values above three standard deviations, and once for values below three standard deviations, with two participant values being replaced (the mean SSRT after this trimming was 219ms (SD =

172.41)). These procedures are as described in Tabachnick and Fidell (2001). This resulted in the distribution being reasonably normal. The reliability was excellent (.99).

Shape-Matching Task

Data were trimmed by condition for each participant individually, looking at correct responses only (5.1% of the responses were incorrect). Response time values that were more than three standard deviations away from the condition mean for a participant were replaced by the cut-off value (of three standard deviations from the condition mean). This was done for 1.41% of trials. A 2 x 2 repeated-measures ANOVA was used to examine the data. The factors were distracter (present v not present) and match (did the target match the green shape, yes v no) and the measure was the mean response time for each condition.

There was an effect of distracter ($F_{(1, 141)} = 122.73$, $p < .01$, $\eta_p^2 = .47$). If the distracter shape was present, the response time was significantly longer than when the distracter shape was not present (distracter = 1296.10ms > no distracter = 973.07ms). There was also an effect of match ($F_{(1, 141)} = 25.13$, $p < .01$, $\eta_p^2 = .15$). Participants took significantly longer to respond to trials in which the green shape did not match the white shape than to trials where they did not (match = 1093.12ms < no match = 1176.05ms). There was an interaction between distraction and matching ($F_{(1, 141)} = 4.65$, $p < .05$, $\eta_p^2 = .03$), as can be noted in Figure A3. Errors were similar for all conditions (proportion of errors in distracter no match condition = .07, in distracter match condition = .05, in no distracter no match condition = .04, in no distracter match condition = .04), indicating that there was no particular trade-off between speed and accuracy.

For Review Only

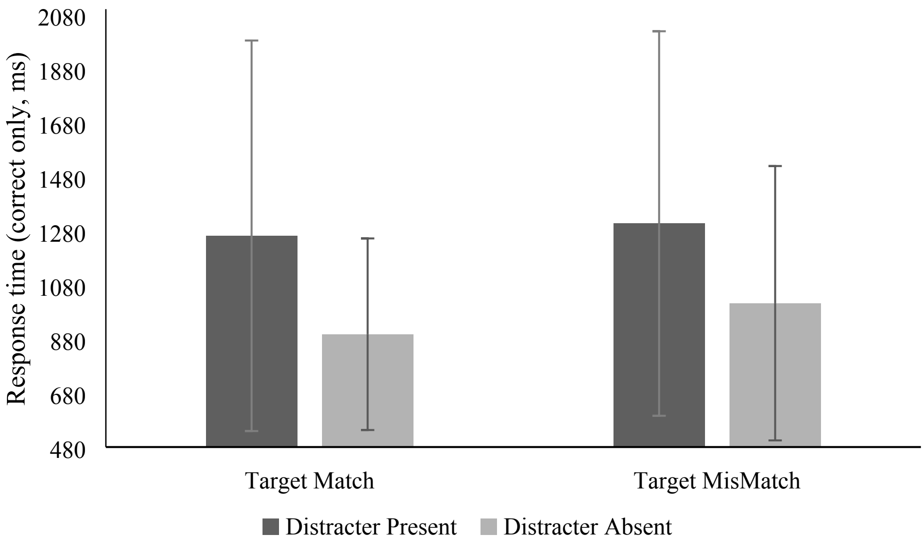


Figure A3. Response time by condition (bars = standard deviations)

For Review Only

Simple main effects showed that for both match and mismatch conditions, participants took significantly longer to respond to trials where the distracter was present (p's $< .01$). When the distracter was present, there was no difference between the match and mismatch conditions ($p = .08$), but participants took longer to respond to the mismatch condition when the distracter was not present ($p < .01$).

The critical dependent variable was the difference in mean response time between the distracter no-match condition and the no distracter match condition. The distracter no-match condition appeared to have the largest amount of interference from the distracter shape, as this shape matched the target. The no distracter match condition had the least amount of interference, and gives a baseline measure of the response time to the simple matching process. The difference between these conditions is therefore a measure of the interference caused by the distracter shape only, which should require inhibitory control. Any values more than three standard deviations away from the mean were replaced by the cut-off value (of three standard deviations above or below the mean). The distribution of the critical shape matching task dependent variable was extremely non-normal, so underwent a log 10 transformation resulting in a normal distribution. Reliability was adequate (.66).

Cued Recall

As the data analysed for the cued recall task were the number of correct trials, they were not trimmed (9.5% of the responses were incorrect). Control and interference trial performance was analysed, using a 2-way repeated measures ANOVA. In both trial types participants were cued to recall a target word from a list (the second of two). In interference trial a foil word was also present (in the first list of the two). The first list of words was always read aloud, and the second set was always read silently. Before the analysis was carried out, missing values for two participants[4] on task were estimated through imputation and regression estimation procedures. The result of the estimations from these two procedures did not differ significantly, so the estimates from the imputation method using the program NORM (Schafer, 1999) were used.

Participants were more accurate in the control condition than in the interference condition ($F_{(1, 141)} = 191.38$, $p < .01$, $\eta_p^2 = .59$; see Figure A4). This effect is probably due to the effect of the foil word in the interference trials affecting recall of the target word, as expected. The critical dependent variable was calculated as the difference between the number of correct control trials and the number of correct interference trials. The mixed modality of the presentation and distracter task were common to both trials, so this dependent variable should remove any contribution from that. The dependent variable should therefore only measure the level of interference caused by the foil word in recalling the target word, interference which

[4] These participants failed to complete the cued recall task due to time constraints (arrived late to session).

For Review Only

should be mediated by inhibitory control. The distribution of the critical dependent measure was normal, so no transformations were required. However, the reliability of the measure was extremely low (.12), so this measure was excluded from the final analyses.

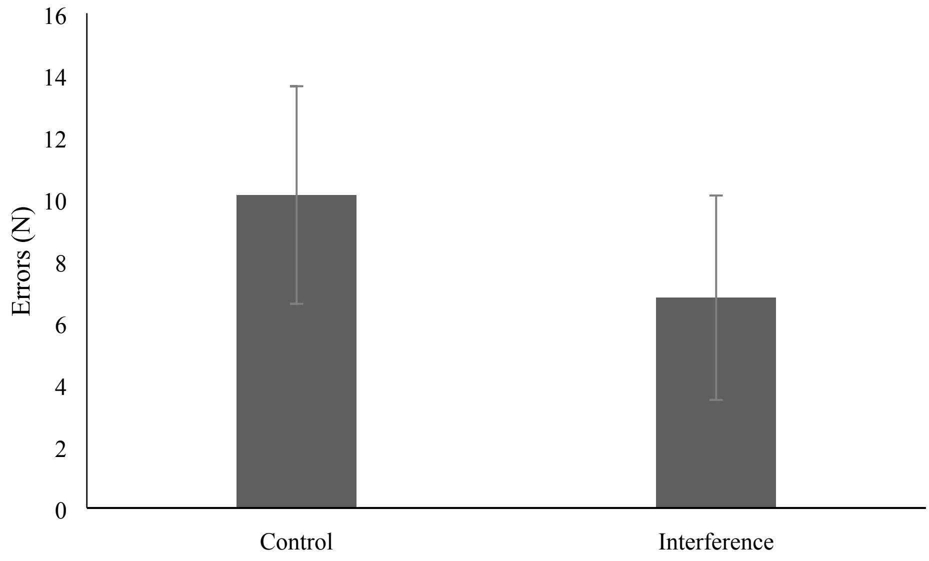


Figure A4. Errors (N; bars = standard deviation)

Simon task

The data were trimmed by condition for each participant for correct responses only (5.3% were incorrect responses). Response time values that were more than three standard deviations away from the condition mean for a participant were replaced by the cut-off value (of three standard deviations from the condition mean). A 2 (position; left x right) x 2 (orientation; left x right) repeated-measures ANOVA was used to examine the data, with the dependent variable mean response time for correct trials.

There was no effect of the position of the arrow on the response time of participants ($F_{(1, 152)} = .77$, $p = .38$, $\eta_p^2 = .01$), whilst there was a significant effect of orientation ($F_{(1, 152)} = 36.34$, $p \leq .01$, $\eta_p^2 = .19$), with responses to left-orientated arrows significantly slower than to rightorientated arrows. This may be due to the majority of the sample being right-handed participants.

There was a significant interaction between position and orientation (F $_{(1, 152)}$ = 567.16, $p \leq .01$, $\eta_p^2 = .79$), as shown in Figure . Simple main effects were conducted to analyse this further.

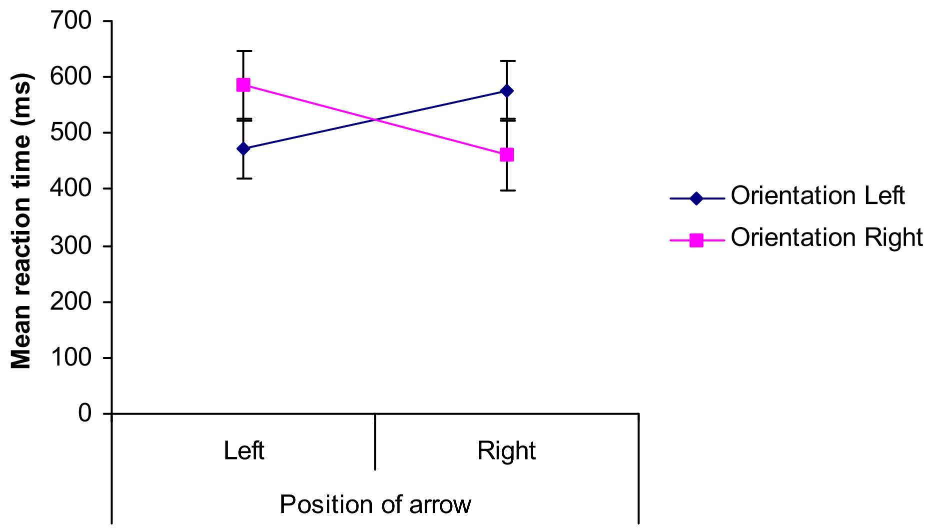


Figure A5: Mean response times (and standard error) to differing positions and orientations of arrow

w Only

Simple main effects indicated that when the orientation and position of the arrow was congruent, response times were significantly faster, relative to when they were incongruent (all p's < .01). The pattern of errors showed a similar effect, with more errors made when position and orientation were incongruent (.13) compared to when they were congruent (.35), indicating no speed – accuracy trade-off.

The critical dependent variable of the Simon task was the difference between the mean response times to incongruent stimuli and congruent stimuli. The reliability of this measure was satisfactory (.92). This was taken to measure the interference between the dominant response and correct response to the incongruent stimuli, interference that should require inhibitory control to resolve. The distribution was examined, was found to be skewed, so was square root transformed, resulting in a normal distribution.

Theory of Mind tasks

L1 VPT

Data trimming was conducted on correct responses only on the matching trials (4.78% of the data were incorrect responses). Response times that were 2.5 standard deviations away from the mean were eliminated (2.04% of the data; in line with Qureshi et al., 2010) and so were response omissions due to the timeout procedure (responses over 2000ms; 0.47% of the data) for a total of 3.92% of all trials. The data were analysed using a 2 x 2 repeated-measures ANOVA, with consistency (consistent v inconsistent) and perspective (self v other) as the factors. The measure first investigated was response time in each of the conditions.

There was a main effect of consistency ($F_{(1, 141)} = 288.59$, $p < .01$, $\eta_p^2 = 0.67$), with the response time in the inconsistent condition significantly higher than that in the consistent condition. There was no effect of perspective ($F_{(1, 141)} = .85$, $p = .36$, $\eta_p^2 =.01$). There was a significant interaction between consistency and perspective ($F_{(1, 141)} = 63.52$, $p < .01$, $\eta_p^2 = 0.31$; see Figure A6) that was analysed further using simple main effects.

For both other and self-perspective judgments, response times to inconsistent trials were significantly longer than for consistent trials (p's < .01). For consistent trials, self-perspective judgements were significantly longer than other perspective judgement trials ($p < .01$),

For Review Only

whereas for inconsistent trials, other perspective judgements were significantly longer than self-perspective judgements ($p < .01$)

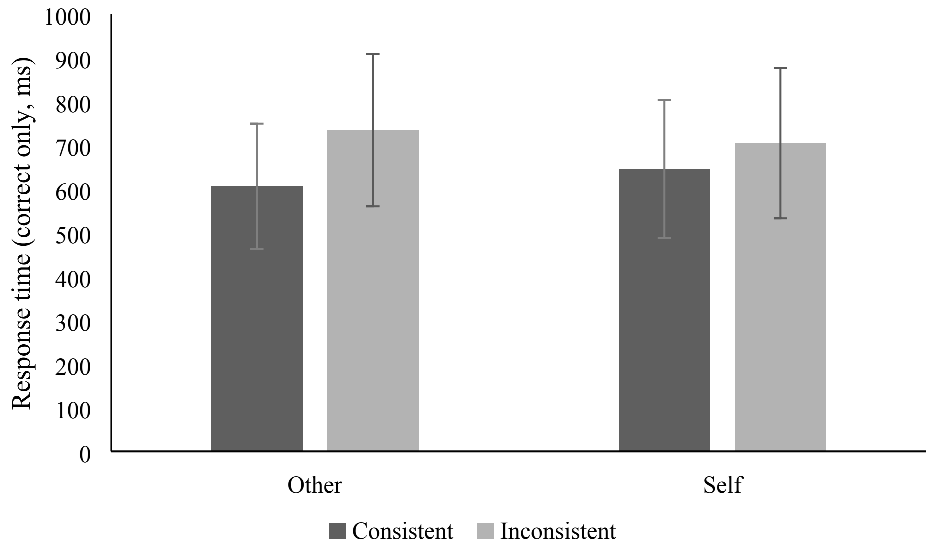


Figure A6: Response times (bars = standard deviation) by condition and trial

The errors (not trimmed) in each condition were examined using an identical 2 x 2 repeated measures ANOVA. There was a main effect of consistency ($F_{(1, 141)} = 176.31$, $p < .01$, $\eta_p^2 = .56$), with more errors in the inconsistent condition than in the consistent condition. There was again no effect of perspective ($F_{(1, 141)} = 1.09$, $p = .30$, $\eta_p^2 = .01$). There was a significant interaction between consistency and perspective ($F_{(1, 141)} = 14.46$, $p < .01$, $\eta_p^2 = .09$), investigated further using simple main effects (see Figure A7).

For both other and self-perspective judgments, there were more errors on inconsistent trials than on consistent trials (p's $< .01$). For consistent trials, there were more errors for self-perspective judgements than for other perspective judgement trials ($p < .01$), whereas for inconsistent trials, there were more errors for other perspective judgements than for self-perspective judgements ($p < .05$)

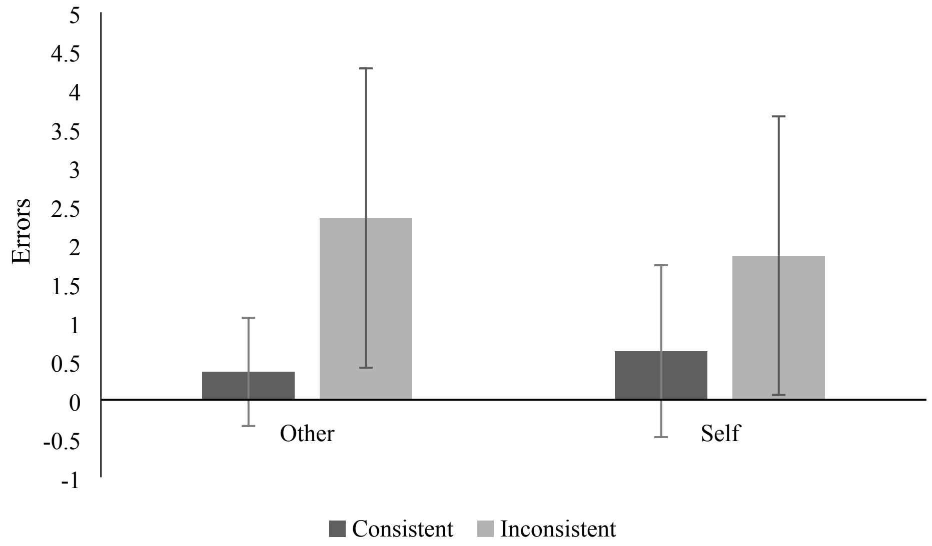


Figure A7: Visual perspective error proportions (bars = standard deviation) by condition and trial

There was no difference in either response time or error rate between the participants taking their own perspective (self condition) or taking the avatar's perspective (other condition), indicating that in general there was no difficulty in taking one perspective over another. Other perspective trials show higher response times and error rates than self perspective trials, possibly due to greater interference of the self perspective in the inconsistent trials only.

This suggests that the L1 VPT was showing similar patterns to the original study of Samson et al. (2010).

The critical dependent variables for the task were calculated as per Bukowski & Samson (2017), resulting in a Conflict Index (inconsistent – consistent perspectives) and a Focus Index (self-perspective – other perspective), using Inverse Efficiency Scores (IES; response

Review Only

time / 1 – error rate)[5]. A higher value in the Conflict index indicated greater difficulty in handling conflicting perspectives, while positive values in the Focus index indicated better performance in taking the other person's perspective than the self- perspective (more altercentric rather than egocentric). Results showed that the mean Conflict Index for participants was .21 (SD = .13), with an excellent reliability of .85, and the mean Focus Index was -.004 (SD = .13), with again an excellent reliability of .88. The variances were adjusted to within a 10:1 ratio of the other task measures (Kline, 2005).

For Review Only

Director Task

For the response time analyses, the data were trimmed by condition for each participant individually, looking at correct responses only (12.5% were incorrect responses to the competitor item, 1.4% were responses to another object). Response time values that were more than three standard deviations away from the condition mean for a participant were replaced by the cut-off value (of three standard deviations from the condition mean). This was done for 0.48% of trials. The data were analysed using a 2 x 2 repeated measures ANOVA, trial type (control v experimental) and condition (ambiguous v relational). Response time was looked at first, with correct responses only used. There was no effect of trial type ($F_{(1, 141)} = .12$, $p = .73$, $\eta_p^2 = .00$). There was an effect of condition ($F_{(1, 141)} = 187.23$, $p < .01$, $\eta_p^2 = .59$), with the response time to ambiguous trials being significantly faster than to relational trials (see Figure A8). There was also a marginal significant

[5] Use of IES is not recommended if speed-accuracy trade-offs are shown (indicated by no positive correlation between response times and error rates) and if the average error rate is above .10 (c.f. Bukowski & Samson, 2017). The correlation between response time and error rate was positive (r (142) = .20, $p < .01$), and the mean error rate across conditions was 0.06 (standard deviation = 0.07), meaning both recommendations for using IES were met.

interaction between trial type and condition ($F_{(1, 141)} = 3.71$, $p = 0.06$, $\eta_p^2 = .02$). These findings, compared against the error findings described below, show no indication of a speed-accuracy trade-off.

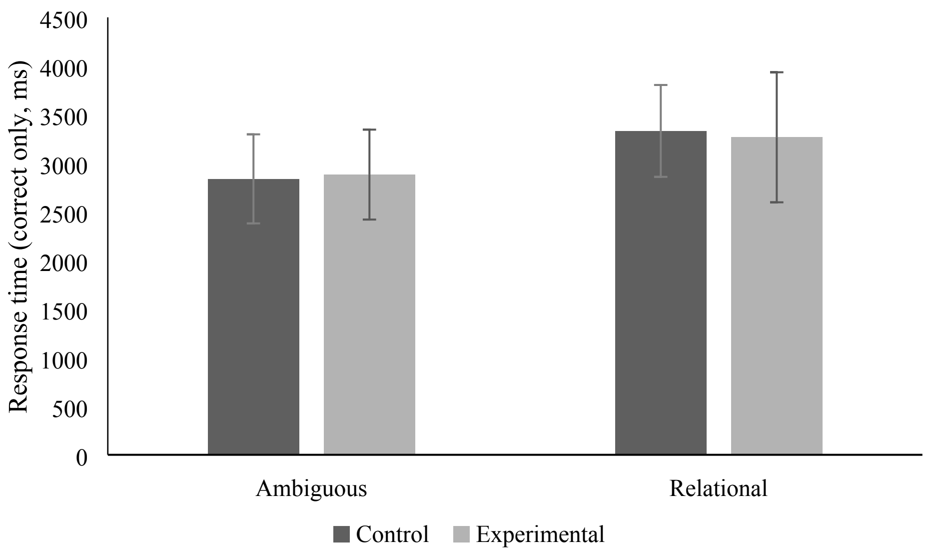


Figure A8. Mean response times (and standard error) in Director task by condition and trial type

The error rate was then examined with a series of t-tests, as no errors made in the control conditions (error data were not trimmed). A one-sample t-test (with a theoretical value of zero) showed that significantly more errors than zero were made in the experimental condition (t (141) =12.89, $p < .01$, with a mean difference of 3.94 (SD = 3.59)). A paired-sample t-test showed there was also an effect of condition in the experimental trials (t (141) = -8.88, $p < .01$), with significantly more errors in the relational condition than in the ambiguous condition.

For the ambiguous trials there were significantly more errors than zero (t (141) = 11.36, $p < .01$, with a mean difference of 1.17 (SD = 1.21)). There were significantly more errors than zero in the relational trials (t (141) = 12.12, $p < .01$, with a mean difference of 2.77 (SD = 2.68)). These differences can be seen in Figure A9.

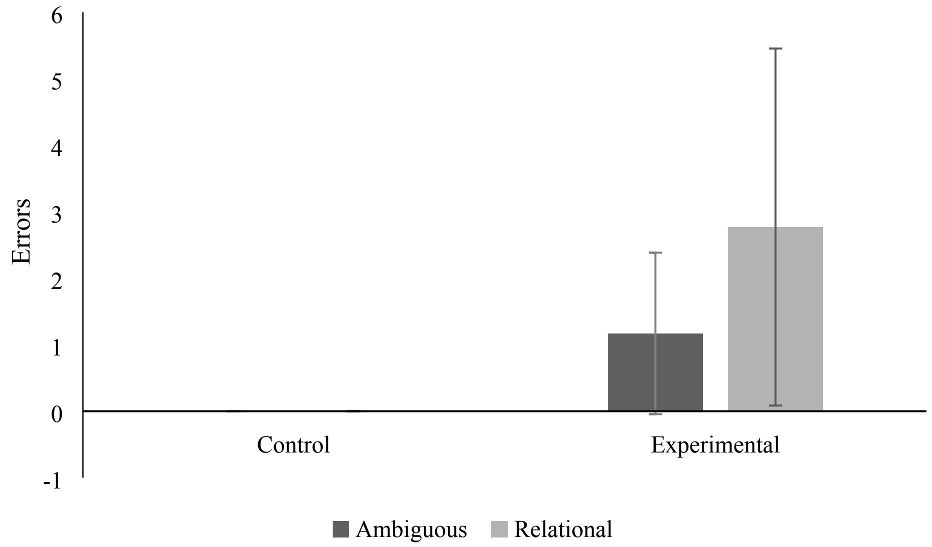


Figure A9: Mean number of errors (and standard error) in Director task by condition and trial type

The control condition was expected to have no errors and a faster response time due to there being no competing item for the item description given by the instructor. The relational trials were slower and resulted in more errors than the ambiguous trials. This may be due to there being more items that the participant is cued to in the relational trials (for example in the experimental grid where the critical instruction is 'move the small ball…', there are three different sized balls present) compared to in the ambiguous trials (where in an experimental grid, where the critical instruction is 'move the mouse…', there is a computer mouse and a mouse, so there are only two potential items). The additional item that the participant must consider in relational trials may explain the increased response times, and perhaps the increased error rate.

The critical dependent variables for the Director task were the error rates of the relational and ambiguous trials. These were expected to measure the participants' failure to use their theory of mind. Both the distributions of the dependent variables were skewed, and they were both square root transformed (ambiguous with a constant of one to remove negative values). This resulted in adequately normal distributions for both dependent variables. The reliability of the ambiguous measure was borderline satisfactory (.59), though the reliability of the relational measure was excellent (.83).

For Review Only

Appendix A4

Executive task discussion

An initial correlation matrix including the Simon task is shown in Table 1. There were no correlations between the Simon task and any other task, so this was dropped from the final analyses.

Table 1. Correlation matrix between variables

	Simon task	Shape Matching	Go NoGo	Go NoGo (picture)	Stop-Signal	Director Task (ambiguous)	Director Task (relational)	Visual Perspective (Conflict)
Shape-Matching	.07							
Go NoGo	.11	-.04	-					
Go NoGo (picture)	.03	.00	.23**	-				
Stop-Signal	.14	.15	.18*	.08	-			
Director (ambiguous)	-.09	.09	.00	.34**	.02	-		
Director (relational)	.04	.29**	.01	.25**	.05	.65**	-	
Visual Perspective (Conflict)	.12	.06	.14	.00	.24**	-.03	-.04	-
Visual Perspective (Focus)	.07	-.02	.08	-.04	.02	.02	-.03	-.18*

For Review Only

Final ~~R~~results from the path analyses showed a correlation between the Go / No-Go (picture) and Go / No-Go tasks, and between the Go / No-Go (picture) and the SST. There were no correlations with the Shape-matching task. While we predicted that the executive tasks would all intercorrelate as they all require response-distracter inhibition (as per Friedman & Miyake, 2004), the specific nature of the inhibition required by these varying tasks may explain this pattern of correlations. The Go / No-Go tasks both involve selecting relevant *responses* (i.e., to respond (go) or not (no-go) to the stimuli). Likewise the SST involves selecting the appropriate response to a given stimuli (whether to respond (go) or inhibit their response (stop) depending on whether there is a stop-signal). On the other hand, the Shape-matching task involves selection of relevant and ignoring of irrelevant *information*, whereby the participant needs to ignore the distracter shape and select their response based on the relevant target shape. This may explain why responses on shape matching were not related with the other executive measures.

The nature of the ToM tasks may also explain the relationships found with the different executive tasks. As noted, L1 VPT performance, particularly on the Conflict index, may be related to the effect of self/other consistency on selection of a response (Apperly, 2010; Qureshi et al., 2010; McCleery et al., 2011), rather than on selection of information or any integration. As such it is perhaps unsurprising that Conflict index of the L1 VPT is only related to the SST. However, the Director task requires integration of perspective and instruction, which may involve selection of relevant information and also response selection. This perspective would explain the relationship between to the Go / No-Go (picture) and Shape-matching task. The latter may be related to Relational trials of the Director task due to there being potentially three referent objects (rather than two as in ambiguous trials).

For Review Only

References

Apperly, I. A. (2010). *Mindreaders: the cognitive basis of "theory of mind"*. Taylor & Francis: Psychology Press: London

Apperly, I.A., Carroll, D.J., Samson, D., Qureshi, A., Humphreys, G.W. & Moffatt, G. (2010). Why are there limits on theory of mind use? Evidence from adults' ability to follow instructions from an ignorant speaker. *Quarterly Journal of Experimental Psychology*, *63*(6): 1201 – 1217

Barr, D. (2008). Pragmatic expectations and linguistic evidence: listeners anticipate but do not integrate common ground. *Cognition,* 109(1): 18 – 40

DeSchepper, B. & Treisman, A. (1996). Visual Memory for Novel Shapes: Implicit Coding Without Attention. *Journal of Experimental Psychology: Learning, Memory & Cognition, 22* (1): 27 – 47

Endo, N., Saiki, J. & Saito, H. (2001). Determinants of occurrence of negative priming for novel shapes with matching paradigm. *Japanese Journal of Psychology*, 72 (3), 204 – 212

Friedman, N. P. & Miyake, A. (2004). The Relations Among Inhibition and Interference Control Functions: A Latent-Variable Analysis. *Journal of Experimental Psychology:General, 133* (1): 101 – 135

Kline, R. B. (2005). *Principles and practice of structural equation modelling (2nd ed.)*. New York: Guilford Press

McCleery, J.P., Surtees, A.D.R., Graham, K.A., Richards, J.E. & Apperly, I.A. (2011). The Neural and Cognitive Time Course of Theory of Mind. *The Journal of Neuroscience,* 31(36): 12849 - 12854

Oberauer, K. (2005). Binding and inhibition in working memory: individual and age differences in short-term recognition. *Journal of Experimental Psychology: General*, 134(3): 368 – 87

Peterson, B. S., Kane, M. J., Alexander, G. M., Lacadie, C., Skudlarski, P., Leung, H-C., May, J., Gore, J. C. (2002) An event-related functional MRI study comparing interference effects in the Simon and Stroop tasks. *Cognitive Brain Research*, 13, 427–440

For Review Only

Qureshi, A. W., Apperly, I. A. & Samson, D. (2010). Executive function is necessary for perspective selection, not Level-1 visual perspective calculation: Evidence from a dual-task study of adults. *Cognition*, *117*, 230 – 236

Salthouse, T. A., Siedlecki, K. L. & Krueger, L. E. (2006). An individual differences analysis of memory control. *Journal of Memory and Language*, 55 (1), 102 – 125

Schmitt, B. M., Münte, T. F. & Kutas, M. (2000). Electrophysiological estimates of the time course of semantic and phonological encoding during implicit picture naming. *Psychophysiology, 37*: 473–484

Tabachnick, B. G. & Fidell, L. S. (2001). *Using multivariate statistics (4th ed.)*. MA: Allyn & Bacon

Tolan, G. A. & Tehan, G. (1999). Determinants of Short-term Forgetting: Decay, Retroactive Interference, or Proactive Interference. *International Journal of Psychology, 34 (5/6),* 285 – 292

van Turennout, M., Hagoort, P. & Brown, C. M. (1997). Electrophysiological Evidence on the Time Course of Semantic and Phonological Processes in Speech Production. *Journal of Experimental Psychology: Learning, Memory & Cognition*, *23* (4): 787 – 806

For Review Only

Appendix B (models prior to final)

Initial Model

The initial model shown in Figure 1 showed no relationship between the Go / No-Go or Stop-Signal tasks and either of the Director task variables. The Shape-Matching task only predicted performance on the relational variable of the Director task. The Go / No-Go (picture) task predicted performance on both variables of the Director task. The Stop-Signal Task predicted performance on the L1 VPT Conflict Index. All parameters (standardised and unstandardized path coefficients, covariances and correlations, variances and squared multiple correlations) are shown in Table 1.

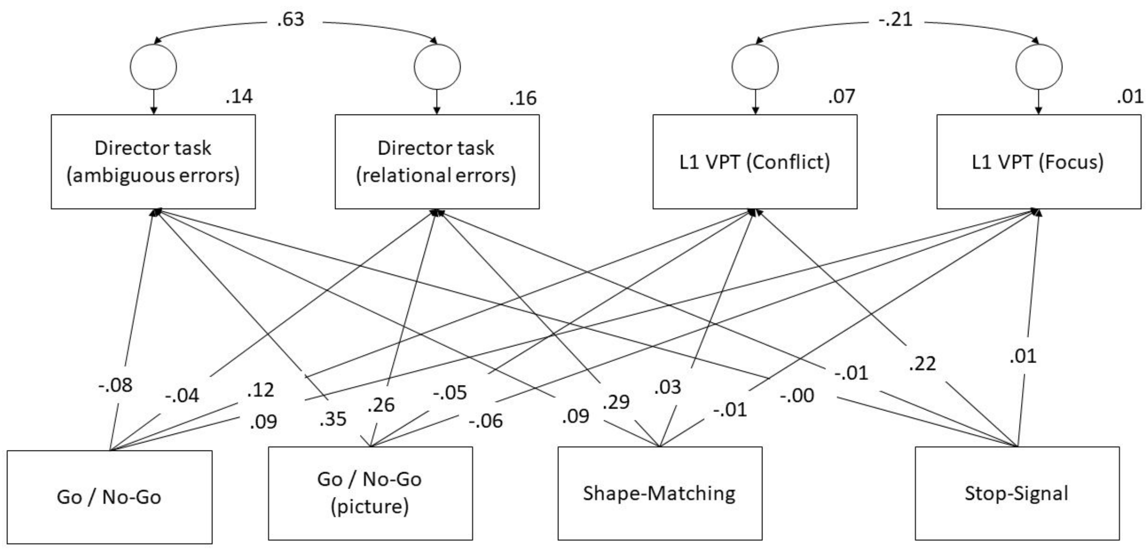


Figure 1. Initial Model (standardised coefficients)

Table 1. Parameter estimates for initial model

Path	Unstandardised Estimate (Standard Error)	Standardised Estimate
Shape-matching →Director (relational)	.16 (.04)	.29***
Go/No-Go → Director (relational)	-.05 (.10)	-.04

Go/No-Go (picture) → Director (relational)	.23 (.07)	.26***
SST → Director (relational)	-.00 (.03)	-.01
Shape-matching →Director (ambiguous)	.04 (.04)	.09
Go/No-Go → Director (ambiguous)	-.09 (.09)	-.08
Go/No-Go (picture) → Director (ambiguous)	.28 (.06)	.35***
SST → Director (ambiguous)	-.00 (.03)	-.00
Shape-matching →L1 VPT (Focus)	-.01 (.04)	-.02
Go/No-Go → L1 VPT (Focus)	.09 (.08)	.09
Go/No-Go (picture) → L1 VPT (Focus)	-.04 (.06)	-.06
SST → L1 VPT (Focus)	.00 (.03)	.01
Shape-matching → L1 VPT (Conflict)	.02 (.04)	.03
Go/No-Go → L1 VPT (Conflict)	.12 (.09)	.12
Go/No-Go (picture) → L1 VPT (Conflict)	-.03 (.06)	-.05
SST → L1 VPT (Conflict)	.07 (.03)	.22**
Covariances	**Unstandardised Estimate (Standard Error)**	**Standardised Estimate**
Director (relational) ↔ Director (ambiguous)	3.42 (.54)	.63***
L1 VPT (Conflict) ↔L1 VPT (Focus)	-.93 (.39)	-.21*
Variances	**Estimate (Standard Error)**	
Shape-Matching	23.60 (2.81)***	
Go/No-Go	4.48 (.53)***	
Go/No-Go (picture)	9.22 (1.10)***	
SST	45.20 (5.38)***	
e1 (Director (relational))	5.96 (.71)***	
e2 (Director (ambiguous))	5.02 (.60) ***	
e3 (L1 VPT (Conflict))	4.47 (.53)***	
e4 (L1 VPT (Focus))	4.66 (.56)***	
Squared Multiple Correlations		
Director (relational)	.16	
Director (ambiguous)	.14	
L1 VPT (Conflict)	.07	
L1 VPT (Focus)	.01	

*$p < .05$, **$p < .01$, ***$p < .001$

For Review Only

Table 1 shows significant correlations between the pairs of variables for each ToM task, and that the error terms and variances were all significant. Approximately 16% of the relational variable and 14% of the ambiguous variable were accounted for by this initial model. The model (**Error! Reference source not found.**) however, was not a good fit to the actual data.

Model 2

Non-significant paths were removed, resulting in the model shown in Figure 2.

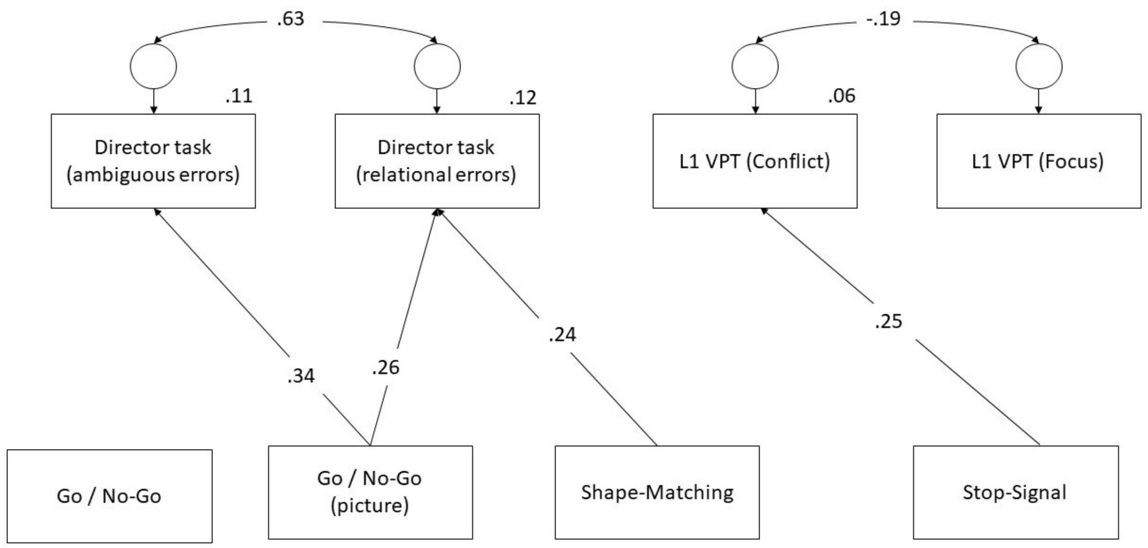


Figure 2. Model 2 (standardised coefficients)

Table 2 shows the model parameters in full. Removing the non-significant parameters did not affect the individual parameters or the model fit (**Error! Reference source not found.**), with the AIC value reducing from the initial model.

iew Only

Table 2. Parameter estimates for Model 2

Path	Unstandardised Estimate (Standard Error)	Standardised Estimate
Shape-matching →Director (relational)	.13 (.03)	.24***
Go/No-Go → Director (relational)		
Go/No-Go (picture) → Director (relational)	.22 (.07)	.26***
SST → Director (relational)		
Shape-matching →Director (ambiguous)		
Go/No-Go → Director (ambiguous)		
Go/No-Go (picture) → Director (ambiguous)	.26 (.06)	.34***
SST → Director (ambiguous)		
Shape-matching →L1 VPT (Focus)		
Go/No-Go → L1 VPT (Focus)		
Go/No-Go (picture) → L1 VPT (Focus)		
SST → L1 VPT (Focus)		
Shape-matching → L1 VPT (Conflict)		
Go/No-Go → L1 VPT (Conflict)		
Go/No-Go (picture) → L1 VPT (Conflict)		
SST → L1 VPT (Conflict)	.08 (.03)	.25**

Covariances	Unstandardised Estimate (Standard Error)	Standardised Estimate
Director (relational) ↔ Director (ambiguous)	3.54 (.56)	.63***
L1 VPT (Conflict) ↔L1 VPT (Focus)	-.89 (.40)	-.19*

Variances	Estimate (Standard Error)
Shape-Matching	23.60 (2.81)***
Go/No-Go	4.48 (.53)***
Go/No-Go (picture)	9.22 (1.10)***
SST	45.20 (5.38)***
e1 (Director (relational))	5.99 (.71)***
e2 (Director (ambiguous))	5.10 (.61)***
e3 (L1 VPT (Conflict))	4.51 (.54)***
e4 (L1 VPT (Focus))	4.73 (.56)***

Squared Multiple Correlations	
Director (relational)	.12
Director (ambiguous)	.11
L1 VPT (Conflict)	.06
L1 VPT (Focus)	.00

* $p < .05$, ** $p < .01$, *** $p < .001$

For Review Only

Appendix C (all models)

The following analyses used interference measures for self (altercentric interference from the avatar perspective) and other (egocentric interference from the self perspective) for the L1 VPT task. These were calculated based on reaction times for correct trials only and were "inconsistent perspective condition – consistent condition" for self and other trials separately.

After variance adjustment and transformations, the final descriptive statistics for the dependent and independent variables were as shown in Table 1.

Table 1. Final task descriptives

Measure	M (SD)	Variance	Reliability[1]
Shape-Matching (ms) (square root transformation)	21.56 (4.88)	23.77	.92
Go / No Go (FAR)	4.28 (2.12)	4.51	.93
Go / No Go (picture) (FAR; Bird/Mammal)[2]	6.17 (3.05)	9.28	.74
Stop-Signal (SSRT) (log 10 transformation)	16.33 (6.74)	45.52	.99
Director Task (errors) (log 10 transformation)			
Ambiguous	2.34 (2.41)	5.79	.59
Relational	2.75 (2.66)	7.07	.83
Visual Perspective measures (ms)			
Egocentric interference	4.88 (5.62)	31.63	.63
Altercentric interference	8.52 (6.33)	40.08	.43

An initial correlation matrix between the variables is shown in

Table 2. For the inhibitory control tasks, positive correlations were shown between the Shape-Matching and Stop-Signal tasks, the Go/No-Go and Go/No-Go (picture) tasks and the Go/No-Go and Stop-Signal tasks. A positive correlation was shown between the Director task variables, and a negative one between the L1 VPT variables. The Shape-Matching task had

[1] Split-half reliabilities

[2] As FAR for the bird and mammal trials were significantly correlated and followed the same pattern, they were collapsed to form a single FAR for the go / no go (picture) task

For Review Only

positive correlations with both the Director (relational) variable, while the Go/No-Go (picture) task had positive correlations with both measures of the Director task. The Stop-Signal task also had a positive correlation with the L1 VPT (Conflict) variable.

Table 2. Correlation matrix between variables

	Shape Matching	Go NoGo	Go NoGo (picture)	Stop-Signal	Director Task (ambiguous)	Director Task (relational)	Visual Perspective (Conflict)
Go NoGo	-.04	-					
Go NoGo (picture)	.00	.23**	-				
Stop-Signal	.15	.18*	.08	-			
Director (ambiguous)	.09	.00	.34**	.02	-		
Director (relational)	.29**	.01	.25**	.05	.65**	-	
Visual Perspective (Egocentric)	.19*	.10	.04	.24**	.05	.08	-
Visual Perspective Altercentric	.16	.03	.01	.14	-.12	-.07	.24**

* $p < .05$, ** $p < .01$

Models

Analyses were carried out using AMOS 25.

Initial Model

The initial model shown in Figure 1 showed no relationship between the Go / No-Go or Stop-Signal tasks and either of the Director task variables. The Shape-Matching task predicted performance on the relational variable of the Director task. The Go / No-Go (picture) task predicted performance on both variables of the Director task. The Stop-Signal Task and Shape-Matching task predicted performance egocentric interference. All parameters

For Review Only

(standardised and unstandardized path coefficients, covariances and correlations, variances and squared multiple correlations) are shown in Table 1.

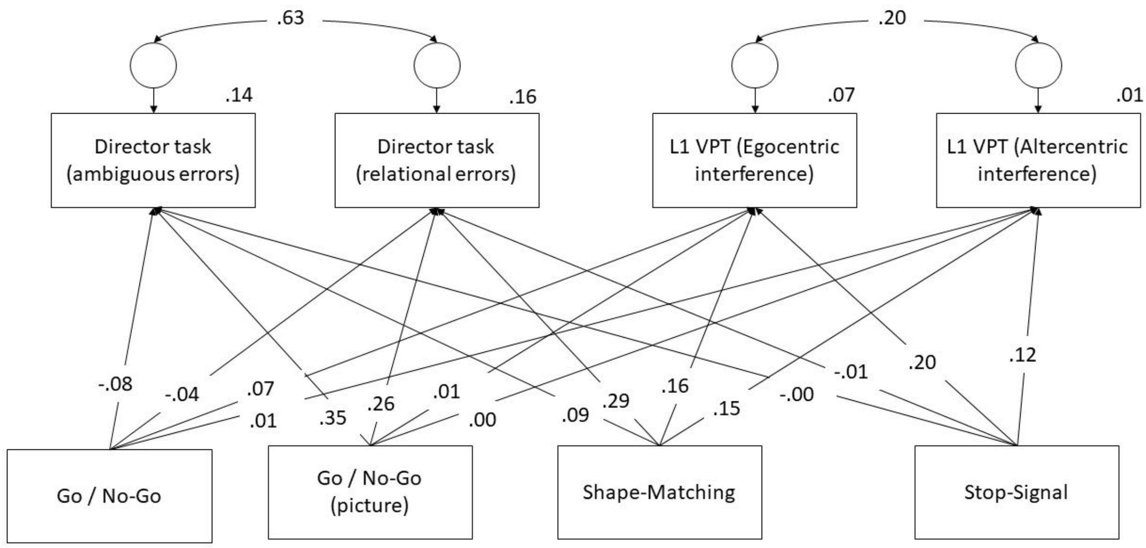


Figure 1. Initial Model (standardised coefficients)

iew Only

Table 3. Parameter estimates for initial model

Path	Unstandardised Estimate (Standard Error)	Standardised Estimate
Shape-matching →Director (relational)	.16 (.04)	.29***
Go/No-Go → Director (relational)	-.05 (.10)	-.04
Go/No-Go (picture) → Director (relational)	.23 (.07)	.26***
SST → Director (relational)	-.00 (.03)	-.01
Shape-matching →Director (ambiguous)	.04 (.04)	.09
Go/No-Go → Director (ambiguous)	-.09 (.09)	-.08
Go/No-Go (picture) → Director (ambiguous)	.28 (.06)	.35***
SST → Director (ambiguous)	-.00 (.03)	-.00
Shape-matching →L1 VPT (Altercentric interference)	.17 (.10)	.15
Go/No-Go → L1 VPT (Altercentric interference)	.04 (.22)	.01
Go/No-Go (picture) → L1 VPT (Altercentric interference)	.00 (.15)	.00
SST → L1 VPT (Altercentric interference)	.10 (.07)	.12

Shape-matching → L1 VPT (Egocentric interference)	.21 (.11)	.16*
Go/No-Go → L1 VPT (Egocentric interference)	.20 (.24)	.07
Go/No-Go (picture) → L1 VPT (Egocentric interference)	.02 (.17)	.01
SST → L1 VPT (Egocentric interference)	.19 (.08)	.20*
Covariances	**Unstandardised Estimate (Standard Error)**	**Standardised Estimate**
Director (relational) ↔ Director (ambiguous)	3.42 (.54)	.63***
L1 VPT (Egocentric interference) ↔L1 VPT (Altercentric interference)	6.57 (.2.84)	.20*
Variances	**Estimate (Standard Error)**	
Shape-Matching	23.60 (2.81)***	
Go/No-Go	4.48 (.53)***	
Go/No-Go (picture)	9.22 (1.10)***	
SST	45.20 (5.38)***	
e1 (Director (relational))	5.96 (.71)***	
e2 (Director (ambiguous))	5.02 (.60) ***	
e3 (L1 VPT (Egocentric interference))	30.11 (3.59)***	
e4 (L1 VPT (Altercentric interference))	36.43 (4.34)***	
Squared Multiple Correlations		
Director (relational)	.16	
Director (ambiguous)	.14	
L1 VPT (Egocentric interference)	.07	
L1 VPT (Altercentric interference)	.04	

*p < .05, **p < .01, ***p < .001

For Review Only

Table 3 shows significant correlations between the pairs of variables for each ToM task, and that the error terms and variances were all significant. Approximately 16% of the relational variable and 14% of the ambiguous variable were accounted for by this initial model. The model (see Table 6) however, was not a good fit to the actual data.

Model 2

Non-significant paths were removed, resulting in the model shown in Figure 2. Doing this did not significantly affect the model fit (X^2 (11) = 8.85, *p* = .63).

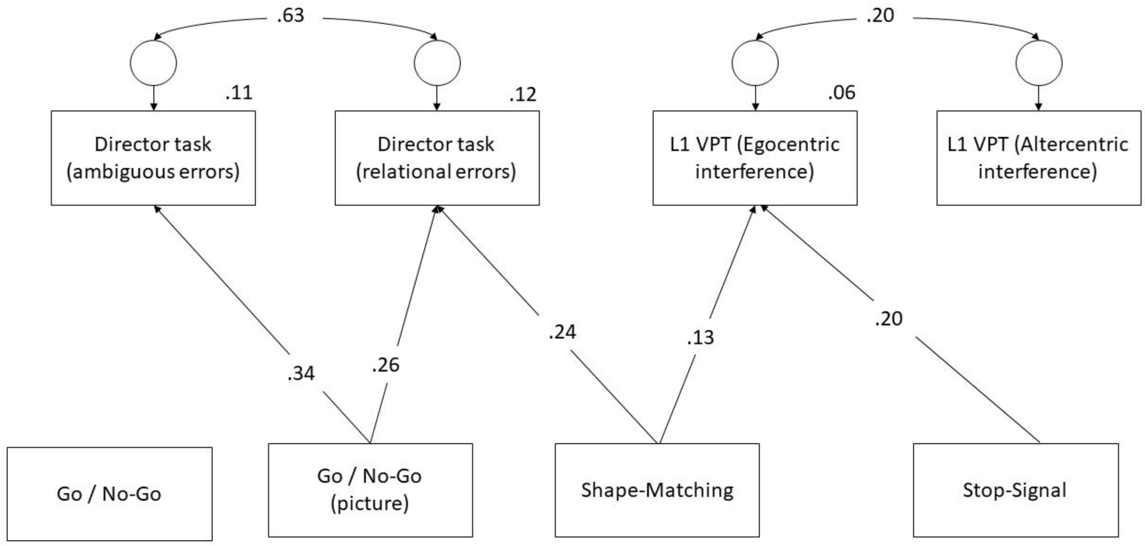


Figure 2. Model 2 (standardised coefficients)

Only

Table 2 shows the model parameters in full. Removing the non-significant parameters did not affect the individual parameters or the model fit (Table 6), with the AIC value reducing from the initial model.

Table 4. Parameter estimates for Model 2

Path	Unstandardised Estimate (Standard Error)	Standardised Estimate
Shape-matching →Director (relational)	.13 (.03)	.24***
Go/No-Go → Director (relational)		
Go/No-Go (picture) → Director (relational)	.22 (.07)	.26***
SST → Director (relational)		

Shape-matching →Director (ambiguous)		
Go/No-Go → Director (ambiguous)		
Go/No-Go (picture) → Director (ambiguous)	.26 (.06)	.34***
SST → Director (ambiguous)		
Shape-matching →L1 VPT (Altercentric interference)		
Go/No-Go → L1 VPT (Altercentric interference)		
Go/No-Go (picture) → L1 VPT (Altercentric interference)		
SST → L1 VPT (Altercentric interference)		
Shape-matching → L1 VPT (Egocentric interference)	.16 (.10)	.13
Go/No-Go → L1 VPT (Egocentric interference)		
Go/No-Go (picture) → L1 VPT (Egocentric interference)		
SST → L1 VPT (Egocentric interference)	.18 (.07)	.19*

Covariances	**Unstandardised Estimate (Standard Error)**	**Standardised Estimate**
Director (relational) ↔ Director (ambiguous)	3.47 (.55)	.63***
L1 VPT (Egocentric interference) ↔L1 VPT (Altercentric interference)	6.89 (2.92)	.20*

Variances	**Estimate (Standard Error)**
Shape-Matching	23.60 (2.81)***
Go/No-Go	4.48 (.53)***
Go/No-Go (picture)	9.22 (1.10)***
SST	45.20 (5.38)***
e1 (Director (relational))	5.99 (.71)***
e2 (Director (ambiguous))	5.10 (.61)***
e3 (L1 VPT (Egocentric interference))	31.40 (3.74)***
e4 (L1 VPT (Altercentric interference))	36.68 (4.37)***

Squared Multiple Correlations	
Director (relational)	.12
Director (ambiguous)	.11
L1 VPT (Egocentric interference)	.05
L1 VPT (Altercentric interference)	.00

* $p < .05$, ** $p < .01$, *** $p < .001$

For Review Only

Model 3 (Final)

The non-significant path between the Shape-Matching task and egocentric interference was removed, and correlations added between the Go / No-Go (picture) and Go / No-Go tasks and also between the Go / No-Go and Stop-Signal tasks (as per modification indices), resulting in the final model in Figure 3. The model fit (as per AIC values) was improved.

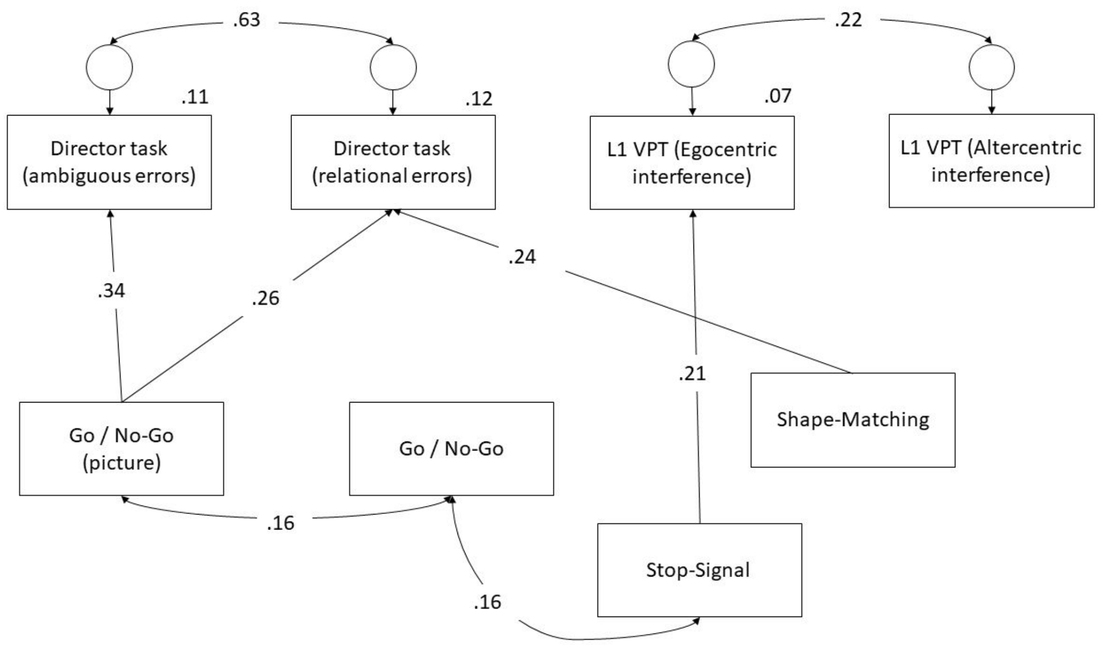


Figure 3. Final model (standardised coefficients)

Table 5 shows the model parameters in full.

Table 5. Parameter estimates for Model 3 (Final)

Path	Unstandardised Estimate (Standard Error)	Standardised Estimate
Shape-matching →Director (relational)	.13 (.03)	.24***
Go/No-Go → Director (relational)		
Go/No-Go (picture) → Director (relational)	.22 (.07)	.26***
SST → Director (relational)		

Shape-matching →Director (ambiguous)		
Go/No-Go → Director (ambiguous)		
Go/No-Go (picture) → Director (ambiguous)	.26 (.06)	.34***
SST → Director (ambiguous)		
Shape-matching →L1 VPT (Altercentric interference)		
Go/No-Go → L1 VPT (Altercentric interference)		
Go/No-Go (picture) → L1 VPT (Altercentric interference)		
SST → L1 VPT (Altercentric interference)		
Shape-matching → L1 VPT (Egocentric interference)		
Go/No-Go → L1 VPT (Egocentric interference)		
Go/No-Go (picture) → L1 VPT (Egocentric interference)		
SST → L1 VPT (Egocentric interference)	.20 (.08)	.21**

Covariances	**Unstandardised Estimate (Standard Error)**	**Standardised Estimate**
Director (relational) ↔ Director (ambiguous)	3.47 (.55)	.63***
L1 VPT (Egocentric interference) ↔L1 VPT (Altercentric interference)	7.52 (2.96)	.22*
Go/No-Go ↔ SST	2.29 (1.18)	.16*
Go/No-Go ↔ Go/No-Go (picture)	1.38 (.55)	.22*

Variances	**Estimate (Standard Error)**
Shape-Matching	23.60 (2.81)***
Go/No-Go	4.48 (.53)***
Go/No-Go (picture)	9.22 (1.10)***
SST	45.20 (5.38)***
e1 (Director (relational))	5.99 (.71)***
e2 (Director (ambiguous))	5.10 (.61)***
e3 (L1 VPT (Egocentric interference))	31.40 (3.74)***
e4 (L1 VPT (Altercentric interference))	37.57 (4.47)***

Squared Multiple Correlations	
Director (relational)	.12
Director (ambiguous)	.11
L1 VPT (Egocentric interference)	.04
L1 VPT (Altercentric interference)	.00

* $p < .05$, ** $p < .01$, *** $p < .001$

Model fit values are shown in Table 6.

For Review Only

Table 6. Model fit parameters

	CMIN	df	*p*	NFI	CFI	AIC	RMSEA	Low	High	Bollen-Stine Bootstrap *p*
Model 1	20.32	10	<.05	.87	.92	72.32	.09	.03	.14	.06
Model 2	29.17	21	.11	.81	.94	59.17	.05	.00	.10	.18
Model 3	20.16	20	.45	.87	.99	52.16	.01	.00	.07	.52

For Review Only