Summer 2023

Individual Differences in Perspective-Taking During Language Comprehension

Kanan Benjamin Luce

INDIVIDUAL DIFFERENCES IN PERSPECTIVE-TAKING DURING LANGUAGE COMPREHENSION

by

Kanan Benjamin Luce

Bachelor of Arts
University of California, Berkeley, 2018

Submitted in Partial Fulfillment of the Requirements

For the Degree of Doctor of Philosophy in

Linguistics

College of Arts and Sciences

University of South Carolina

2023

Accepted by:

Amit Almor, Major Professor

Anne Bezuidenhout, Committee Member

Brett Sherman, Committee Member

Jessica Klusek, Committee Member

Anne Vail, Dean of the Graduate School

ABSTRACT

Previous research has found mixed results for a link between executive function and perspective-taking. One proposed reason for this is that perspective-taking during comprehension tasks may not be internally reliable. This dissertation presents two large individual differences experiments with multiple perspective-taking during comprehension and executive function measures. We aimed to see (1) whether people were consistent in their ability to take perspectives during comprehension, (2) if the typical tasks used in this area were reliable, and (3) whether executive function measures predicted perspective-taking ability. We found a lack of reliability of some of the most commonly used perspective-taking during comprehension tasks, and no evidence that inhibition control or working memory predict perspective-taking ability. Sparse correlations between the perspective-taking tasks lead us to propose that perspective-taking during comprehension should be considered in a more granular way.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

CI..Confidence Interval

E1..Experiment 1

E2..Experiment 2

RT..Response Time

ToM..Theory of Mind

VPT..Visual Perspective-Taking

WCS..Wisconsin Card Sort

CHAPTER 1

PERSPECTIVE-TAKING

I begin with a review of the previous literature on perspective-taking in comprehension, including comparison to the related psychological constructs of theory of mind and visual perspective-taking, different spatial and information mismatch (non-spatial) perspective-taking methodologies, the effortful nature of perspective-taking, and the difficulties that come with predicting perspective-taking ability.

PERSPECTIVE-TAKING IN COMPREHENSION

Using language successfully in conversation requires maintaining some model of what your interlocutors know. Asking questions requires us to assume our interlocutor has some privileged knowledge that we lack, referring expressions require mutual knowledge in order to be sure that we are referring to the same object, and accomplishing a joint task requires us to adopt partner-specific entrained phrases and schemas.

This dissertation will mainly focus on perspective-taking as it is required for comprehending referring expressions. In order to accurately comprehend referring expressions that are being produced for us, we require some model of what is and is not mutually known. Consider the following referring expressions:

1. Meet me at the Empire State Building.
2. That screwdriver is dangerous.
3. Hand me the green screwdriver.
4. Hand me the screwdriver to the right.

Sentence (1) assumes that we have a shared cultural knowledge of that famous building. (2) assumes that there is some salient screwdriver in the context, either from a previous linguistic mention or in the physical context, perhaps based on simultaneous pointing to the target screwdriver. Sentence (3) contrasts one screwdriver (the green one) from other screwdrivers that are mutually known or visible to the interlocutors. (4) requires the comprehender to visualize how the scene looks from the producer's point of view, which is not a difference in what is mutually known to the interlocutors, but how perspective differences structure that information differently.

Failing to consider the interlocutor's perspective in these cases leads to the referring expression failing. If one of us lacks the knowledge of what the Empire State Building is, we may end up waiting forever for the other to show up. In (2), if we do not have a mutual understanding of which screwdriver is dangerous, one of us may end up hurt. In (3) if there are multiple green screwdrivers visible, then this sentence would fail to disambiguate in the context. In (4), if the comprehender did not take the perspective of the producer, they may end up passing the wrong screwdriver, the one on their right rather than the one on the producer’s right.

This dissertation is concerned with how comprehenders take perspective-taking into account when interpreting referring expressions, whether perspective-taking ability is really one ability or a constellation of related processes, and what cognitive resources are recruited in order to complete this complex process.

LINGUISTIC PERSPECTIVE-TAKING, THEORY OF MIND, AND VISUAL PERSPECTIVE-TAKING

Linguistic perspective-taking is closely related to the psychological constructs of theory of mind and visual perspective-taking. In this section, I will briefly explain each of these in turn as well as some current research into the relations between these three abilities.

Theory of Mind (ToM)

Theory of mind (ToM) is the ability to attribute mental states (including desires, beliefs, and emotions) to another person. The study of ToM is often focused on beliefs, and the belief aspect of ToM is split into two levels. Level 1 theory of mind has to do with different beliefs and the presence or absence of belief, whereas level 2 theory of mind has to do with false belief; that people may be incorrect in their beliefs.

Level 2 theory of mind is often tested using the classic Sally-Anne test, in which one person (Sally) hides an object and leaves the scene. A different person (Anne) then moves the object to another hiding spot, and then the hider returns. The task is to select where the original hider (Sally) will look for the object, as they have the false belief that the object should still reside where they originally hid it.

Aspects of ToM may actually be several distinct but related abilities. Wellman and Liu (2004) provide a more granular concept of theory of mind, in which several ToM abilities are learned in sequential order in child development. They find that children come to understand others' desires before beliefs, and that people have diverse beliefs before they understand that people have false beliefs, and finally that real and apparent emotions may differ.

Visual Perspective-Taking (VPT)

Visual perspective-taking is the ability to know whether and how an object is seen by another person. Visual perspective-taking differs from theory of mind in that it is explicitly spatial, and seems to require general spatial resources. VPT, like theory of mind, can also be divided into level 1 and 2. Level 1 VPT has to do with deciding whether or not an object is visible to another person, whereas level 2 VPT involves figuring out how an object looks from another's viewpoint.

These two levels of VPT seem to be dissociable abilities. Michelon and Zacks (2006) suggest that different processes are performed in level 1 and level 2 VPT. In four experiments, they provide evidence that level 2 VPT requires a perspective transformation, where you spatially rotate your viewpoint to match the target, while level 1 VPT only requires tracing the line of sight of the target. So like ToM, there seems to be a constellation of abilities that take the heading VPT.

Further studies have suggested that level 1 VPT can be broken down into two parts: calculation and selection of the visual perspective. Qureshi et al. (2010) and Samson et al. (2010) provide evidence that in VPT tasks, the calculation of perspectives (both for self and others) is autonomous and relatively effortless, while the selection of perspective is effortful. Qureshi et al. did a dual level 1 VPT/executive function task and found that the executive function task interfered with VPT trials that had inconsistent perspectives both when participants were asked to select an answer based on their perspective or another's perspective. They take this as showing that participants were calculating the other person's perspective even when it was not necessary, and that the processing cost came from selecting between the inconsistent perspectives.

Samson et al. (2010) similarly suggest that people automatically calculate the perspective of an avatar even when that perspective is not necessary to complete the task. These findings contrast the usual suggestion that VPT and ToM are effortful and require executive function resources, by suggesting that only the selection of perspective and not calculation of perspective is effortful.

Furthermore, VPT is distinct from object rotations, as evidenced by Hegarty and Waller (2004) and Huttenlocher and Preston (1973, 1979). The spatial perspective rotations seem to rely on embodied transformations of the self, while object rotation does not (Kessler & Thomson, 2010). Level 2 VPT and object mental rotation also show differing activation of brain structures. While both recruit areas associated with visuospatial reasoning, level 2 VPT also recruits areas associated with social cognition, most importantly the temporoparietal junction (Gunia et al., 2021).

Linguistic Perspective-Taking

Linguistic perspective-taking deals in terms of common and privileged ground. The term common ground was formalized by Stalnaker (1978) and refers to the information that is shared between interlocutors and updated through linguistic contributions. Privileged ground, on the other hand, refers to the private information known to only one interlocutor or the other. Theories of linguistic perspective-taking rely on the notion of common ground and the fact that people must maintain some kind of representation of what is mutually known in order to communicate. This is done by using common ground in order to restrict the domain of interpretation. Although traditionally the focus was on common ground, more recent work has also given more attention to

how privileged ground plays a role in communication (Brown-Schmidt, 2009; Heller et al., 2016).

The most influential account taking this notion of common ground into psycholinguistics comes from Clark and colleagues (Clark & Marshall, 1978, 1981; Clark, 1996). This account has several parts.

The first claim is that common ground is built through heuristic assumptions. This is a required step, as to build mutual knowledge would otherwise seem to logically require an infinitely recursive stack of checking what the other person knows. That is to say that in order for me to refer to something I have to know that you know that I know that you know… (etc.) what it is I am referring to. In order to get around this we can appeal to heuristics like assuming that because both you and your interlocutor can see an object that that object must be mutually known. This allows us a shortcut to build knowledge in a psychologically plausible way that can be processed relatively quickly.

The second aspect is that common ground can be based on information that is linguistically, physically, or culturally co-present. A linguistic assertion can update the common ground in the Stalnaker way by making some information co-present and mutually known. However, that is not the only way to update the common ground. The objects in the room that we are speaking in would also be physically co-present to us and included into the common ground even if they have yet to be mentioned in the dialogue. Likewise, speakers likely have some cultural common ground that can be assumed based on shared community membership (like Americans knowing who the American president is).

The third claim is that we have a specialized explicit representation of the common ground from which we base our perspective-taking, as opposed to other theories where common ground arises out of other processes. This takes the form of a “reference diary” that tracks events where common ground evidence is established with other people. Once common ground is established, it is relative to that specific pair (or group) of interlocutors and is retained for future use with those interlocutors.

An alternative explanation of common ground is that it arises from ordinary memory processes rather than a specific common ground representation (Horton & Gerrig, 2005a, 2016). In this view, ordinary episodic memory retrieval is all that is needed in order to get the evidence of co-presence with an interlocutor that is required for common ground. You can use your interlocutor as a retrieval clue to look back and remember that they were at a certain event, then you can use that as common ground in language production and comprehension. People who you have a lot of interaction with will also have more robust memory association and therefore more common ground.

Horton and Gerrig divide the process of audience design into two parts: commonality assessment and message formation. Commonality assessment being the process of determining whether some information is shared with your interlocutor, and message formation being deciding what the best words are given the shared information (what referring expression will best direct your interlocutor to the right referent). Both of these processes seem to be automatic at times and strategic and effortful at times. The automatic aspect is explained through appealing to memory resonance, which allows fast and automatic probe of memory, while the strategic aspect requires a dedicated, effortful search of memory.

This view potentially allows us to get the benefits of the reference diary proposed by Clark, but in a domain-general way without proposing a special common ground representation. Horton and Gerrig further argue that the reference diary would be too computationally taxing to be plausible for use in language.

Another alternate theory of linguistic perspective-taking is Pickering and Garrod's (2004) Interactive Alignment model. In this approach, interlocutors become aligned with one another at every linguistic level through priming. Pickering and Garrod distinguish between two types of common ground; implicit common ground, which arises automatically through priming and alignment, and regular common ground, where you explicitly model your interlocutor's mental state, which is a costly process that is only employed if there is a failure of alignment that needs to be repaired. They argue that much of common ground is implicit, and arises from priming at the lower linguistic levels that eventually causes interlocutors to become aligned at the dialogue and situation model levels. Because this process is automatic and relatively effortless, it explains how we so readily make use of common ground in everyday conversation.

Given these three theories, it is clear that the mechanism underlying the phenomenon of perspective-taking is still under debate. It could be a specialized reference diary, emergent from ordinary memory, or due to alignment. The fact that there is no settled explanation is relevant to issues later in the dissertation, where it is not clear which (or whether) cognitive resources predict perspective-taking ability, or even whether perspective-taking is domain-general between production and comprehension (Ryskin et al., 2015).

Linguistic Perspective-Taking in relation to ToM, and VPT

In the linguistics literature, perspective-taking is taken to be the application of theory of mind by adults to language production and comprehension. One of the main findings in the psycholinguistics literature is that perspective information is not always used and seems to be effortful (Keysar et al., 2003). This contrasts with typical psychological research, which focuses on the development of theory of mind in children, as adults perform at ceiling with the usual false belief tasks. It seems that even though adults have a robust theory of mind, they do not always employ this ability during language processing.

Because of this, linguistic perspective taking tends to focus on level 1 theory of mind, where we look only at the absence of knowledge that our interlocutor might have, rather than on level 2 theory of mind (Brown-Schmidt & Heller, 2018). This can be seen in the most prototypical linguistic perspective-taking task, the cubbies referential communication task (Keysar et al., 2000; Hanna et al., 2003; Heller et al., 2008). In this task, objects are placed between the two participants, with some objects visible to both (in the common ground) and some visible to only one participant or the other (in their privileged ground). These studies then see how the ground status of objects will affect how referring expressions are produced and comprehended. This task has shown that listeners use perspective information from initial stages of processing to disambiguate temporally ambiguous referring expressions, based on their eye movements (Heller et al., 2008).

However not all studies use these level 1 theory of mind tasks to study perspective-taking. One potential reason for inconsistent results in the literature is that

many linguistic perspective-taking tasks also make use of VPT abilities, as they focus not on common or privileged information, but on how an object looks to an interlocutor. These tasks rely on participants' ability to produce and comprehend utterances that rely on spatial VPT transformations, rather than ToM in the traditional sense.

For example, many studies include the map task, where two participants are given maps that differ in orientation or which landmarks are included, and one participant has to give directions to the other without knowing exactly what their map looks like (Brown, 1995; Ryskin et al., 2014). In this task, if the orientation is the only thing different, then this perspective-taking is not about absence of knowledge, but it is more akin to visual perspective-taking, as it requires some kind of spatial orientation. Other tasks have similar conceits, where the only perspective manipulation is spatial, such Ryskin et al. (2015)'s experiment 2, where the speaker's and listener's viewpoints of a grids are rotated 180 degrees, or Greenberg et al. (2013), which is a visual perspective-taking task without any major linguistic component.

Even when including spatial perspective-taking tasks, these linguistics studies typically make no theoretical distinction between absence-of-belief theory of mind tasks, and spatial visual perspective-taking tasks. Both are often implicitly considered under the unified heading of perspective-taking.

However, I argue that this is a mistake and that we need to, at least, carefully distinguish between spatial and information-mismatch perspective-taking. As mentioned above, theory of mind can be considered a multifaceted rather than unitary ability, and is probably distinct from visual perspective-taking. So linguistics perspective-taking should consider how the details of our perspective-taking tasks may tap into both different

aspects of theory of mind on the one hand, and spatial ability and visual perspective-taking on the other hand. Most importantly this means distinguishing spatial from information-mismatch perspective-taking. But it could also mean differentiating between different scaling components of theory of mind (absence of belief vs false belief, first and second order false belief etc).

PERSPECTIVE-TAKING METHODS: SPATIAL AND INFORMATION-MISMATCH

There have been a variety of methods used to measure perspective-taking ability in comprehension. This section will review some of these tasks in detail, and divide them into spatial and information-mismatch (non-spatial) categories.

The most prototypical information-mismatch perspective-taking task is the referential communication “cubbies” task (Keysar et al., 2000; Hanna et al., 2003; Heller et al., 2008). In this task, participants see a grid, usually 3x3, of cubbyholes, with an interlocutor on the other side of the cubbies. The cubbies contain an array of objects, and can be open, such that the objects are visible to both the participant and the interlocutor on either side of the cubbies, or they can be obscured either to the participant, or to the interlocutor. If the cubbies are open, then the object in that cubby will be in the common ground between the two partners. If the cubby is obscured to either the participant or the interlocutor, then the object in that cubby will be in the privileged ground of one of partners. In this way, the scene can be manipulated to change which objects are visually co-present and in the common ground, and which objects are privileged and known by only one partner.

The task requires producing or comprehending a referring expression, given the context of the common and privileged objects in the display. In the comprehension version of this task, the interlocutor will produce a referring expression, and the participant must select a referent based on what the interlocutor said. For example, the interlocutor will say “Pick up the big hammer”, and the participant has to decide which hammer is being referred to.

In critical trials, there will be three items of the same type as the target, a small hammer, a medium, hammer and a large hammer, as well as some other distractor items. In the baseline condition, all of the relevant items are in the common ground cubbies, so that when the instruction “pick up the big hammer” is given, the largest of the three hammers should be the target that the comprehender selects. In the perspective-mismatch condition, however, the big hammer is privileged so that only the comprehender and not the producer can see it. So therefore, when the instruction to pick up the big hammer is given, the comprehender should pick up the medium hammer, as that is the biggest hammer in the common ground that the producer could see.

A slightly different version of the task uses two sets of two objects (a large/small hammer and a large/small comb), with one set of objects both in the common ground, and the large item of the other set privileged to the comprehender (Heller et al., 2008’s two contrast trials). This way, when the instruction “pick up the large X” is given, the sentence should be disambiguated at the word “large” if the comprehender is considering the producer’s perspective, as from their point of view there is only one large object in the common ground. If the comprehender is egocentric, then they will wait until the whole sentence is produced instead of selecting the referent after hearing “large”.

This task is typically performed while tracking the eye movements of the comprehender, and the dependent variable of interest is the proportion of looks to the target (the medium hammer) over the egocentric competitor (the large hammer) in the perspective-mismatch condition. However, eye tracking is not required to see an effect of perspective in this task, as there are also differences in accuracy, with participants regularly reaching for or picking up the egocentric competitor in the perspective-mismatch condition (Keysar et al., 2000; Heller et al., 2008).

Another commonly used task, this time with a spatial component, is the map task. In this task, one participant provides directions to another participant based on individualized maps. The director and follower each have their own maps, and the information on the maps may differ. In one version of the task (Brown, 1995) the director has a more detailed map and a marked route, which they must guide the follower through on their less detailed map. In another version of the task (E1 in Ryskin et al., 2014) the director and follower have the same map, but the orientation of the map is manipulated so that in one condition the maps match, and in another, the follower’s map is rotated 180 degrees. In both versions of the task, the errors made in the follower's route are measured. The idea of both tasks is that participants must work together to adopt a shared perspective in order to accurately convey the information of the route from the director to the follower.

Another spatial task, the “table” task, (Schober, 1995; Ryskin et al., 2016) has a director and matcher looking at a circular table with a variety of objects on it. The director gives the matcher directions about which objects to pick up based on the relative location of the objects (e.g. “pick up the red triangle to the right of the purple square”).

The degree of perspective match/mismatch was manipulated such that sometimes the director and matcher share the same perspective of the table, and sometimes their perspectives are mismatched via various rotations around the table. Another more straightforward version of this task uses a rectangular grid of objects, and only has two conditions of perspective-match or 180-degree perspective-mismatch (E2 in Ryskin et al., 2014).

Part of the problem with these three tasks is that they seem to be measuring different things, even though they all fall under the heading of "perspective-taking". The spatial tasks seem to rely on a type of level-2 VPT, where the participant is required to mentally rotate themselves to take the producer's perspective in order to figure out how the series of objects is structured from their point-of-view. Only then can they calculate which way is left or right from the others' perspective.

In contrast, the "cubbies" task does not require level-2 VPT. It may involve level-1 VPT, with participants deciding whether or not the producer can see the privileged objects, but it seems to rely more on a general theory of mind. In order to complete the task in the "non-egocentric" way, the participant must recognize that the producer does not have the same knowledge as them as to the contents of the privileged cubbies. There seems to be a fundamental difference between these spatial and information-mismatch perspective-taking tasks that needs to be carefully distinguished, and which will be explored further in the dissertation.

PERSPECTIVE-TAKING IS EFFORTFUL

One of the key findings in the perspective-taking literature is that perspective-taking is not always employed and seems to be effortful. This is despite the fact that

perspective-taking is often required in order to produce and comprehend sentences accurately and informatively.

Early evidence for this comes from Horton and Keysar (1996) who used a referential communication task to look at perspective-taking in production. They found that when participants were not pressed to respond quickly, they were able to distinguish between contextual objects that were in the common ground and those that were in the privileged ground, and incorporate that information correctly into their utterances for their listener (as measured by their adjective usage). However, when participants were required to respond quickly, they failed to distinguish common and privileged objects when producing their utterances. Horton and Keysar interpret this to mean that perspective-taking follows a "monitoring and adjustment" model, where grounding information is only incorporated only after the initial planning of an utterance, when we realize that our initial utterance may lead to a misunderstanding and needs to be revised. This is as opposed to an "initial design" model, where grounding information is available from the earliest stages of planning and utterance. Because speed pressure interfered with participants' use of grounding, their results point toward the "monitoring and adjustment" model, where the effortful incorporation of perspective information is left out when there is not enough time for processing.

Keysar et al. (2000) extended this finding into perspective-taking in comprehension by playing a referential communication game while tracking the eye-movements of the comprehender. They found that comprehenders regularly fixated on objects that were privileged to them when listening to instructions from an interlocutor, even though the interlocutor could not be referring to objects that they did not know

about. The authors propose that in comprehension as well, we utilize an egocentric heuristic first, and only incorporate perspective when it becomes necessary. The reasoning for this is that perspective-taking is an effortful and slower process that is not always required. Therefore, it benefits us to default to egocentrism, which is quicker to calculate.

More recent research has disfavored the idea that perspective-taking follows an egocentric-first approach, instead opting for a constraint-based approach where perspective information is incorporated immediately. The first study using this approach was Hanna et al. (2003). This study, like Keysar et al. (2000) used eye-tracking to measure perspective-taking in comprehension in a referential communication task. However, they tweaked the objects in the display to get rid of a potential confound (that the privileged object was the best perceptual match in one condition but not the other) and also reinforced the common ground through linguistic cues. They found that while participants' eye movements showed interference from a privileged object, they still preferred the common ground object from the earliest moments of processing. This suggests that there is not an egocentric-first heuristic that is only monitored and controlled later, but that grounding information is taken into account right away. Furthermore, the results of Hanna et al. and Keysar et al. can be reconciled by adopting a constraint-based model of perspective-taking, where common ground is one of many constraints that may be weighed in the context. In the Hanna experiment, the context more strongly favored the use of common ground information than in the Keysar experiment.

This line of constraint-based perspective-taking research has since been expanded by Brown-Schmidt, Heller, and others, and has shown that (among other things): comprehenders use perspective from initial processing to prefer considering privileged objects in response to wh-questions (Brown-Schmidt et al., 2008); comprehenders use perspective to make predictions about temporally ambiguous utterances (Heller et al., 2008); and common ground consists of a rich gradient representation rather than a binary presence/absence (Brown-Schmidt, 2012).

The constraint-based model of perspective-taking was formalized and expanded by Heller et al. (2016). They created a Bayesian model for perspective-taking in comprehending referring expressions like those used in Keysar et al. (2000) and Heller et al. (2008). The model incorporates both egocentric and common ground perspectives, as both appear to play a role, rather than only one being available. They also incorporate the idea that referring expressions will change based on the other objects in the domain (unlike Hanna et al.) and importantly, they include both an egocentric and a common ground domain and assume that both are considered, not just one. A coefficient alpha determines the relative weight given to the egocentric domain over the common ground domain.

While there has been disagreement about whether people use an egocentric-first heuristic, or incorporate perspective from initial processing in a constraint-based model, the important point for this dissertation is that there is agreement that perspective-taking appears to be effortful and not always considered. The egocentric-first model of course incorporates this fact due to the nature of its design, which gives the egocentric heuristic the place of a fast automatic process and perspective considerations the role of a monitor

and reanalysis. The constraint-based approach also recognizes that the presence of a privileged object interferes with calculating the correct referent (because we get eye movements to privileged objects even when they are not available referents). And that the extent to which perspective information is considered will depend on how strongly the context supports that particular constraint.

Furthermore, we need not even rely on eye movements to see that perspective taking is difficult. In both Keysar et al. (2000) and Heller et al. (2008), participants regularly reached for privileged objects in response to hearing a referring expression such as "pass me the big screwdriver" (24% and 17% of the time respectively). This shows a relatively severe lack of consideration of perspective in interpreting these referring expressions, much more than eye fixations to a particular referent.

In sum, perspective-taking is effortful. And that effortfulness means that it is not always employed, despite its necessity.

PREDICTING PERSPECTIVE-TAKING

One question that arises is whether some people are better or worse perspective-takers and if perspective-taking ability is predicted by lower-level cognitive resources like inhibition control and working memory. Previous research has suggested that having greater inhibition control and working memory, both components of executive function, may facilitate some people to utilize perspective information more quickly and more often during language production and comprehension. Inhibition control is thought to be important because perspective-taking often requires a conflict between multiple perspectives, one of which will need to be inhibited. And working memory because it

would allow someone the ability to manipulate, compare, or maintain multiple perspectives or potential referents.

Brown-Schmidt (2009) provided early evidence for the link between perspective-taking in comprehension and inhibition control. They found that a verbal measure of inhibition control (Stroop task) predicted perspective use in comprehension during a communication reference task. In the task, people with higher inhibition control scores showed a greater proportion of fixations to the target object in the array than to the competitor object (which was privileged), which equated to perspective-taking ability. A non-verbal inhibition control task did not significantly predict perspective-taking ability.

Working memory has also been shown to predict perspective-taking ability. Lin et al. (2010) used an operation span test to get a measure of working memory, and then compared participants in the top 20% of working memory scores to those in the bottom 20% on a perspective-taking in comprehension task. They found that the high working memory group participants were significantly quicker to reach for the target item when a privileged competitor item was present. High working memory participants also showed significantly reduced fixations to the privileged competitor item when it was present.

In a second experiment, Lin et al. (2010) had participants complete the same referential communication task while also performing a memory load task where they had to memorize numbers in between trials of the referential communication task. On each trial, the participant could either be given a high memory load (memorizing 4 two-digit numbers) or a low memory load (memorizing 1 two-digit number). They found that in trials that had both a high memory load and a privileged competitor object, participants were slower to reach for the target object and fixated more on the privileged competitor

object. In trials without the privileged competitor, there was no difference, showing that perspective-taking specifically was affected by memory load, rather than there being a general delay.

If we turn to perspective-taking in production, Wardlow (2013) also found that both inhibition control (as measured by the Flanker task) and working memory (as measured by the digit span task) predicted perspective-taking ability (as measured by adjective-usage) in a production version of a referential communication task. Ryskin et al. (2015) also found that working memory (operation span) predicted perspective-taking in production but not in comprehension.

Besides executive function, bilingualism (Greenberg et al., 2013) and culture (Wu & Keysar, 2007) have also been shown to influence perspective-taking ability. Greenberg et al. (2013) did a spatial perspective-taking experiment with bilingual and monolingual children and found that bilingual children performed more accurately than monolingual children, showing that bilingualism may make one a better perspective-taker. However, one explanation for this bilingual advantage is that bilinguals may have greater inhibition control than monolinguals, so this bilingual advantage in perspective-taking may really be due to differences in inhibition control. Wu and Keysar (2007) did a referential communication task with both Chinese and American students and found that the Chinese students exhibited greater perspective-taking in their eye movements. They claim that this is due to differences between the more collectivist culture of China and the more individualistic culture of the U.S.

While bilingualism and culture may play a role in perspective-taking, in what follows, I will focus on the link between executive function and perspective-taking, rather than bilingualism and culture.

PROBLEMS WITH PREDICTING PERSPECTIVE-TAKING

The link between perspective-taking and executive function, however, may not be as clear as the studies above suggest. While many studies have found that executive function measures predict perspective-taking ability, many others have failed to find any connection.

Brown-Schmidt (2012) showed no link between two inhibition control (Stroop) tasks and perspective-taking in comprehension in a referential communication task. Ryskin et al. (2014) found that inhibition control inconsistently predicted perspective-taking in comprehension in three experiments (map task, table/grid task, and referential communication task). Ryskin et al. (2015) found that working memory (Operation span) predicted perspective-taking in production but not comprehension, and that inhibition control tasks did not predict perspective-taking in production or comprehension. And Brown-Schmidt and Fraundorf (2015) also failed to find any link between inhibition control (Stroop task) and perspective-taking in comprehension in three referential communication task experiments.

Both Ryskin et al. (2015) and Brown-Schmidt and Fraundorf (2015) propose that a potential reason they failed to find a link between their executive function measures and perspective-taking in comprehension is that they found their perspective-taking in comprehension tasks to have low reliability. This problem seems to be limited only to perspective-taking in comprehension, as Ryskin et al. (2015) found low internal

reliability only for perspective-taking in comprehension, but not for perspective-taking in production or cue generation.

This may mean that there are not consistent individual differences in perspective-taking ability in comprehension. If this is true, it might explain why there has been such inconsistency in the link between perspective-taking and executive function measures, as a lack of consistent individual differences would make perspective-taking in comprehension impossible to predict with any factor. We would need to carefully interpret any study that has previously shown a correlation between perspective-taking and other factors.

This would also have implications for models of reference resolution that have some way of weighting to what extent people use an egocentric or common ground domains of reference (for example, Heller et al., 2016), as we would not be able to include perspective-taking ability as a part of that weighting, although situational factors may still play a role. And finally, it may mean that perspective-taking in comprehension is not the true construct that we are after in these tasks. It is possible that we need to more carefully distinguish the requirements of each task that we use to study perspective-taking in order to make sure that perspective-taking in comprehension is really a valid construct, especially in light of the fact that Ryskin et al. (2015) has already shown that perspective-taking is not domain general between production and comprehension.

To summarize, perspective-taking is a common and necessary ability to comprehend referring expressions. However, what actually underpins the phenomenon of perspective-taking is unclear. There are multiple theories proposing different underlying mechanisms (Clark, 1996; Horton & Gerrig, 2005a; Pickering & Garrod, 2004), there are

conflicting results about whether inhibition control or working memory predict perspective-taking ability, (Brown-Schmidt, 2009; Brown-Schmidt & Fraundorf, 2015), perspective-taking in comprehension and production do not seem to correlate (Ryskin et al., 2015), and perspective-taking in comprehension tasks do not seem to be reliable (Brown-Schmidt & Fraundorf, 2015; Ryskin et al., 2015). I propose that the distinction between spatial and information-mismatch perspective-taking may be one source of confusion around some of the inconsistent results in the literature. However, it may be that the distinctions need to be even more granular to accurately explain perspective-taking.

THE CURRENT STUDY

Experiment 1 of the dissertation seeks to test whether perspective-taking in comprehension is indeed a reliable ability. It does this by having the same group of participants complete multiple perspective-taking in comprehension tasks. This allows us to see if performance in one comprehension task will be correlated with performance in the other tasks, something that previous researchers that found potential unreliability (Brown-Schmidt & Fraundorf, 2015; Ryskin et al., 2015) were unable to do. Furthermore, if perspective-taking in comprehension is reliable, experiment 1 seeks to test which cognitive resources (working memory and inhibition control) are most likely to predict perspective-taking ability. The main finding of this experiment is that the resources required seem to depend on what kind of perspective-taking is measured (spatial or information-mismatch), and that it is therefore a mistake to consider perspective-taking to be a unitary ability that always recruits the same cognitive resources.

This distinction is important as it may help explain the inconsistent replication of a link between inhibition control and the speed and likelihood of perspective-taking (Brown-Schmidt, 2012; Brown-Schmidt & Fraundorf, 2015; cf. Brown-Schmidt, 2009). It may be that the relationship depends on the specific demands of the task and which cognitive resources are recruited, and so that is why many studies have failed to replicate this relationship. Furthermore, previous research has shown that perspective-taking is not domain-general between production and comprehension (Ryskin, et al., 2015). If perspective-taking in comprehension is also not a unitary ability, then we would need to change our conception of what is happening while considering multiple perspectives during language processing. This may also bring the linguistics literature's conception of perspective-taking more in line with the more multifaceted views of theory of mind (Wellman & Liu, 2004) and visual perspective-taking (Michelon & Zacks, 2006) that exist in psychology.

Experiment 2 of the dissertation is structured so as to directly test the spatial vs information-mismatch distinction in perspective-taking that is suggested by Experiment 1. The goal is to show that these two types of perspective-taking do not correlate with one another, and that each recruits a different set of general cognitive resources.

One major goal of the dissertation is to differentiate between these two possible models of perspective taking. The first model, the unified model, is the one implicitly assumed by much of the literature, and takes perspective-taking to be a single unified construct, of which both spatial and information-mismatch perspective-taking tasks are a part. This unified perspective-taking ability may or may not be predicted by executive function.

The alternate model, or the split model, is one in which perspective-taking is composed not of a single construct but a constellation of different but related abilities. For the purposes of this dissertation, I limit this to two likely candidates for dissociable components: spatial and information-mismatch perspective-taking. Under this model, the spatial and information-mismatch aspects of perspective-taking may not correlate with one another. And spatial and information-mismatch aspects may recruit different resources: with spatial perspective-taking likely to recruit general spatial ability and information-mismatch perspective-taking potentially more likely to recruit executive function resources.

CHAPTER 2

EXPERIMENT 1[1]

INTRODUCTION

Purpose and Research Questions

Previous studies disagree about whether inhibition control or working memory predict how often and quickly perspective information is incorporated during language comprehension (Brown-Schmidt, 2009; Lin et al. 2010; compare with Brown-Schmidt, 2012; Ryskin et al., 2014, 2015). Furthermore, Brown-Schmidt and Fraundorf (2015) and Ryskin et al. (2015) suggest that the unreliability of perspective-taking in comprehension tasks may make it impossible to predict perspective-taking during comprehension. If these tasks are unreliable and there are not consistent individual differences in perspective-taking during comprehension then nothing will predict perspective-taking ability consistently. Debate over a link between perspective-taking (in comprehension) and executive function is therefore getting ahead of itself unless we first get reliable measures of perspective-taking during comprehension or determine that there are simply not individual differences in perspective-taking ability during comprehension.

[1] This chapter, except for the introduction, was first published as Luce, K., & Almor, A. (2023). Inconsistency in Perspective-Taking during Comprehension. *Psychonomic Bulletin & Review*. https://doi.org/10.3758/s13423-023-02315-0. It is reproduced with permission from Springer Nature. See Appendix A.

It is therefore important to check whether there really are consistent individual differences in perspective-taking ability during comprehension. One way to do this is to have the same group of participants complete multiple perspective-taking in comprehension tasks. This would allow us to assess both the reliability of these tasks by themselves, but also see whether performance in one task predicts performance in the other tasks. Previous studies have had different groups of participants complete different perspective-taking in comprehension tasks (Brown-Schmidt & Fraundorf, 2015). Or have the same group of participants complete perspective-taking tasks in production, comprehension, and cued recall (Ryskin et al. 2015). However, to my knowledge, there hasn't been a study that compares the same group of participants in multiple perspective-taking in comprehension tasks.

Experiment 1 was designed to answer the questions (1) are there consistent individual differences in perspective-taking ability during comprehension, and (2) if there are, are they predicted by inhibition control or working memory. In order to achieve this we recruited a large group of participants to complete three different perspective-taking in comprehension tasks, as well as measures of inhibition control and working memory. We sought to test the first question by looking at the reliability of each of the three perspective-taking tasks and seeing whether participants' scores in the tasks correlate with one another. If the scores in the three perspective-taking tasks do correlate, that would be strong evidence that there are consistent individual differences, as a good score in one task correlates with a good score in the other tasks. Because we need to make sure that the tasks are all measuring the same thing, we used tasks that have been used in previous research and shown effects of perspective.

We test whether executive function predicts perspective-taking during comprehension by also including measures of inhibition control and working memory. If these components of executive function are necessary for perspective-taking, we should see them correlate with our three perspective-taking measures.

EXPERIMENT METHODOLOGY

Selection of Tasks

We selected tasks based on two criteria. The first being a previous history of showing perspective effects in comprehension, and the second being coverage of different aspects of perspective-taking. We selected tasks that vary on linguistic/non-linguistic and spatial/non-spatial dimensions so as to represent different types of perspective-taking tasks, as we believe it is important to consider a variety of tasks in order to isolate the construct of perspective-taking rather than the measure of a single monolithic task. Despite their variance on these dimensions, all of these tasks have fallen under the heading of perspective-taking in the literature.

Our first task (Blocks) required general (non-linguistic) visual-spatial perspective-taking and was modeled after Greenberg et al., (2013) who used this task to show that bilingualism affects perspective taking in children. In this task, participants must select how a partner views a pattern of blocks from a different spatial position than their own. While the original task measured the accuracy of the children’s selections, we will also measure response time with the idea that this will be a more sensitive measure when using adult participants, who may make less errors than children.

Our second task (Cubbies) was linguistic and non-spatial, requiring information-mismatch rather than a spatial perspective-taking in the context of language

comprehension. This task used a version of the referential communication task with a grid of cubbies that has been used widely in the perspective-taking literature, and has provided consistent effects of perspective (e.g., Keysar et al., 2000; Heller et al., 2016). Because our tasks were administered online, we again use RT as our dependent variable for this task, instead of the eye-tracking measures that are often used for this task. Given that previous studies show differences in accuracy between conditions for this task (Keysar et al., 2000; Heller et al., 2016) and other perspective tasks (Greenberg et al., 2013; Ryskin et al. 2014), we assume that RT should also be sensitive enough to capture perspective-taking behavior. Furthermore, we believe that it is beneficial to use RT rather than eye-tracking for our purposes in this experiment. Previous studies have shown low reliability for perspective-taking in comprehension, which may be due to the noisy nature of eye-tracking data. RT may be better for looking at individual differences, as well as being better suited to the primary issue of cognitive demand.

Our third perspective-taking task (Tables) was similar to (but far from a direct replication of) experiment 2 in Ryskin et al. (2014). This was a linguistic and spatial perspective-taking language comprehension task that manipulated whether direction terms (left and right) were matched or mirrored between the speaker and the comprehender. We again use RT as our dependent measure for this task.

These three tasks provide a good representation of the use of perspective-taking in comprehension and thus allowed us to see if perspective-taking ability is consistent for participants across these different perspective-taking in comprehension tasks. However, note that, because our tasks involved taking the perspective of a cartoon owl, they are less interactive than if participants were responding to a live person. We believe that we

should still see perspective-taking effects, given that previous research has shown perspective-taking with dolls (Michelon & Zacks, 2006) and arrows (Ryskin et al., 2016). However, these perspective effects have been shown to be attenuated in non-interactive tasks (Brown-Schmidt, 2009), which could limit the interpretation of our results.

We also include two executive function tasks. Although, ideally, we might include several tasks for each component of executive function, we had to limit them to two in order to keep the experiment at a reasonable length. Because of this, we chose tasks that were likely to show a correlation with perspective-taking given previous results. We used a version of the Stroop task for measuring inhibition control, as this task has been found to predict perspective-taking in comprehension before (Brown-Schmidt, 2009). And we used the operation span task to measure working memory, which Ryskin et al. (2015) found to predict perspective-taking ability in production and cue generation (but not comprehension).

Participants

Two groups of participants were recruited. All participants had normal or corrected-to-normal vision and were native English speakers.

The first group of participants were students at an Anonymous U.S. University who were given psychology class credit for their participation. This group included 56 participants (50 female, average age 19.9, standard deviation 1.4). Two participants were dropped because they indicated during the pre-experiment survey that they were not native English speakers.

The second group of participants were recruited via Amazon Mechanical Turk and were paid $7.25 for their participation. The participants were Amazon Masters and

required to be located in the United States. This group included 67 participants (29 female, average age 43.7, standard deviation 11.7). One participant was dropped because they indicated during the pre-experiment survey that they were not a native English speaker.

Materials

The experiment was written and hosted through the Penn Controller for Ibex Farm (Zehr & Schwarz, 2018). Participants completed the experiment on their own computers through a web link. In each of the five tasks (three perspective-taking and two executive function), they were presented with visual and written stimuli and responded either by clicking the screen with their cursor or pressing a button on their keyboard.

Procedure

Participants completed three perspective-taking tasks, followed by a Stroop task and an operation span task. The order of the three perspective-taking tasks was balanced across participants. The entire experiment took approximately one hour to complete.

Pre-Experiment Survey: Before completing the experiment, participants were asked to complete a short survey that gathered information about their age, gender, native language(s), whether they spoke other languages, whether they had normal or corrected-to-normal vision, and whether they had any problems seeing colors.

Perspective Task 1: "Blocks" On each trial of this task, participants were presented with a screen that included four colored blocks (colored red, blue, green, and yellow) arranged in a diamond pattern, as well as a cartoon character, Mr. Owl, who viewed the pattern of blocks from a different angle (rotated either 90, 180, or 270 degrees from the participant's perspective). On each trial, the participant would respond to the

question "What does Mr. Owl see from where he is standing?" by selecting one of four provided answers. The pattern of blocks, question, and answers were all presented on the same screen (see Figure 2.1). We measured the selection that the participant made, as well as their reaction time.

The four choices shown to participants were pictures of different arrangements of the same four colored blocks. They included 1) the correct arrangement (as seen by Mr. Owl), 2) an arrangement that contained a structural error (where two blocks switched position) but the proper rotation, 3) an arrangement that contained a rotational error (rotated 90 instead of 180 degrees) but the proper structure, and 4) an egocentric arrangement that displayed the blocks in the same pattern as originally viewed by the participant. The four answers were presented in random order. Participants made their selection by clicking on the image of their chosen answer.

Because this task does not contain multiple conditions[2], we treat it as an individual differences measure, with the only prediction being that people who are better at perspective-taking will respond more quickly.

Each participant saw 21 critical trials, presented in a random order. One third of the trials had Mr. Owl standing at a 90 degree rotation, one third at a 180 degree rotation and one third at a 270 degree rotation. There were 6 practice trials, during which

---

[2] We did find that 180 degree trials were responded to more slowly than 90 and 270 degree trials. However, due to the lack of a baseline perspective-match condition, we did not distinguish these angles as separate conditions, as they all require a perspective-mismatch. We did attempt to analyze the Blocks task with a mixed-effects model in the same fashion as the other tasks, using degree rotation as the fixed-effect, however when doing this, the participant slopes for condition did not correlate with any of the other tasks.

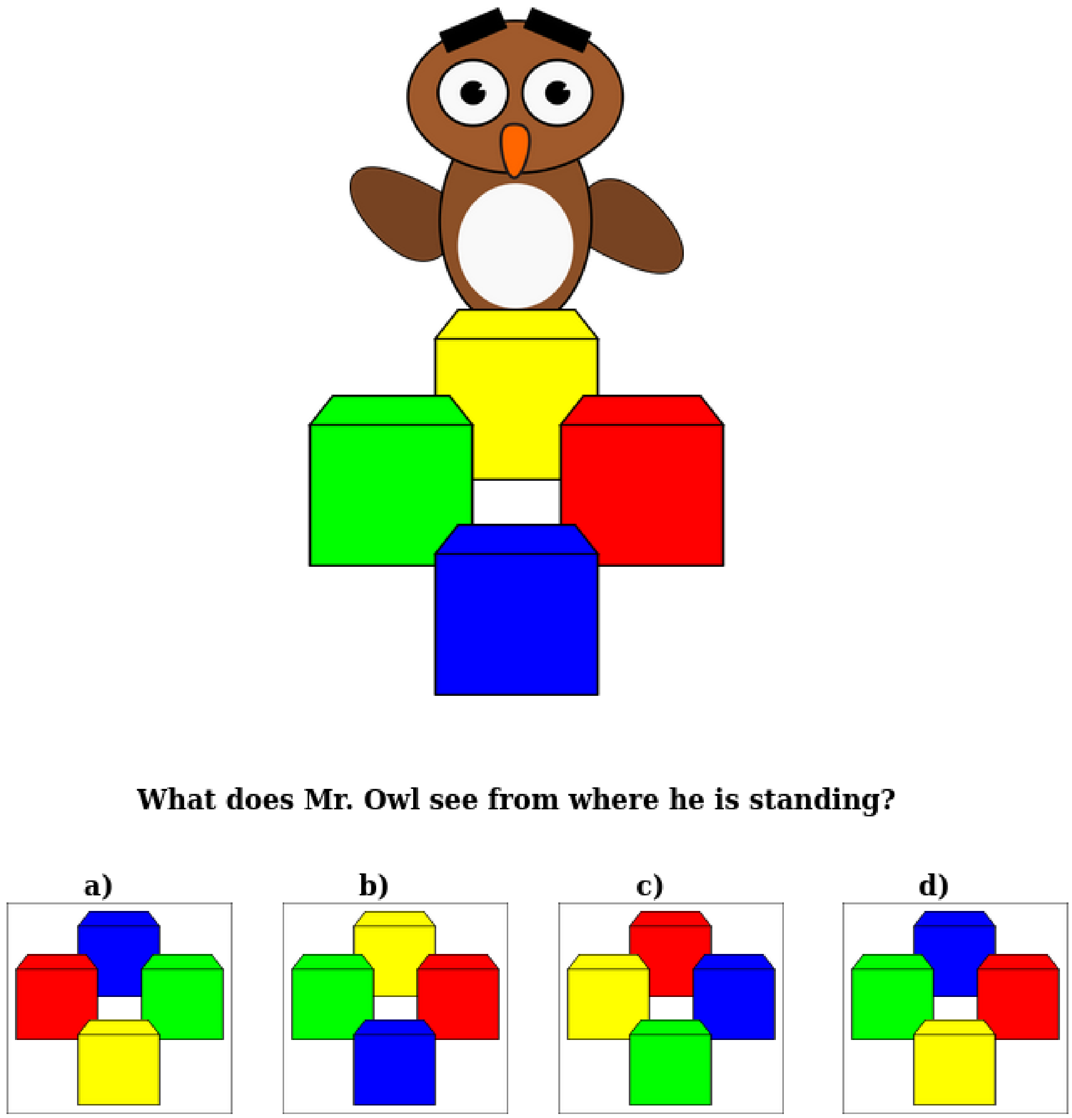

Figure 2.1 A trial from the "Blocks" task

participants were given feedback about whether they made the correct selection. No feedback was provided during critical trials.

Perspective Task 2: "Cubbies" On each trial, participants were presented with a screen that included a 3x3 grid of cubby-holes which was filled with objects, as well as a cartoon character, Mr. Owl, who faced the participant from the other side of the cubby-holes. On critical trials, Mr. Owl would give the participant directions in the form of "Click on the large X", where X was one of the objects in the cubbies. The participant's selection and reaction time was recorded.

In all critical trials, there were four objects in the cubbies. One object that appeared in three different sizes (e.g., a small duck, a medium duck, and a large duck) and another object that appeared in only one size (e.g., a medium sponge).

There were two critical conditions, which varied who could see each object. In the perspective-match condition (Figure 2.2, left), all four of the objects were in common ground cubbies (visible to both the participant and Mr. Owl). In the perspective-mismatch condition (Figure 2.2, right), the largest object was in a privileged ground cubby (visible only to the participant and not Mr. Owl) while the rest of the objects were in common ground cubbies. A "curtain" blocked the view of cubbies that were hidden to either the participant or Mr. Owl.

The prediction is that participants will take longer to select an object in the perspective-mismatch condition than in the perspective-match condition. This is because in the perspective-mismatch condition, there is some ambiguity in Mr. Owl's instruction to "Click on the large X". From an egocentric perspective, this would be the largest object in the display. However, if the participant is considering Mr. Owl's perspective,

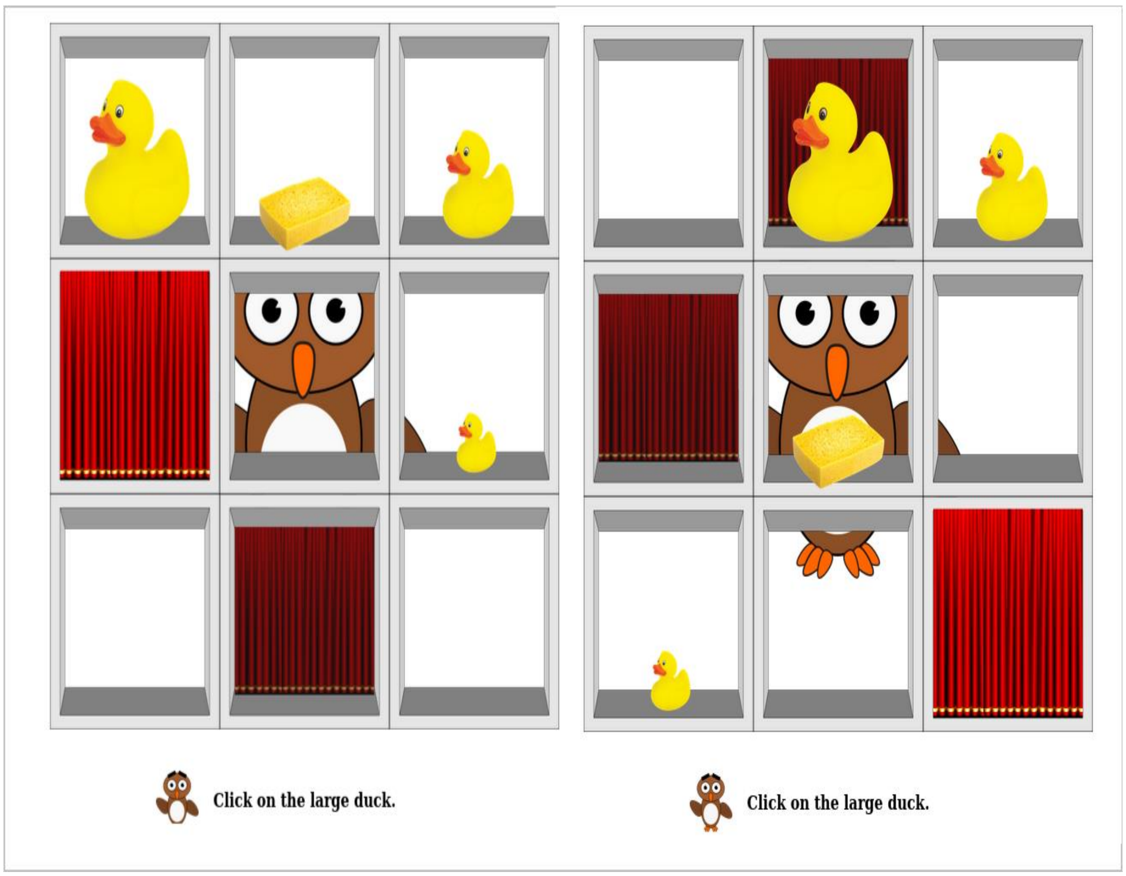

Figure 2.2 The two conditions of the "Cubbies" task

they would realize that the largest object is hidden from Mr. Owl in this condition and that he therefore must mean the medium object, which appears comparatively large based on the common ground context set. This conflict between the egocentric and allocentric perspective should slow participants compared to the perspective-match condition, where all the information in the scene is in the common ground.

Each participant saw 16 critical trials, 8 in each condition. Half of the critical trials in each condition used “small” as the descriptive adjective and the other half “large”. Each critical trial used a different target item. The positions of all objects and the cubby types were randomized on each trial. The trials were presented in random order, including fillers.

Each participant also saw 20 filler trials. 12 of the fillers manipulated the color of the objects rather than the size. The other 8 fillers directed the participants to click on one of the non-critical items (e.g., the singleton object, the medium object).

There were 8 practice trials, which used different target items. These were presented in random order, and feedback was given to the participant about whether they made the correct choice. These practice trials were used to enforce the idea that Mr. Owl did not know what was in the privileged cubbies, and that the “correct” selection in perspective-mismatch cases was to select the medium object.

Perspective Task 3: “Tables” On each trial, participants were presented with a screen that included a 5x3 array of objects on a table as well as a cartoon character, Mr. Owl. Mr. Owl instructed the participants to click on one of the objects in the array, with directions in the form of: “Click on the duck to the right of the sponge.” The participant’s selection and reaction time was recorded.

There were two conditions. In the perspective-match condition (Figure 2.3, left), Mr. Owl appeared at the bottom half of the table, sharing the participant's view of the table. In the perspective-mismatch condition (Figure 2.3, right), Mr. Owl appears at the top edge of the table, facing the table and the participant from the opposite side.

The prediction was that reaction times would be faster in the perspective-match than in the perspective-mismatch conditions. This is because when Mr. Owl says "to the right", in the perspective-match condition this is to the participant's right as well, while in the perspective-mismatch condition, Mr. Owl's perspective mirrors the participant's and so the target object is to the left.

The critical item always had the same item to the left and right of it. In this way, the instructions given by Mr. Owl were ambiguous as there was the same object both to the left and right of the target, and information about the relative perspective had to be used.

Each participant saw 32 critical trials, 16 in each condition. Trials alternated between conditions. Each object appeared once in each condition as a target item, and once in each condition as the reference item. Distractor items were reused target/reference items from other trials. There were no filler trials.

There were 8 practice trials, 4 in each condition. These used different critical items. Feedback was given during practice trials. Before the practice trials, participants were given explicit instructions about the mirrored behavior of Mr. Owl (that when he was on the opposite side of the table and said "right" that would mean your left).

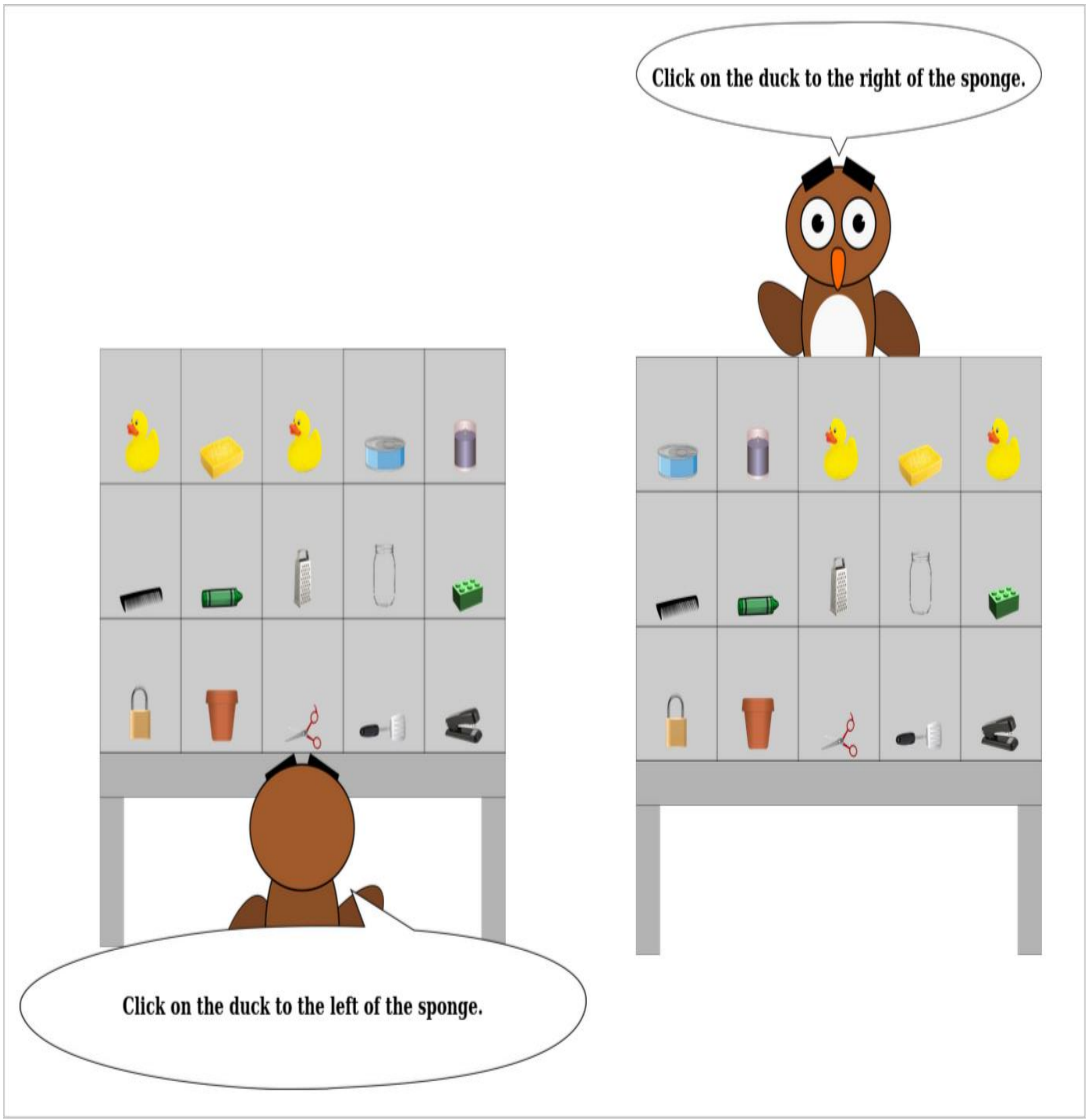

Figure 2.3 The two conditions of the "Tables" task

Stroop Task This was a version of the Stroop task (Stroop, 1935). On each trial, participants were presented with a screen that included one of three color words (green, red, or blue) in one of three font colors (green, red, or blue). Their task was to press the 'g' key if the font color was green, the 'r' key if red, and the 'b' key if blue. The participant's selection and reaction time was measured.

There were two conditions. On congruent trials, the color of the word and the font matched. On incongruent trials, the color of the word and the font did not match.

Each participant saw 120 trials, of which 90 were congruent trials (30 for each color pair) and 30 were incongruent trials (5 for each color pair, counting red-blue and blue-red as different pairs). After each trial, participants were required to wait 1 second before the next trial would begin automatically. During this time, participants were given feedback about their answer that either said "Correct!" or "Incorrect - press the key that matches the COLOR of the text, not the meaning of the word".

There were 10 practice trials, 5 congruent and 5 incongruent. Feedback was provided both during the practice trials and the critical trials. During practice, participants were forced to wait 2 seconds after each trial and were given the same feedback as in the critical trials. The order of the trials was random during practice and critical trials.

The prediction was that the incongruent trials will take participants longer than congruent trials.

Operation Span Task This was a version of the operation span task (Turner & Engle, 1989). On each trial, participants were presented with a screen that included a math question in the form of "(X * Y) + Z = ?". After 12 seconds or clicking a button, they were presented with a new screen with a potential solution to the math problem.

They then had to decide whether that solution was correct or incorrect by clicking a button on the screen. After making their selection, they were given a letter to remember. This process repeated 3-7 times, after which participants were taken to a screen that contained 12 letters with text boxes. Participants were required to write numbers next to the letters that they memorized in the correct order.

Each participant saw 15 trials, 3 each of the different span lengths between 3 and 7. Letters were randomized out of 12 possible letters. The numbers of the math problems were random numbers between 1 and 99. Order of span lengths was randomized. Feedback was given after each span was completed about how many math problems were correctly solved and how many letters were correctly remembered.

There were 3 different practices for this task. The first practice was just for the letters part of the task, the second practice was just for the math part of the task, and the final practice combined the two.

RESULTS

Subject Source

Because we recruited participants from two different groups, subject source was entered as a random effect in our analysis for the tasks that had multiple conditions. For the “Blocks” and Operation Span tasks, we collapse all participants into a single group, although we note that participants from Amazon Mechanical Turk were slower to respond in the Blocks task, but more accurate in the Operation Span task.

Perspective Task 1: “Blocks”

Our dependent variable of interest for the Blocks task was reaction time for correct responses. Participants who had accuracy less than 60% were excluded from the

analysis, on the basis that they did not understand the task well and that their responses would therefore not be reflective of perspective-taking ability. This affected 1 participant. From the remaining 119 participants, we removed all incorrect responses from the analysis (3.6% of responses). We also removed outlier RTs that were above 2.5 or below -2.5 standard deviations from the mean (0.1%). After outlier removal, the mean RT was 5356 ms (SD: 4117 ms). No main effects are provided for this task, as there was only a single condition.

Perspective Task 2: "Cubbies"

Our dependent variable of interest for the Cubbies task was reaction time for correct responses. Participants who had accuracy less than 60% were excluded from the analysis. This affected 6 participants. From the remaining 114 participants, we removed all incorrect responses from the analysis (4.1% of responses). We also removed outlier RTs that were above 2.5 or below -2.5 standard deviations from the mean (0.7%). After outlier removal, the mean RTs were 2749 ms (SD: 1425 ms) for the perspective-match condition, and 2974 ms (SD: 1418 ms) for the perspective-mismatch condition.

Raw[3] RTs were analyzed using linear mixed effects modeling, with the lme4 package (Bates et al., 2015) in R (R Core Team, 2023). Condition (perspective-match vs perspective-mismatch) was entered as a fixed effect. Subjects, Items, and Subject-Source were entered as random effects, using the maximal random effects structure that allowed the model to converge (Barr, Levy, Scheepers, and Tily, 2013). Results of the full model

---

[3] We use raw RTs rather than log-transformed RTs due to the concern that log-transforming the data may obfuscate real individual differences in right skewed data when doing the correlation analysis. However, using raw RTs may somewhat inflate the reliability of these measures (See Staub, 2021).

are reported in Table 2.1. As can be seen from the table, perspective-mismatch trials were responded to more slowly than perspective-match trials.

Perspective Task 3: "Tables"

Our dependent variable of interest for the Tables task was reaction time for correct responses. Participants who had accuracy less than 60% were excluded from the analysis. This affected 1 participant. From the remaining 119 participants, we removed all incorrect responses from the analysis (4.4% of responses). We also removed outlier RTs that were above 2.5 or below -2.5 standard deviations from the mean (0.8%). After outlier removal, the mean RTs were 5443 ms (SD: 4100 ms) for the perspective-match condition, and 6319 ms (SD: 4126 ms) for the perspective-mismatch condition.

Raw RTs were analyzed using linear mixed effects modeling, with the lme4 package (Bates et al., 2015) in R (R Core Team, 2023). Condition (perspective-match vs perspective-mismatch) was entered as a fixed effect. Subjects, Items, and Subject-Source were entered as random effects, using the maximal random effects structure that allowed the model to converge (Barr et al., 2013). Results of the full model are reported in Table 2.2. As can be seen from the table, perspective-mismatch trials were responded to more slowly than perspective-match trials.

Stroop Task

Our dependent variable of interest for the Stroop task was reaction time for correct responses. Participants who had accuracy less than 60% were excluded from the analysis. This affected 0 participants. From the remaining 120 participants, we removed all incorrect responses from the analysis (1.4% of responses). We also removed outlier RTs that were above 2.5 or below -2.5 standard deviations from the mean (0.3%). After

Table 2.1 E1 Parameter Estimates for Perspective Task 2: Cubbies

Fixed Effects	β	SE	*t*	*p*
(Intercept)	2740.145	231.341	11.845	0.005*
perspective-mismatch	216.852	63.924	3.392	0.001*
Random Effects	Variance			
Subject				
(Intercept)	756340			
perspective-mismatch	198986			
Subject-Source				
(Intercept)	89212			
Item				
(Intercept)	33452			
Observations: 1737	Groups:	Subjects: 114	Items: 32	Subject-Source: 2

Significant effects at a $p \leq .05$ level are marked with an *.

Table 2.2 E1 Parameter Estimates for Perspective Task 3: Tables

Fixed Effects	β	SE	*t*	*p*
(Intercept)	5427.896	194.600	27.893	0.001*
perspective-mismatch	870.815	136.504	6.379	0.000*
Random Effects	Variance			
Subject				
(Intercept)	2647395			
perspective-mismatch	385694			
Subject-Source				
(Intercept)	15781			
Item				
(Intercept)	94211			
Observations: 3613	Groups:	Subjects: 119	Items: 64	Subject-Source: 2

Significant effects at a $p \leq .05$ level are marked with an *.

outlier removal, the mean RTs were 745 ms (SD: 448 ms) for the congruent condition, and 948 ms (SD: 500 ms) for the incongruent condition.

Raw RTs were analyzed using linear mixed effects modeling, with the lme4 package (Bates et al., 2015) in R (R Core Team, 2023). Condition (congruent vs incongruent) was entered as a fixed effect. Subjects, Items, and Subject-Source were entered as random effects, using the maximal random effects structure that allowed the model to converge (Barr et al., 2013). Results of the full model are reported in Table 2.3. As can be seen from the table, perspective-mismatch trials were responded to more slowly than perspective-match trials.

Operation Span Task

Our dependent variable of interest for the Operation Span task was the percentage of spans in which a participant remembered all the letters correctly. 4 participants were removed for not getting any spans correct. For the remaining 116 participants, the mean for accurate spans was .72 (SD: .45).

Correlations between Tasks

In addition to the above experimental effects, we also want to see if there are consistent individual differences in each task, and if these differences correlate across different tasks.

In order to carry out the correlation analysis, one measure was used to represent each task, and averaged for each participant. For Perspective Task 1: “Blocks”, the average reaction time of correct responses of each participant was used. For the Operation Span task, the percentage of trials in which a participant remembered all the letters in a span was used.

Table 2.3 E1 Parameter Estimates for the Stroop Task

Fixed Effects	β	SE	*t*	*p*
(Intercept)	748.45	32.38	23.12	0.002*
Incongruent	203.30	10.27	19.80	0.000*
Random Effects	**Variance**			
Subject				
(Intercept)	20582			
Incongruent	7795			
Subject-Source				
(Intercept)	1731			
Observations: 14153	Groups:	Subjects: 120		Subject-Source: 2

Significant effects at a $p \leq .05$ level are marked with an *.

For Perspective Task 2: “Cubbies”, Perspective Task 3: “Tables”, and the Stroop task, we extracted subject slopes for condition from each of the mixed effects models, and used these subject slopes as the measure for how good the participant was at the task.

In order to assess the reliability of these model-based measures we split the data for each task into two halves, and then fit a separate model for each half, using the same fixed and random effects structure as the full model. We then compared the correlation between the subject slopes for each half. We originally split the data based on odd/even trial numbers. However, this split led to a singular convergence error for the “Cubbies” model and so for this task we report the results from splitting the data by every two trials instead of odd/even trial numbers. We also split the data by every two trials for the “Tables” task, as in this task, all odd trials and all even trials were in the same condition. The Stroop task was split based on odd/even trial numbers.

The Spearman-Brown adjusted correlation between the by-subject slopes for conditions of the two models for the “Cubbies” task was p = .01. This reflects very poor reliability for this task. For the “Tables” task, the Spearman-Brown adjusted correlation was p = .79. And for the Stroop task, the Spearman-Brown adjusted correlation was p = .79. We take the reliability of these two tasks to be good but not highly reliable.

For the “Blocks” task, we calculated the average Spearman-Brown adjusted correlation between 5000 random split-halves to be p = .9 (with a 95% confidence interval of [.86, .92]), making this task very reliable, with the caveat that this number reflects overall reaction time, rather than a difference between conditions. For the Operation Span task, we calculated the average Spearman-Brown adjusted correlation

between 5000 random split-haves to be p = .68 (with a 95% confidence interval of [.59, .76]), making this task acceptably reliable.

Due to our accuracy threshold, each task had a different number of participants, so for each correlation between pairs of tasks, we used the maximum number of participants that had data for both tasks. The number of participants contributing to each correlation is shown underneath the correlation coefficients in Table 2.4.

The correlation table is shown below. After using a Bonferroni correction to account for multiple comparisons, only one significant correlation was found. This significant correlation was between the measures of perspective-taking in the “Tables” and “Blocks” tasks ($r = 0.53$, $p < 0.001$). The positive correlation suggests that perspective taking ability in the two tasks is linked.

We note that there are multiple correlations present in the results if we maintain an alpha of .05 and do not correct for multiple comparisons. This includes a negative correlation between the “Cubbies” and “Tables” tasks and a correlation between the “Blocks” and “Stroop” tasks. However, we take these to not be meaningful correlations due to 1) the very low reliability of the “Cubbies” task, 2) the odd negative direction of the correlation, and 3) the lower strength of the correlations not being significant at a .005 level.

DISCUSSION

Our results affirm previous research that shows that people take perspective information into account during comprehension, and that this maintenance of multiple perspectives is effortful. We see this in Perspective Task 2: “Cubbies" and Perspective Task 3: “Tables”, where reaction times during perspective-mismatch conditions are

Table 2.4 E1 Correlation Matrix

	Blocks (.90)	Cubbies (.01)	Tables (.79)	Inhibition (.79)	Working Memory (.68)
Blocks (.90)	---				
Cubbies (.01)	-.16 (*p* = .099, n = 113)	---			
Tables (.79)	**.53* (*p* < .001, n = 118)**	-.25 (*p* = .007, n = 113)	---		
Inhibition (.79)	.19 (*p* = .044, n = 119)	-.16 (*p* = .095, n = 114)	.17 (*p* = .061, n = 119)	---	
Working Memory (.68)	-.04 (*p* = .674, n = 115)	-.15 (*p* = .118, n = 110)	.01 (*p* = .896, n = 115)	-.16 (*p* = .087, n = 116)	---

Significant correlations at a $p \leq .005$ level are marked with an *. This reflects an α of .05 with a Bonferroni correction for 10 comparisons. The reliability appears in parentheses under the name of each task.

slower than during perspective-match conditions. This confirms that these tasks were indeed providing us the expected effects of perspective-taking.

As for the correlations between tasks, the results are quite sparse, with only one significant correlation between the “Tables” and “Blocks” tasks. Our results diverge from our predictions based on the previous literature in two important areas. Firstly, that there was a lack of correlation between one of our measures of perspective-taking and the other two, although the very low reliability of the “Cubbies” task makes it hard to draw much inference from this.

The second area of divergence was that we found both 1) internal reliability of (some of) our perspective-taking measures and 2) a lack of correlation between those measures and inhibition control or working memory. This contradicts previous suggestions that the tenuous relationship between inhibition control and perspective-taking is due to a lack of internal reliability in perspective-taking, at least for the “Tables” and “Blocks” task, although this explanation is obviously still very relevant for the “Cubbies” task.

Relationship between Different Measures of Perspective-Taking in Comprehension

We found that our measure of perspective ability on the “Cubbies" task had very poor internal reliability, despite the strong experimental effect. This means that participants did not show much in the way of individual differences on this task. This likely explains why we did not find performance on the “Cubbies” task to correlate with perspective-taking ability in either the “Tables” or “Blocks” tasks. The “Tables” and “Blocks” tasks, on the other hand, both showed acceptable levels of reliability, and correlated significantly with one another.

The fact that the three perspective-taking tasks did not all correlate with one another is perhaps not too surprising given these reliability numbers. However, it leads us to question why our “Cubbies” task was so much less reliable than the “Tables” or “Blocks” tasks. Previous research (Brown-Schmidt & Fraundorf, 2015; Ryskin et al., 2015) has suggested that perspective-taking in comprehension may simply not show consistent individual differences in participants.

However, this explanation cannot be the whole story, given that two of our three perspective-taking tasks showed good internal reliability. If perspective-taking in comprehension was simply not reliable, all three tasks should have shown low reliability and performance in none of the three tasks should have been correlated. Thus, we should take a more nuanced view on perspective-taking in comprehension.

Specifically, we ask why the “Blocks" and “Tables” tasks were reliable and correlated, and how they differed from the “Cubbies” task. The most likely explanation seems to be the fact that both the “Blocks” and “Tables” tasks required visuo-spatial perspective-taking while the “Cubbies” task did not. The “Blocks” and “Tables” tasks required participants to spatially rotate or mirror their perspective. The “Cubbies” task on the other hand, required participants to keep track of common and privileged information, but required no spatial rotation of their perspective. Although all three of these tasks are supposed measures of perspective-taking in comprehension, the differences between the three tasks suggest that the type of perspective-taking required by the spatial tasks may be fundamentally different from the type required by the information-mismatch task.

An alternative explanation for the correlation between “Blocks” and “Tables” is suggested by the high reliability of the “Blocks” task. This high reliability could reflect

the fact that, as this task did not have multiple conditions, its results may reflect individual differences in participants' baseline response times rather than their perspective-taking effort. This could make it more likely that the "Blocks" task will correlate with other tasks that also contain a strong response time component, making it difficult to say whether the correlation is actually due to participants overall response speed or due to their perspective-taking ability. However, the fact that "Blocks" correlated with "Tables" but not with the Stroop task, which also reflects response time and showed high reliability, makes this alternative explanation less likely and lends more credence to the interpretation of the correlation between "Blocks" and "Table" as actually reflecting perspective-taking ability.

Whether either explanation is correct will require further research. However, our results suggest that perspective-taking in comprehension may not be a unitary construct. Previously, Ryskin et al. (2015) have shown dissociations between perspective-taking ability in comprehension, production, and cue generation, leading them to conclude that perspective-taking is not domain-general. Our results suggest that we were also incorrect to assume a general perspective-taking ability in comprehension. It seems that perspective-taking ability in comprehension may actually be decomposable into several different components, such as spatial and information mismatch components, which are engaged to different degrees by different tasks. This explanation might also account for the inconsistent replication of a link between inhibition control and perspective-taking, as different perspective-taking tasks may recruit different cognitive resources, or show differences in reliability. This would have major consequences for theories of perspective-taking in comprehension, as what is usually conceptualized as a single ability

would now need to be subdivided into different components that may depend on different cognitive resources. Further research is needed to uncover what exactly the components of perspective-taking in comprehension may be.

Relationship between Perspective-Taking and Executive Function

None of our three measures of perspective-taking ability correlated with inhibition control (as measured by the Stroop task) or working memory (as measured by the Operation Span task). This finding contributes another negative result to a literature that has found evidence both for and against the correlation of these abilities.

Previous studies have suggested that the tenuous relationship between inhibition control and perspective-taking ability in comprehension is due to the lack of internal reliability of perspective-taking in comprehension (Brown-Schmidt and Fraundorf, 2015; Ryskin et al., 2015). We also find this to potentially be true in our “Cubbies” task. However, this suggestion cannot explain the results of our “Tables” and “Blocks” tasks, both of which found a lack of any relationship between executive function measures and perspective-taking ability, while still having decent internal reliability. This means that the lack of correlation between executive function and perspective-taking cannot be pinned on a lack of internal reliability for these tasks. It would either need to be explained in another way, or it would suggest that these executive function measures do not predict perspective-taking ability.

LIMITATIONS

We must consider a few important limitations of this study when interpreting the null correlations. First, the perspective-taking tasks in this study were all non-interactive. Although we may still expect to see some perspective effects in non-interactive tasks,

they may be attenuated compared to interactive ones (Brown-Schmidt, 2009). Interactive tasks tend to have a more naturalistic reason for perspective-taking to be necessary compared to non-interactive tasks. Secondly, the perspective-taking tasks in this study used a non-human interlocutor (Mr. Owl). While there is evidence that perspective-taking happens even with non-human interlocutors such as a doll (Michelon & Zacks, 2006) or an arrow (Ryskin et al., 2016), the non-human interlocutor is still only a stand-in for a human, which adds a layer of representation in comparison to a natural dialogue with a living person. Third, the online modality of the experiment as well as the written instructions may be an attenuating factor for perspective-taking compared to experiments that happen live in a lab with oral instructions. And finally, the experiment included only one measure each of inhibition control and working memory, which limits the extent to which we can get at these constructs of interest, due to the measurement error of the tasks. Future research would benefit from multiple measures of each construct.

In conclusion, we found that perspective-taking in comprehension may not be one unitary construct but rather consists of dissociable components, based on the consistent individual differences found in our two spatial perspective-taking tasks but not in the "Cubbies" task. This calls into question a unitary theoretical conception of perspective-taking and necessitates further research into the structure of perspective-taking in comprehension.

CHAPTER 3

EXPERIMENT 2

INTRODUCTION

Rationale for Experiment 2

One explanation for the lack of correlation between the perspective-taking measures in Experiment 1 is that these tasks reflect different processes. It could be that the “Cubbies” task does not correlate with the “Blocks” and “Tables” tasks because it is measuring a different underlying process. If this is the case and we still want to call all three of those tasks “perspective-taking” tasks, then it would mean that perspective-taking needs to be decomposed into dissociable components that recruit different resources. While Experiment 1 (Luce & Almor, 2023) provides some initial suggestions as to how these components differ, the experiment was set up under the prevailing belief that perspective-taking is a unitary ability, and so it could not directly test what resources go into the different components of perspective-taking. The goal of Experiment 2 was to investigate (1) whether perspective-taking in comprehension should be decomposed into dissociable components, and (2) whether these components recruit different cognitive resources.

The crucial distinction suggested by Experiment 1 is between spatial and information-mismatch perspective-taking. By spatial perspective-taking, I mean tasks that require participants to 1) recognize the different viewpoint of their interlocutor and 2) calculate how a visible array of objects looks from that different viewpoint. This may

involve some kind of spatial rotation to be performed, in order to get from the participant's own viewpoint to the interlocutor's. This is equivalent to visual perspective-taking, at least in the tasks that do not require a linguistic component.

In contrast, I take information-mismatch perspective-taking to be used in tasks where the participant and the interlocutor know different things about the scene. In this way the perspective is not just about differing rotations or viewpoints, but about differing access to knowledge. In information-mismatch perspective tasks, the participant will have to maintain some kind of representation of which knowledge is in the common ground between themselves and the interlocutor, and which knowledge is privileged only to them. This ability comes down to theory of mind.

Based on Experiment 1, neither type of perspective-taking correlated with our measures of executive function. However, given previous literature exploring the relation between executive function measures and perspective-taking, it seems prudent to continue to test whether and when we can find this relationship. Of the two proposed types of perspective-taking, it would seem that information-mismatch perspective-taking would be more likely to recruit executive function, as it requires the inhibition of an egocentric point of view in order to limit the search of referents to only those in the common ground. The spatial perspective-tasks, on the other hand, do not necessarily require one to inhibit their own point of view, but to rotate and manipulate that point of view to match the interlocutor. Spatial perspective-taking also would potentially recruit more general spatial abilities that information mismatch perspective-taking would not need.

The decomposition of perspective-taking in comprehension into different components that recruit spatial and executive-function resources differently would have significant consequences for perspective-taking theory. It may serve to explain the puzzling inconsistency of the link between inhibition control and perspective-taking ability (Brown-Schmidt, 2009; Brown-Schmidt & Fraundorf 2015), as executive function may only be recruited for some types of perspective-taking but not others. It also would be interesting in the context of Ryskin et al. (2015)'s finding that perspective-taking does not appear to be domain general across production and comprehension, as such a finding would need to be more carefully evaluated if perspective-taking does turn out to be a cluster of different but related abilities.

Purpose and Research Questions

Experiment 2 was designed to answer three questions. The first question is whether perspective-taking is a unitary ability, or if it should be dissociated into different components. The second question is: if perspective-taking can be dissociated into components, is spatial vs information-mismatch perspective-taking the correct distinction. The third question is: do these components recruit different cognitive resources?

In order to answer these questions, we had the same set of participants complete four perspective-taking tasks. These four perspective tasks differed in whether they were spatial or non-spatial and linguistic or non-linguistic. We chose to vary the spatial dimension given the initial evidence from Experiment 1 that spatial and information-mismatch (non-spatial) perspective-taking may be different processes. We varied the linguistic dimension in order to see if linguistic perspective-taking recruits more domain

general theory of mind or has its own particular processes. The non-linguistic spatial task is equivalent to a VPT task, while the non-linguistic non-spatial task is equivalent to a typical theory of mind false belief task. This four way distinction allows us to explore how linguistic perspective-taking relates to more general VPT and ToM abilities.

We sought to answer our first question by having the same participants complete all four perspective-taking tasks and seeing how scores in one task correlate with scores in the other tasks. If perspective-taking is a unitary ability that is recruited the same in all four tasks, we would expect that task performance in all four tasks will correlate. If instead, performance doesn't correlate between some or all of the tasks, it would mean that perspective-taking is not specific enough of a concept to describe what participants are doing in each of the tasks.

We tested our second question through our split of spatial and non-spatial perspective-taking tasks. Because each of our non-spatial tasks involve information-mismatch perspective-taking, we predict that the two spatial tasks will correlate, the two non-spatial tasks will correlate, but that the spatial and non-spatial tasks will not correlate. If these results are obtained, it would confirm that the spatial/information-mismatch distinction is the correct decomposition of perspective-taking.

Finally, we tested our third question through the inclusion of five cognitive resource tasks. Two of these tasks are measures of general spatial ability and three are measures of components of executive function. We hypothesize that the spatial perspective-taking tasks will correlate with the spatial but not executive function resource tasks, while the information-mismatch perspective-taking tasks will correlate with the executive function but not spatial resource tasks. If these results are obtained, it would

mean that not only is there a spatial/information-mismatch distinction in perspective-taking, but that each of these components recruit different cognitive resources.

METHODOLOGY

Selection of Tasks

Four perspective-taking tasks were used in total. The first three of these tasks are the same as those used in Experiment 1 (Luce & Almor, 2023). These three tasks varied in their spatial and linguistic dimensions, but did not form a complete paradigm. Here, a fourth task is added in order to complete the paradigm. The three reused perspective-taking tasks are the “Tables” task (which is spatial and linguistic), the “Blocks” task (spatial, non-linguistic), and the “Cubbies” task (non-spatial, linguistic). To these three, we add a false belief task as an example of non-spatial and non-linguistic perspective-taking. All four of these tasks were chosen as they have been shown to be reliable tests of perspective-taking ability in previous studies.

In addition to the four perspective-taking tasks, we also included five tasks measuring general cognitive resources. Two of these are designed to measure spatial memory (the Corsi task and the Spatial Two-Back). These are added after Experiment 1’s finding that spatial perspective-taking tasks differed from other types of perspective-taking, in order to see if they recruit more general spatial resources.

Three tasks are also included that measure different aspects of executive function (working memory, inhibition control, and cognitive flexibility). We included these tasks based on conflicting evidence from previous studies that perspective-taking may recruit executive function. Typically either inhibition control or working memory are the executive function components considered most likely to predict perspective-taking

ability. However, here we also added a measure of the third part of executive function, cognitive flexibility, which was not included in Experiment 1, in order to check all aspects of executive control. We chose the Digit Span task to measure working memory, the Flanker task to measure inhibition control, and the Wisconsin Card Sort task to measure cognitive flexibility. The Digit Span and Flanker tasks were selected as they were both shown to correlate with perspective-taking in production by Wardlow (2013).

Participants

Two groups of participants were recruited. All participants had normal or corrected-to-normal vision and were native English speakers.

The first group of participants were students at the University of South Carolina, who were given psychology course credit for their participation. This group included 165 participants (136 female, 1 nonbinary, average age 20.38, standard deviation 2.96). 9 participants were dropped because they indicated during the pre-experiment survey that they were not native English speakers or did not have normal or corrected-to-normal vision.

The second group of participants were recruited online through the participant recruiter Prolific, and were paid $12 for their participation. These participants were limited to be under the age of 35. This group included 67 participants (26 female, 2 nonbinary, average age 28.55, standard deviation 5.45). 1 participant was dropped because they indicated during the pre-experiment survey that they were not native English speakers or did not have normal or corrected-to-normal vision.

Procedure

Participants completed four perspective-taking tasks followed by five resources tasks. The order of the perspective and resource tasks was balanced across participants. All tasks were presented online on the participant's computer screen, and participants responded either by clicking the screen or pressing a button. Each of the nine tasks are described below. The complete procedure took around 90 minutes to complete. The experiment was written and hosted through the Penn Controller for Ibex Farm (Zehr & Schwarz, 2018).

Pre-Experiment Survey:

Before completing the experiment, participants were asked to complete a short survey that gathered information about their age, gender, native language(s), whether they spoke other languages, whether they had normal or corrected-to-normal vision, and whether they had any problems seeing colors.

Perspective Task 1 "Tables" (Spatial, Linguistic):

The purpose of this task was to measure perspective-taking that requires both linguistic and spatial components. This task was adapted from experiment 2 of Ryskin et al. (2014). This task is the same as in Experiment 1 (Luce & Almor, 2023).

On each trial, participants were presented with a screen that included a 5x3 array of objects on a table as well as a cartoon character, Mr. Owl. Mr. Owl instructed the participants to click on one of the objects in the array, with directions in the form of: "Click on the duck to the right of the sponge." The participant's selection and reaction time was recorded.

There were two conditions. In the perspective-match condition, Mr. Owl appeared at the bottom half of the table, sharing the participant's view of the table. In the perspective-mismatch condition, Mr. Owl appears at the top edge of the table, facing the table and the participant from the opposite side.

The prediction was that reaction times would be faster in the perspective-match than in the perspective-mismatch conditions. This is because when Mr. Owl says "to the right", in the perspective-match condition this is to the participant's right as well, while in the perspective-mismatch condition, Mr. Owl's perspective mirrors the participant's and so the target object is to the left.

The critical item always had the same item to the left and right of it. In this way, the instructions given by Mr. Owl were ambiguous as there was the same object both to the left and right of the target, and information about the relative perspective had to be used.

Each participant saw 32 critical trials, 16 in each condition. Trials alternated between conditions. Each object appeared once in each condition as a target item, and once in each condition as the reference item. Distractor items were reused target/reference items from other trials. There were no filler trials.

There were 8 practice trials, 4 in each condition. These used different critical items. Feedback was given during practice trials. Before the practice trials, participants were given explicit instructions about the mirrored behavior of Mr. Owl (that when he was on the opposite side of the table and said "right" that would mean your left).

Perspective Task 2 "Blocks" (Spatial, Nonlinguistic):

The purpose of this task was to measure spatial perspective-taking in a non-linguistic way. It can be considered to measure visual perspective-taking (VPT). This task was adapted from Greenberg et al. (2013). This task is similar to Experiment 1 (Luce & Almor, 2023). However, whereas the "Blocks" task in Experiment one only included trials at a 90, 180, or 270 degree rotation, here we also include a control condition where there is no rotation.

On each trial of this task, participants were presented with a screen that includes four colored blocks (colored red, blue, green, and yellow) arranged in a diamond pattern, as well as a cartoon character, Mr. Owl, who viewed the pattern of blocks from a different angle (rotated either 0, 90, 180, or 270 degrees from the participant's perspective). On each trial, the participant must respond to the question "What does Mr. Owl see from where he is standing?" by selecting one of four provided answers. The pattern of blocks, questions, and answers are all presented on the same screen. We measured the selection that the participant made, as well as their reaction time.

The four choices shown to participants were pictures of different arrangements of the same four colored blocks. They include 1) the correct arrangement (as seen by Mr. Owl), 2) an arrangement that contained a structural error (where two blocks switched position) but the proper rotation, 3) an arrangement that contained a rotational error (rotated 90 instead of 180 degrees) but the proper structure, and 4) an egocentric arrangement that displayed the blocks in the same pattern as originally viewed by the participant. The four answers were presented in random order. Participants made their selection by clicking on the image of their chosen answer.

Each participant saw 28 critical trials, presented in a random order. The number of trials was spread evenly amongst the 4 possible positions of Mr. Owl (0, 90, 180, and 270 degree rotation from the participant's point of view). There were 6 practice trials, during which participants were given feedback about whether they made the correct selection. No feedback was provided during critical trials.

Perspective Task 3 "Cubbies" (Nonspatial, Linguistic):

The purpose of this task was to measure the linguistic information-mismatch perspective-taking. It is a version of the commonly-used cubbies referential communication task (e.g. Keysar et al., 2000; Heller et al., 2016). This task is the same as in Experiment 1 (Luce & Almor, 2023).

On each trial, participants were presented with a screen that included a 3x3 grid of cubby-holes which was filled with objects, as well as a cartoon character, Mr. Owl, who faced the participant from the other side of the cubby-holes. On critical trials, Mr. Owl would give the participant directions in the form of "Click on the large X", where X was one of the objects in the cubbies. The participant's selection and reaction time was recorded.

In all critical trials, there were four objects in the cubbies. One object that appeared in three different sizes (e.g., a small duck, a medium duck, and a large duck) and another object that appeared in only one size (e.g., a medium sponge).

There were two critical conditions, which varied who could see each object. In the perspective-match condition, all four of the objects were in common ground cubbies (visible to both the participant and Mr. Owl). In the perspective-mismatch condition, the largest object was in a privileged ground cubby (visible only to the participant and not

Mr. Owl) while the rest of the objects were in common ground cubbies. A “curtain” blocked the view of cubbies that were hidden to either the participant or Mr. Owl.

The prediction is that participants will take longer to select an object in the perspective-mismatch condition than in the perspective-match condition. This is because in the perspective-mismatch condition, there is some ambiguity in Mr. Owl’s instruction to “Click on the large X”. From an egocentric perspective, this would be the largest object in the display. However, if the participant is considering Mr. Owl’s perspective, they would realize that the largest object is hidden from Mr. Owl in this condition and that he therefore must mean the medium object, which appears comparatively large based on the common ground context set. This conflict between the egocentric and allocentric perspective should slow participants compared to the perspective-match condition, where all the information in the scene is in the common ground.

Each participant saw 16 critical trials, 8 in each condition. Half of the critical trials in each condition used “small” as the descriptive adjective and the other half “large”. Each critical trial used a different target item. The positions of all objects and the cubby types were randomized on each trial. The trials were presented in random order, including fillers.

Each participant also saw 20 filler trials. 12 of the fillers manipulated the color of the objects rather than the size. The other 8 fillers directed the participants to click on one of the non-critical items (e.g., the singleton object, the medium object).

There were 8 practice trials, which used different target items. These were presented in random order, and feedback was given to the participant about whether they made the correct choice. These practice trials were used to enforce the idea that Mr. Owl

did not know what was in the privileged cubbies, and that the "correct" selection in perspective-mismatch cases was to select the medium object.

Perspective Task 4 "False-Belief" (Nonspatial, Nonlinguistic):

The purpose of this task was to measure participant's non-linguistic and non-spatial theory of mind. It is a version of a typical "Sally-Anne" false-belief task modified for adults by measuring reaction time in addition to accuracy (Baron-Cohen et al., 1985).

On each trial of this task, participants viewed a short story in images. In the first image, Mr. Owl placed an object in one of four containers on the screen. Mr. Owl then either (a) walked away and turned his back so that he could not see (perspective-mismatch condition) or (b) walked away but watched from afar, where he could still see (perspective-match condition). Mr. Penguin then appeared and either moved the object to another container or left it where it was. Mr. Penguin then left and Mr. Owl reappeared. The participant was asked: "Where will Mr. Owl search for his object?" The participant then selected one of the four containers. An example perspective-mismatch trial can be seen in Figure 3.1. The container that the participant selected and the reaction time were measured. The idea of this task is that participants will be slower to select their answer on the perspective-mismatch conditions, as they have to overcome their own curse of knowledge to select the hiding spot where Mr. Owl originally placed his object.

Each participant saw 32 critical trials, half in each condition, presented in a random order. The number of trials was spread evenly amongst the 4 possible hiding positions, such that in three fourths of the trials, Mr. Penguin would move the object, and in one fourth would keep the object in its original location. There were 16 different objects that Mr. Owl could hide, which were repeated once in each condition. Each

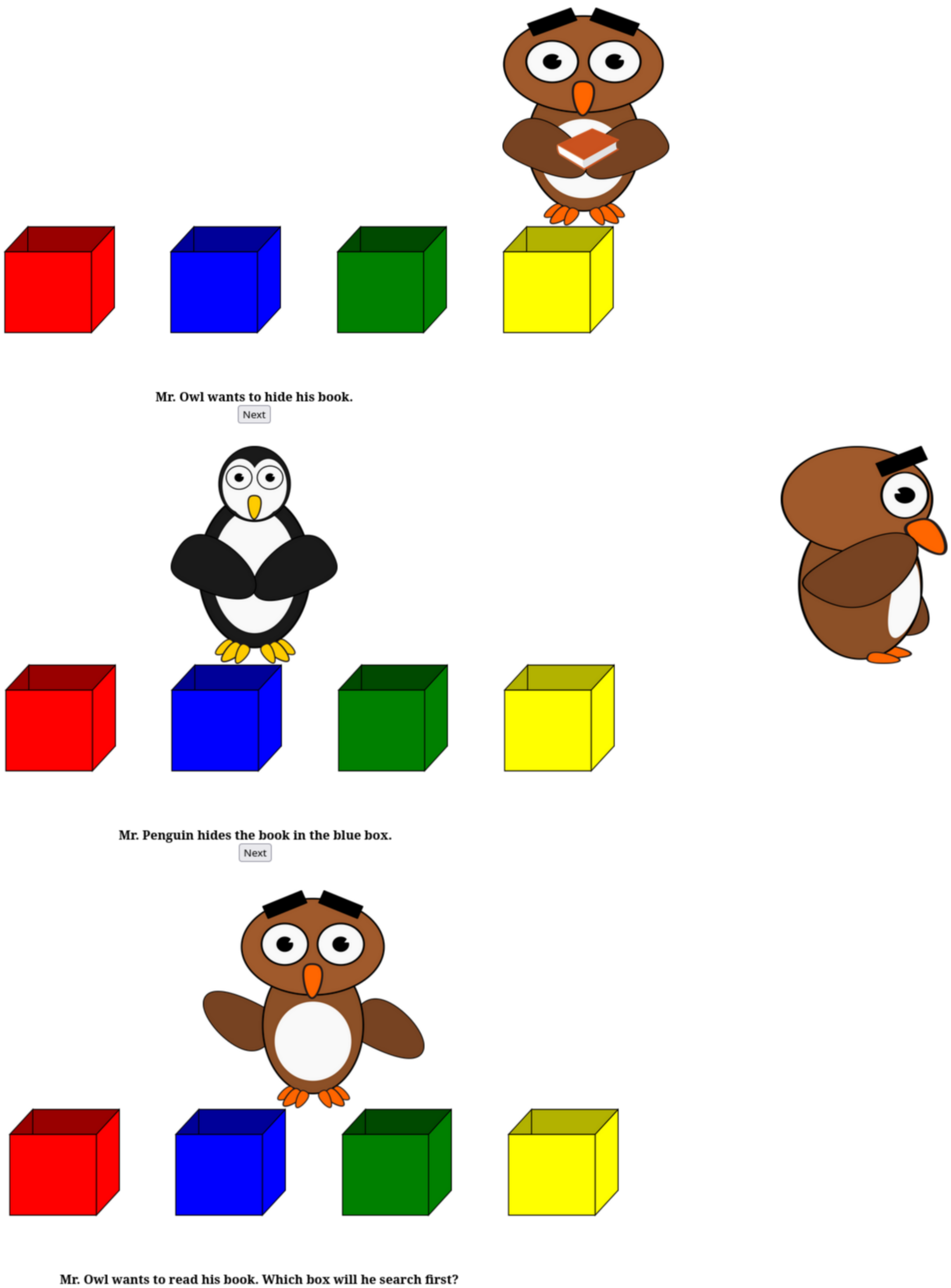

Figure 3.1 Three images of a perspective-mismatch trial of the "False Belief" task

participant saw 8 practice trials before beginning the regular trials. Feedback was provided during the practice but not the regular trials.

Resource Task 1 "Corsi" (Spatial Memory):

The purpose of this task was to measure the spatial memory of the participant. It is a version of the Corsi task (Corsi, 1972).

On each trial of this task, participants were presented with a screen that included 9 boxes. A random series of boxes were then highlighted for a brief moment one after the other on the screen. The participant's task was then to click on the boxes in the order that they were highlighted. The span length of boxes that the participant must remember started at 2 and increases by 1 every two trials that they correctly recalled. When the participant got two trials incorrect in a row, or reached the max span length of 9, the forward version of the task ended and the task was repeated, but with participants now having to click on the spans in reverse order to how they were shown. Each participant completed 4 practice trials before both the forward and backward trials. Practice trials included two trials at span length 2 and two at span length 3. Feedback was provided during the practice trials. Their last correct span length both forward and backwards was recorded. The sum of the participant's forward and backward Corsi span was used as their score for this task.

Resource Task 2 "Two-Back" (Spatial Memory):

The purpose of this task was to measure the visuospatial memory of the participant. It is a spatial version of the Two-Back task (Kirchner, 1958).

Participants were presented with a blank screen. On each trial, a single red square would appear on the screen for 5000 ms. The square would appear in one of 9 preset

locations. After the 5000 ms, the square would disappear and participants were asked to respond by pressing the ‘f’ key if the square just seen does not match the location of the square from two trials ago, and the ‘j’ key if it did. Participants had 4000 ms to respond before the next trial would begin whether they responded or not. Feedback was provided (“Correct/Incorrect”) for each trial during the task, after the participant responded but during the 4000 ms response window.

Each participant saw the same sequence of 60 trials, of which 17 trials were hits (where the location was the same as two trials ago). Participants’ accuracy was measured. Before the task, participants completed 12 practice trials. If participants missed at least two practice trials, they were required to complete the 12 practice trials a second time, in order to ensure that they understood the task. Feedback was provided during practice and regular trials.

Resource Task 3 “Flanker” (Inhibition Control):

The purpose of this task was to measure the inhibition control of the participant. It is a version of the Flanker task (Eriksen & Eriksen, 1974).

On each trial of this task, participants were presented with a screen that included 5 arrows, which faced either left or right. Their goal was to attend to only the central arrow, and to respond with the left arrow key if the arrow faced left and the right arrow key if the arrow faced right. There were two conditions, in the congruent condition, the four “flanker” arrows faced the same direction as the central arrow that the participant was told to attend to. In the incongruent condition, the four flanker arrows faced the opposite direction as the central arrow. Both the accuracy and the reaction time of the participants were measured.

The prediction is that participants are less accurate and slower to respond to the incongruent condition trials, as the distractor arrows interfere in the incongruent but not congruent condition.

Each participant saw 120 trials, of which 90 were congruent trials and 30 were incongruent trials. The order of the trials was randomized. There were also 10 practice trials. Feedback was provided during the practice trials.

Resource Task 4 "Digit Span" (Working Memory):

The purpose of this task was to measure the working memory capacity of the participant. It is a version of the backward digit span task.

On each trial of this task, participants were asked to remember a series of digits, in reverse order from how they were presented. Single digits were presented on the screen in sequence for 500ms each. After a sequence was presented, the participant saw the digits 1-9 on the screen, and had to click on the digits in the reverse order that they saw them. The trials continued until the participants failed to input the correct sequence twice in a row. The span length began at 2. Every two trials, the span length increased by one. The final correct span length that the participant was able to remember was recorded. There were 4 practice trials, two at span length 2 and two at span length 3. Feedback was provided during the practice trials.

Resource Task 5 "Wisconsin Card Sort" (Cognitive Flexibility):

The purpose of this task was to measure the cognitive flexibility of the participant. It is a version of the Wisconsin Card Sort task (Grant & Berg, 1948).

On each trial of this task, participants saw four central cards and one matching card. Their goal was to match their card with one of the four central cards, by clicking on

the card they think is a match. They were then given feedback about whether the card they clicked on was a match or not.

Each card had a number of colored shapes on them. The shapes could be stars, squares, triangles or circles. The colors were red, green, blue, and yellow. The number of shapes was one, two, three, or four.

In order to complete the task, the participant must discover which rule is currently required in order for the cards to match. Cards could match based on color, shape, or number. Every ten trials the rule changed, and the participant had to figure out what the new rule was. The total number of errors (incorrect matches) was measured, as well as the number of preservation errors (matches that are made based on the previous rule and not the current rule). Participants with better cognitive flexibility should make less errors, as they find it easier to adapt to the changing rules.

Each participant completed 120 trials. Each participant also completed 15 practice trials, during which the rule changed every 5 trials. Feedback was given during both practice and regular trials.

RESULTS

Perspective Task 1 "Tables" (Spatial, Linguistic):

Our dependent variable of interest for the Tables task was reaction time for correct responses. Participants who had accuracy less than 60% were excluded from the analysis. This affected 10 participants. From the remaining 212 participants, we removed all incorrect responses from the analysis (3.9% of responses). We removed trials with RTs that were above 60000 ms (0.3%). For each participant, we then removed outlier RTs that were above or below 2.5 standard deviations from the participant's mean

(2.9%). After outlier removal, the mean RTs were 4898 ms (SD: 3476 ms) for the perspective-match condition, and 5494 ms (SD: 3823 ms) for the perspective-mismatch condition.

Log[4] RTs were analyzed using linear mixed effects modeling, with the lme4 package (Bates et al., 2015) in R (R Core Team, 2023). Condition (perspective-match vs perspective-mismatch) was entered as a fixed effect. Subjects and Items were entered as random effects, using the maximal random effects structure that allowed the model to converge (Barr et al., 2013). Results of the full model are reported in Table 3.1. As can be seen from the table, perspective-mismatch trials were responded to more slowly than perspective-match trials.

Perspective Task 2 "Blocks" (Spatial, Nonlinguistic):

Our dependent variable of interest for the Blocks task was reaction time for correct responses. Participants who had accuracy less than 60% were excluded from the analysis. This affected 13 participants. From the remaining 209 participants, we removed all incorrect responses from the analysis (3.7% of responses). We removed trials with RTs that were above 60000 ms (0.2%). For each participant, we then removed outlier RTs that were above or below 2.5 standard deviations from the participant's mean (2.9%). After outlier removal, the mean RTs were 3912 ms (SD: 2315 ms) for the perspective-match condition, and 5115 ms (SD: 3304 ms) for the perspective-mismatch condition.

---

[4] We use log RTs instead of raw RTs due to the extreme right skew of the data which would violate assumptions of the linear regression model.

Table 3.1 E2 Parameter Estimates for Perspective Task 1: Tables

Fixed Effects	β	SE	*t*	*p*
(Intercept)	8.399	0.036	234.336	0.000*
perspective-mismatch	0.118	0.017	6.996	0.000*
Random Effects	**Variance**			
Subject				
(Intercept)	0.076			
perspective-mismatch	0.003			
Subject-Source				
(Intercept)	0.002			
Item				
(Intercept)	0.003			
Observations: 6306	Groups:	Subjects: 212	Items: 64	Subject-Source: 2

Significant effects at a $p \leq .05$ level are marked with an *.

Log RTs were analyzed using linear mixed effects modeling, with the lme4 package (Bates et al., 2015) in R (R Core Team, 2023). Condition (perspective-match vs 90-and-270-mismatch vs 180-mismatch) was entered as a fixed effect. Subjects and Items were entered as random effects, using the maximal random effects structure that allowed the model to converge (Barr et al., 2013). Results of the full model are reported in Table 3.2. As can be seen from the table, perspective-mismatch trials were responded to more slowly than perspective-match trials.

Perspective Task 3 "Cubbies" (Nonspatial, Linguistic):

Our dependent variable of interest for the Cubbies task was reaction time for correct responses. Participants who had accuracy less than 60% were excluded from the analysis. This affected 20 participants. From the remaining 202 participants, we removed all incorrect responses from the analysis (3.7% of responses). We removed trials with RTs that were above 60000 ms (<0.1%). For each participant, we then removed outlier RTs that were above or below 2.5 standard deviations from the participant's mean (2.3%). After outlier removal, the mean RTs were 2428 ms (SD: 1006 ms) for the perspective-match condition, and 2672 ms (SD: 946 ms) for the perspective-mismatch condition.

Log RTs were analyzed using linear mixed effects modeling, with the lme4 package (Bates et al., 2015) in R (R Core Team, 2023). Condition (perspective-match vs perspective-mismatch) was entered as a fixed effect. Subjects and Items were entered as random effects, using the maximal random effects structure that allowed the model to converge (Barr et al., 2013). Results of the full model are reported in Table 3.3. As can be

Table 3.2 E2 Parameter Estimates for Perspective Task 2: Blocks

Fixed Effects	β	SE	*t*	*p*
(Intercept)	8.198	0.039	212.27	0.000*
perspective-mismatch	0.225	0.014	16.28	0.000*
Random Effects	Variance			
Subject				
(Intercept)	0.091			
perspective-mismatch	0.020			
Subject-Source				
(Intercept)	0.002			
Item				
(Intercept)	0.000			
Observations: 5462	Groups:	Subjects: 209	Items: 21	Subject-Source: 2

Significant effects at a $p \leq .05$ level are marked with an *.

Table 3.3 E2 Parameter Estimates for Perspective Task 3: Cubbies

Fixed Effects	β	SE	*t*	*p*
(Intercept)	7.759	0.049	156.37	0.000*
perspective-mismatch	0.100	0.009	10.06	0.000*
Random Effects	**Variance**			
Subject				
(Intercept)	0.048			
perspective-mismatch	0.005			
Subject-Source				
(Intercept)	0.004			
Item				
(Intercept)	0.002			
Observations: 3039	Groups:	Subjects: 202	Items: 32	Subject-Source: 2

Significant effects at a $p \leq .05$ level are marked with an *.

seen from the table, perspective-mismatch trials were responded to more slowly than perspective-match trials.

Perspective Task 4 "False Belief" (Nonspatial, Nonlinguistic):

Our dependent variable of interest for the False Belief task was accuracy. The mean accuracy was 0.96 (SD: 0.23) for the perspective-match condition, and 0.93 (SD: 0.30 ms) for the perspective-mismatch condition.

Accuracy was analyzed using logistic mixed effects modeling, with the lme4 package (Bates et al., 2015) in R (R Core Team, 2023). Condition (perspective-match vs perspective-mismatch) was entered as a fixed effect. Subjects and Items were entered as random effects, using the maximal random effects structure that allowed the model to converge (Barr et al., 2013). Results of the full model are reported in Table 3.4. As can be seen from the table, perspective-mismatch trials were responded to less accurately than perspective-match trials.

Resource Task 1 "Corsi" (Spatial Memory):

Our dependent variable of interest for the Corsi task was the sum of each participant's forward and backward Corsi span. The mean sum of forward and backward Corsi spans was 10.9 (SD: 2.9).

Resource Task 2 "Two-Back" (Spatial Memory):

Our dependent variable of interest for the Two-Back task was percentage accuracy. The mean accuracy was 0.71 (SD: 0.19).

Resource Task 3 "Flanker" (Inhibition Control):

Our dependent variable of interest for the Flanker task was reaction time for correct responses. Participants who had accuracy less than 60% were excluded from the

Table 3.4 E2 Parameter Estimates for Perspective Task 4: False Belief

Fixed Effects	β	SE	z value	*p*
(Intercept)	8.938	0.8734	10.233	0.000*
perspective-mismatch	-3.565	0.0915	-3.897	0.000*
Random Effects	Variance			
Subject				
(Intercept)	37.09			
perspective-mismatch	44.38			
Subject-Source				
(Intercept)	0.000			
Observations: 7104	Groups:	Subjects: 222	Subject-Source: 2	

Significant effects at a $p \leq .05$ level are marked with an *.

analysis. This affected 6 participants. From the remaining 216 participants, we removed all incorrect responses from the analysis (2.6% of responses). We removed trials with RTs that were above 60000 ms (<0.1%). For each participant, we then removed outlier RTs that were above or below 2.5 standard deviations from the participant's mean (2.6%). After outlier removal, the mean RTs were 620 ms (SD: 424 ms) for the congruent condition, and 816 ms (SD: 572 ms) for the incongruent condition.

Log RTs were analyzed using linear mixed effects modeling, with the lme4 package (Bates et al., 2015) in R (R Core Team, 2023). Condition (congruent vs incongruent) was entered as a fixed effect. Subjects and Items were entered as random effects, using the maximal random effects structure that allowed the model to converge (Barr et al., 2013). Results of the full model are reported in Table 3.5. As can be seen from the table, incongruent trials were responded to more slowly than congruent trials.

Resource Task 4 "Digit Span" (Working Memory):

The first two participants had their digit span scores removed due to an error in the experiment code that was fixed for later subjects. This left 220 participants for the digit span task. Our dependent variable of interest for the Digit Span task was the participant's backward digit span length. The mean span length was 6.4 (SD: 3.1).

Resource Task 5 "Wisconsin Card Sort" (Cognitive Flexibility):

Our dependent variable of interest for the Wisconsin Card Sort task was percentage accuracy. The mean accuracy was 0.75 (SD: 0.13).

Table 3.5 E2 Parameter Estimates for the Flanker Task

Fixed Effects	β	SE	*t*	*p*
(Intercept)	6.324	0.028	227.46	0.000*
incongruent	0.256	0.010	25.02	0.000*
Random Effects	**Variance**			
Subject				
(Intercept)	0.077			
incongruent	0.019			
Subject-Source				
(Intercept)	0.001			
Observations: 24594	Groups:	Subjects: 216	Subject-Source: 2	

Significant effects at a $p \leq .05$ level are marked with an *.

Correlations between Tasks

In addition to the above experimental effects, we also want to see if there are consistent individual differences in each task, and if these differences correlate across different tasks.

In order to carry out the correlation analysis, one measure was used to represent each task, using the same method as in Experiment 1 (Luce & Almor, 2023).

For Perspective Task 1: "Tables", Perspective Task 2: "Blocks", Perspective Task 3: "Cubbies", Perspective Task 4: "False Belief" and the Flanker task, we extracted subject slopes for condition from each of the mixed effects models, and used these subject slopes as the measure for how good the participant was at the task.

In order to assess the reliability of these model-based measures we split the data for each task into two halves based on odd or even trial numbers, and then fit a separate model for each half, using the same fixed and random effects structure as the full model. We then compared the correlation between the subject slopes for condition for each half (this process for assessing reliability was based on Ryskin et al. (2015), Rouder & Haaf (2019), and Staub (2021) and was used in Luce & Almor (2023)). The "Tables" and "Blocks" tasks were split by every two trials instead of every other trial, as the conditions alternated in a regular way.

The Spearman-Brown adjusted correlation between the by-subject slopes for conditions of the two models for the "Cubbies" task was p = .20. For the "Tables" task, the Spearman-Brown adjusted correlation was p = -.28. These both reflect poor reliability for these tasks. For the "Blocks" task, the Spearman-Brown adjusted correlation was p = .94. For the "False Belief" task, the Spearman-Brown adjusted correlation was p = .92.

And for the Flanker task, the Spearman-Brown adjusted correlation was p = .87. These three tasks were all very reliable.

For the Two-Back and Wisconsin Card Sort tasks, reliability was calculated by the Spearman-Brown adjusted correlation based on 5000 random splits of the data. For the Two-Back task, this correlation was p = 0.92, with a 95% confidence interval of [0.9, 0.94]. For the WCS task, this correlation was p = 0.84, 95% CI [0.81, 0.84].

No reliability is provided for the Corsi and Digit-Span tasks as these tasks rely on the longest span length remembered rather than RT or accuracy of the individual trials.

Due to our accuracy threshold, each task had a different number of participants, so for each correlation between pairs of tasks, we used the maximum number of participants that had data for both tasks. The number of participants contributing to each correlation is shown underneath the correlation coefficients in Table 3.6 and Table 3.7. Table 3.6 shows the correlations between the perspective-taking tasks and the individual differences measures. Table 3.7 shows the correlations amongst the individual differences measures. The reliability of each task is also shown in parentheses underneath the task.

DISCUSSION

We found the expected effects of perspective in all four perspective-taking tasks. Response times were significantly slower in perspective-mismatch trials compared to perspective-match trials for the “Tables”, “Blocks”, and “Cubbies” tasks. And participants were significantly less accurate in perspective-mismatch trials for the “False Belief” task. However, while there were strong experimental effects, there were again sparse correlations between the different perspective-taking tasks and between the perspective tasks and the individual differences measures.

Table 3.6 E2 Correlation Matrix for Perspective-Taking Tasks

	Tables (-0.28)	Blocks (0.94)	Cubbies (0.20)	False Belief (0.92)
Tables (-0.28)	---			
Blocks (0.94)	0.03 (*p* = 0.668 n = 205)	---		
Cubbies (0.20)	-0.06 (*p* = 0.377 n = 197)	**0.19* (*p* = 0.008 n = 193)**	---	
False Belief (0.92)	-0.12 (*p* = 0.083 n = 212)	-0.02 (*p* = 0.746 n = 209)	0.01 (*p* = 0.887 n = 202)	---
Corsi (-)	0.08 (*p* = 0.257 n = 212)	0.10 (*p* = 0.167 n = 209)	0.02 (*p* = 0.786 n = 202)	0.13 (*p* = 0.051 n = 222)
Two Back (0.92)	-0.04 (*p* = 0.612 n = 212)	**0.21* (*p* = 0.002 n = 209)**	0.08 (*p* = 0.232 n = 202)	-0.03 (*p* = 0.656 n = 222)
Flanker (0.87)	-0.07 (*p* = 0.286 n = 210)	0.14 (*p* = 0.051 n = 207)	0.07 (*p* = 0.303 n = 202)	0.12 (*p* = 0.074 n = 216)
Digit Span (-)	0.04 (*p* = 0.531 n = 210)	0.01 (*p* = 0.928 n = 207)	0.06 (*p* = 0.414 n = 200)	0.03 (*p* = 0.676 n = 220)
WCS (0.84)	0.05 (*p* = 0.476 n = 212)	**0.21* (*p* = 0.002 n = 209)**	0.09 (*p* = 0.179 n = 202)	0.15 (*p* = 0.027 n = 222)

Significant correlations at an alpha of 0.05 corrected using the Benjamini-Hochberg procedure for 36 comparisons are marked with an *. The reliability appears in parentheses under the name of each task.

Table 3.7 E2 Correlation Matrix for Individual Differences Measures

	Corsi (-)	Two Back (0.92)	Flanker (0.87)	Digit Span (-)	WCS (0.84)
Corsi (-)	---				
Two Back (0.92)	**0.18* (*p* = 0.008 n = 222)**	---			
Flanker (0.87)	0.16 (*p* = 0.022 n = 216)	**0.28* (*p* < 0.001 n = 216)**	---		
Digit Span (-)	0.14 (*p* = 0.039 n = 220)	**0.23* (*p* < 0.001 n = 220)**	0.06 (*p* = 0.344 n = 214)	---	
WCS (0.84)	**0.30* (*p* < 0.001 n = 222)**	**0.31* (*p* < 0.001 n = 222)**	0.12 (*p* = 0.082 n = 216)	**0.27* (*p* < 0.001 n = 220)**	---

Significant correlations at an alpha of 0.05 corrected using the Benjamini-Hochberg procedure for 36 comparisons are marked with an *. The reliability appears in parentheses under the name of each task.

Internal Consistency of the Perspective-Taking Tasks

We found that the "Tables" and "Cubbies" tasks each had poor reliability while the "Blocks" and "False Belief" tasks had good reliability. This split is interesting as the two that require a linguistic component are the two with poor reliability, while "Blocks" which measures visual perspective-taking and "False Belief" which measures theory of mind are both reliable. This split does suggest that maybe there is something about linguistic perspective-taking in comprehension that is just not reliable for use in individual differences studies. While we consistently get experimental effects from these tasks, this study adds to a now consistent problem with poor reliability for these perspective-taking during comprehension tasks (Brown-Schmidt & Fraundorf, 2015; Ryskin et al., 2015, Luce & Almor, 2023).

It could be that people are simply not better or worse at perspective-taking, which might explain why we get an experimental effect but no individual differences. However, this poses a problem for the many studies that seek to find which cognitive resources affect how quickly and how often people take into account perspective information. If we don't see people with better or worse perspective-taking ability, then this kind of study is not going to get at anything real. This is especially concerning as the most prototypical perspective-taking task, the "Cubbies" task, has now consistently shown to have poor internal reliability.

Our results also show a divergence from Experiment 1, where the "Tables" task showed acceptable reliability. The only differences between the two experiments were different participants, minor changes to how we removed outliers, and the use of log response time instead of raw response time. The use of log response time as the

dependent measure is likely a large part of the lack of reliability in Experiment 2. However, it was necessary to log transform the RT data as the very strong right skew of the data would violate the assumptions of the linear regression models. In Experiment 1, as we had fewer participants, it was not possible to get the models to converge using log RT as the dependent measure, so we used raw RT. However this choice would have inflated the reliability of the task. And we see the consequences of this in Experiment 2 where the "Tables" task shows very poor reliability.

It is also surprising that we see such a difference in reliability between the "Tables" and "Blocks" tasks given that the tasks seem to require such similar ability. Both require either a 0 or 180 degree rotation to "see" some pattern (although the "Blocks" task also requires 90 and 270 degree rotations). The main difference seems to be the linguistic aspect of the "Tables" task, where Mr. Owl says "the X to the right/left of the Y". It could be that this egocentric referring expression is the source of the decreased reliability.

Relationship between Perspective-Taking Tasks

The only correlation between perspective-taking measures that we see is between the "Blocks" and the "Cubbies" task. This relationship goes against our predictions of a spatial/information-mismatch (non-spatial) split in perspective-taking, as the "Blocks" task is a spatial non-linguistic VPT task, and the "Cubbies" is a non-spatial linguistic information-mismatch task. It also is quite different to the results of Experiment 1, which found only one correlation between the spatial tasks "Blocks" and "Tables".

This result means that we likely want to reject a straightforward split of perspective-taking into spatial and information-mismatch components, as we suggested to

explain Experiment 1. The mixed results and low reliabilities of some tasks in Experiment 1 and 2 make it difficult to put forward an ironclad explanation of both experiments. It may be that perspective-taking is indeed separated into components, but that these components are far more granular than a spatial/information-mismatch split.

Relationship between Perspective-Taking and Individual Differences Measures

Experiment 2 included three measures of executive function (one each of working memory, inhibition control, and cognitive flexibility) and two measures of spatial memory. Like Experiment 1, we again find that working memory (as measured by the backward digit span) and inhibition control (as measured by the flanker task) did not correlate with any of the perspective-taking tasks. This is the second null result we have found for the two cognitive resources thought most likely in the literature to predict perspective-taking ability. Like previous null results, this could be explained at least in part by the poor reliability of some of our perspective-taking tasks. However the digit span and flanker tasks also did not correlate with “Blocks” (VPT) or “False Belief” (ToM), which showed high reliability.

In fact, only one of our four perspective-taking tasks correlated at all with the individual differences measures. That was the “Blocks” task, which correlated with the Two Back task for spatial memory and the Wisconsin Card Sort task for cognitive flexibility. These two correlations provide further evidence against our original strong prediction of a spatial/information-mismatch split in perspective-taking. We had predicted that only the spatial tasks would correlate with the two measures of spatial memory and that only the non-spatial information-mismatch tasks would correlate with the executive function measures. However in the “Blocks” task, we see a correlation with

only one of the two spatial measures as well as one of the executive function measures. This pattern of results means that the resources that go into perspective-taking are clearly more granular than we predicted.

The fact that none of the individual differences measures correlated with “Tables” and “Cubbies” is relatively unsurprising given that the low reliabilities of these tasks attenuate any potential correlation. However, it is somewhat surprising that none of the correlations with the “False Belief” task reached significance given the high reliability of that task.

In conclusion, in Experiment 2 we found (1) further evidence of poor reliability in some of the most common perspective-taking in comprehension tasks, (2) no evidence that inhibition control and working memory predict perspective-taking ability, and (3) that our prediction of a simple spatial/information-mismatch split in perspective-taking is likely not tenable due to the correlation between the “Blocks” and “Cubbies” tasks. Instead, we may need to make even more granular distinctions in perspective-taking. The conclusions of both experiments are developed further in the final chapter.

CHAPTER 4

CONCLUSION

RELIABILITY OF PERSPECTIVE-TAKING IN COMPREHENSION

In both experiments, we found problems with the reliability (the internal consistency) of perspective-taking in comprehension tasks. The "Cubbies" task was unreliable in both Experiment 1 and 2. The "Tables" task was unreliable in Experiment 2. These results further the evidence from previous research (Brown-Schmidt & Fraundorf, 2015; Ryskin et al., 2015) that shows that perspective-taking in comprehension may simply not be reliable: meaning that there are not good perspective-takers who consistently perform faster or more accurately at perspective-taking tasks and poor perspective-takers who consistently perform slower or less accurately.

We did, in all perspective-taking tasks in both experiments, find a consistent effect of perspective, such that perspective-mismatch trials took longer or were less accurate than perspective-match trials. This means that the lack of internal consistency of some of these tasks has no effect on the experimental effect. It is only showing that there is a lack of individual differences in perspective-taking ability.

RT vs Eyetracking as the Dependent Measure

One important difference in our experiments compared to previous studies was the use of RT for the dependent measure in the perspective-taking during comprehension tasks. Many perspective-taking in comprehension tasks, especially the prototypical "Cubbies" task, use eyetracking, and have a dependent measure based on the proportion

of looks to target over a distractor. While RT is not the most common measure, we believe it is both valid and has some utility that eyetracking does not.

In terms of the validity of using RT, many perspective-taking tasks have shown a difference in accuracy (Keysar et al., 2000; Greenberg et al., 2013, Ryskin et al. 2014). RT should be more sensitive than these accuracy differences and so we should be able to see a difference in RT between perspective-match and perspective-mismatch conditions. Indeed, this is what we saw in both experiments for all of the perspective-taking tasks that used RT as the dependent measure. It is good to confirm that we still see these perspective-taking effects when using different measures. Furthermore, the RT effect sizes for all of the perspective-taking tasks were on the expected order of magnitude of hundreds of milliseconds.

The benefits to using RT instead of eyetracking are threefold. First, we have seen in Brown-Schmidt and Fraundorf (2015) and Ryskin et al. (2015) that perspective-taking in comprehension has been unreliable. Both of these studies used an eyetracking measure for perspective-taking in comprehension. Eyetracking data tends to be messy, and it is possible that this messiness contributes to the poor reliability of these perspective-taking in comprehension tasks. This possibility seems even more likely given that Ryskin and colleagues also find that perspective-taking in production is reliable. Their production task is very similar to the comprehension task (both versions of the “Cubbies” task, but from opposite sides - either being the producer or the comprehender). But the dependent measure in the production task is simply how many adjectives get produced while the comprehension measure relies on eyetracking while listening to instructions.

Therefore, it is not totally clear whether, as Ryskin et al. (2015)'s results indicate, there is really a difference between perspective-taking in production and comprehension, or whether the different way that their experiment measured perspective-taking in production and comprehension interfered with their inferences. By switching to RT in our study, we can now be more confident that it is not just eyetracking that is causing the poor reliability, as we also see poor reliability with RT.

The second benefit is that RT should be better suited to the individual differences part of the study, as it more closely gets at cognitive demand than does eyetracking. This means that an RT measure of perspective-taking should be better for studying whether perspective-taking is predicted by inhibition control and working memory. The fact that we find no evidence for a link with these executive function measures contributes another null result to the literature where this link has been shown to be tenuous. And our study also allows us to be more confident in this result as we are getting at the construct of perspective-taking using multiple different measures, rather than always sticking with eyetracking.

Finally, using RT as the dependent measure allowed us to deploy the experiment online and recruit far more participants than would be feasible to recruit in person. The increased power from a large group of participants was necessary in order to draw conclusions about individual differences.

No Evidence of a Link with Inhibition Control or Working Memory

A second contribution of our study is another null result for a link between perspective-taking and inhibition control or working memory. In both of our experiments, with large numbers of participants and multiple perspective-taking tasks, we saw no

evidence that inhibition control or working memory measures correlated with perspective-taking ability in comprehension. While, as null results, these should still be considered with caution, our two experiments together with previous studies' results, as well as the general publication bias likely responsible for other such results not being published, lend considerable credence to the conclusion that there may not be reliable connections between inhibition control, working memory, and perspective-taking.

One of the most startling findings of our study was that we found poor reliability in both experiments for the "Cubbies" task, as a version of this task is very commonly used in perspective-taking studies, both for production and comprehension. The lack of reliability of this task is a potential limiting factor in studies that use it to find a link between executive function and perspective-taking ability.

Whether or not there are individual differences in perspective-taking ability, and whether executive function predicts this ability has important theoretical consequences for models of perspective-taking. Take for example the model in Heller et al. (2016), where people interpret referring expressions by considering both egocentric and common ground domains of referents. In this model, α is the weight given to one domain over the other, and it is influenced by situational factors (such as what other objects are present and how good of a referential fit they are). Heller and colleagues only focus on situational factors and don't claim that individual differences in perspective-taking may influence this weighting of egocentric and common ground domains.

It seems reasonable to me that individual differences in perspective-taking ability might influence this weighting of domains, given the persistent claim that a person's inhibition control or working memory capacity may influence how good of a perspective-

taker they are. If someone were a poor perspective-taker, and had a hard time inhibiting their egocentric perspective, it may be harder to get other situational factors to weight α more towards the common ground domain, compared to someone who had better inhibition control or working memory.

However, given the results of this study, I would argue that there is not sufficient evidence to incorporate individual differences into the weighting of domains in any model of perspective-taking, at least for comprehension. The consistent lack of individual differences in perspective-taking during comprehension in this study and in previous research makes this ill-advised.

THE PLACE OF PERSPECTIVE-TAKING IN COMPREHENSION

It is more than spatial vs information-mismatch

The results of Experiment 1, where the two spatial perspective-taking tasks correlated with one another but not with the information-mismatch perspective-taking task, caused us to hypothesize that perspective-taking is better thought of not as a single unified ability, but as two components: spatial and information-mismatch (non-spatial). This result also corresponds to our intuitions that spatial and information-mismatch perspective-taking tasks seem like they are asking different questions. The spatial tasks ask how to resolve a referring expression when the same set of objects looks different from different perspectives, while information-mismatch tasks ask how to resolve a referring expression when the speaker and listener have a different set of objects known to them. These spatial tasks seem more related to VPT while the information-mismatch tasks are more related to theory of mind. These intuitions also caused us to believe that

research in perspective-taking has not adequately acknowledged this division between different types of tasks, which may cause divergent results.

However, given the results of Experiment 2, it seems that our initial hypothesis of a single spatial/information-mismatch split in perspective-taking is likely not tenable. The only correlation between perspective tasks in E2 was between the "Blocks" task, a spatial VPT task, and the "Cubbies" task, a linguistic information-mismatch task. This pattern of results goes against the predictions of a spatial/information-mismatch split, where we would expect "Blocks" and "Tables" to correlate with one another, and "Cubbies" and "False Belief" to correlate with one another.[5]

If we reject a spatial/information-mismatch split, there are two likely alternatives for how we should think of perspective-taking in comprehension. The first is that it is a single unified ability and there is no point in distinguishing components of perspective taking, and the second is that it contains multiple components, but that these components are more granular than our spatial/information-mismatch split.

---

[5] The one interesting difference that may come out of the spatial/information-mismatch split in the tasks stems from us using RT as our dependent measure. We saw that the spatial tasks took longer overall compared to the information-mismatch tasks, and that they had an increased effect size of perspective-taking. The E1 and E2 "Cubbies" task had a mean perspective-match baseline RT in the order of 2 to 3 thousand milliseconds, and a difference between the mean perspective-mismatch and perspective-match times of around 200 ms. The E1 and E2 "Tables" and "Blocks" tasks had mean perspective-match times of around 4 to 6 thousand milliseconds, with differences between the conditions between 500 and 1200 ms. So the two spatial tasks took longer and had a larger difference between the conditions. The overall difference in RT probably reflects that the spatial tasks were simply more difficult than the "Cubbies" task. The increased difference between the two conditions is also going to be influenced by the overall difficulty of the task. But it is also possible that the spatial tasks took longer in the perspective-mismatch trials as performing the spatial rotation of the image takes more time than the "Cubbies" task. Although given the lack of correlation between the two spatial tasks in E2, it is impossible to claim with any certainty that this is what is happening.

That perspective-taking in comprehension is a single unified ability view serves as the null hypothesis given that it is the current default of the field, and we certainly don't provide enough evidence to reject it outright here. However, there is some room for doubt. First is that Ryskin et al. (2015) claims that there is no domain general perspective-taking and that perspective-taking ability in production and comprehension do not correlate with one another. If we do not see a domain general perspective-taking between production and comprehension, that could mean that "perspective-taking" is the wrong level to explain what is happening in those tasks. The same may be true of perspective-taking in comprehension tasks. However the lack of reliability of these tasks makes testing this much more complicated.

The lack of reliability of some but not all of the perspective-taking tasks also serves as a second reason we might doubt that perspective-taking in comprehension is a unified ability. The "Blocks" and "Tables" tasks in E2 are very similar to one another, but one showed solid internal consistency while the other did not. Both of these tasks require considering how a set of objects are arranged from a 180-degree mirrored perspective, with the major differences between them being the linguistic component of the "Tables" task, and the addition of 90 and 270 degree rotations in the "Blocks" task. This may mean that some of the granular components of this task show reliable individual differences while others do not. The inclusion of the language component in the "Tables" task may be what makes the difference.

Perspective-Taking, ToM, and VPT

One reason to think that there may be granular components is the intuition, as well as evidence from E1, that spatial perspective-taking seems to rely on VPT while

information-mismatch perspective-taking relies on theory of mind. While VPT and ToM are overlapping abilities, they also seem to have different components to them. It could be that we are seeing the overlap in E2 (where "Cubbies" and "Blocks" correlate) and the differing components in E1 (where "Tables" and "Blocks" correlate). Both VPT and ToM are also separated into levels, which may recruit different cognitive resources (Michelon & Zacks, 2006; Samson & Apperly, 2016) and are developed at different ages (Flavell et al., 1981).

As Heller and Brown-Schmidt (in press) point out, most perspective-taking research has taken place with tasks that utilize level 1 theory of mind, such as the "Cubbies" task. Meaning that these tasks rely on one interlocutor to have an absence of belief, to not know about, about a certain object. Only a handful of studies (Hanna et al., 2003; Mozuraitis et al., 2018) have looked at perspective-taking that incorporates level 2 ToM, where the two interlocutors have different beliefs about an object. If these perspective-taking tasks are underpinned by different levels of ToM, which recruit different resources, how can we so easily group these together as perspective-taking tasks? It is important to consider whether we are talking about level 1 or level 2 ToM in our perspective-taking tasks. And whether our level 2 tasks are really about differing beliefs (ToM) or are about different spatial viewpoints (VPT). These are issues that need to be clarified in order to continue to make progress on whether there are reliable individual differences in perspective-taking ability during comprehension.

Limitations and Future Directions

This study is instructive in the difficulties of looking at individual differences in this area, where the typically used perspective-taking in comprehension tasks have low

reliabilities or the reliabilities are not reported and there are usually only one or few measures of each construct. Our study to some extent suffers from these limitations, with poor reliability in some of the perspective-taking tasks. And while we made a contribution by having participants complete multiple perspective-taking in comprehension tasks, we only included one measure of inhibition control and working memory in each experiment, which limits our confidence to make very strong claims about executive function. However, many individual differences studies of perspective-taking in comprehension suffer from these problems, and we hope that this study will be instructive in increasing the quality of future individual differences studies in this area. Especially as there exists a potential for bias in publishing only the positive results of a link between executive function and perspective-taking.

As discussed in the limitations of Experiment 1, both experiments also are limited by the online modality, which required the tasks to be non-interactive, use a non-human interlocutor, and use written instructions, all of which may attenuate the perspective-taking effects. While these limitations should be noted, we believe it was worth making this trade-off for the possibility of recruiting large numbers of participants to better understand individual differences during perspective-taking, which few studies have sought to do. We hope that the consistent effects of perspective-taking found in our study, using an online, reaction-time based tasks, encourages future studies to both recruit larger numbers of participants for individual differences research and to try different dependent measures to get at the construct of perspective-taking.

Finally, we foresee future research that thinks of perspective-taking in a more granular way. Our experiment 2 included a linguistic perspective-taking task recruiting

level 1 ToM ("Cubbies"), a linguistic perspective task recruiting level 2 VPT ("Tables"), a level 2 ToM task ("False Belief") and a level 2 VPT task ("Blocks"). While our spatial/information-mismatch hypothesis failed to capture the full pattern of results in our study, we think that there needs to be careful distinctions made in future research between tasks that recruit ToM or VPT, as well as the level of ToM or VPT, as all involve different processes. This dissertation provides only a first step in this direction, but we hope that by decomposing perspective-taking in future work we might be better able to (1) choose between the theories of perspective-taking that propose different underlying mechanisms, (2) explain why perspective-taking may not be domain-general, (3) explain why perspective-taking in comprehension tasks are often not reliable, and (4) explain which cognitive resources underpin perspective-taking.

REFERENCES

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a “theory of mind”?. *Cognition*, *21*(1), 37-46.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. Journal of memory and language, 68(3), 255-278.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1-48. doi:10.18638/jss.v067.i01

Birch, S. A., & Bloom, P. (2007). The curse of knowledge in reasoning about false beliefs. Psychological Science, 18(5), 382-386.

Brown, G. (1995). Speakers, listeners and communication: Explorations in discourse analysis. Cambridge University Press.

Brown-Schmidt, S., Gunlogson, C., & Tanenhaus, M. K. (2008). Addressees distinguish shared from private information when interpreting questions during interactive conversation. Cognition, 107(3), 1122-1134.

Brown-Schmidt, S. (2009). The role of executive function in perspective taking during online language comprehension. Psychonomic bulletin & review, 16(5), 893-900.

Brown-Schmidt, S. (2012). Beyond common and privileged: Gradient representations of common ground in real-time language use. Language and Cognitive Processes, 27(1), 62-89.

Brown-Schmidt, S., & Fraundorf, S. H. (2015). Interpretation of informational questions modulated by joint knowledge and intonational contours. Journal of Memory and Language, 84, 49-74.

Clark, H. H., Marshall, C. R., (1978). Reference diaries. In: Waltz, D.L. (Ed.), *TINLAP-2: Theoretical Issues in Natural Language Processing-2*. Association for Computing Machinery, New York, pp. 57–63.

Clark, H. H., & Marshall, C. R. (1981). Definite knowledge and mutual knowledge. Elements of Discourse Understanding. Cambridge University Press. 10-63.

Clark, H. H. (1996). Using language. Cambridge university press.

Corsi, P. M. (1972). Human memory and the medial temporal region of the brain. Dissertation Abstracts International, 34 (02), 819B.

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. Perception & psychophysics, 16(1), 143-149.

Flavell, J. H., Everett, B. A., Croft, K., & Flavell, E. R. (1981). Young children's knowledge about visual perception: Further evidence for the Level 1–Level 2 distinction. *Developmental psychology, 17*(1), 99.

Grant, D. A., & Berg, E. (1948). A behavioral analysis of degree of reinforcement and ease of shifting to new responses in a Weigl-type card-sorting problem. *Journal of experimental psychology, 38*(4), 404.

Greenberg, A., Bellana, B., & Bialystok, E. (2013). Perspective-taking ability in bilingual children: Extending advantages in executive control to spatial reasoning. Cognitive development, 28(1), 41-50.

Gunia, A., Moraresku, S., & Vlček, K. (2021). Brain mechanisms of visuospatial perspective-taking in relation to object mental rotation and the theory of mind. Behavioural Brain Research, 407, 113247.

Hanna, J. E., Tanenhaus, M. K., & Trueswell, J. C. (2003). The effects of common ground and perspective on domains of referential interpretation. Journal of Memory and Language, 49(1), 43-61.

Hegarty, M., & Waller, D. (2004). A dissociation between mental rotation and perspective-taking spatial abilities. Intelligence, 32(2), 175-191.

Heller, D., & Brown-Schmidt, S. (in press). The multiple perspectives theory of mental states in communication. *Cognitive Science*.

Heller, D., Grodner, D., & Tanenhaus, M. K. (2008). The role of perspective in identifying domains of reference. Cognition, 108(3), 831-836.

Heller, D., Parisien, C., & Stevenson, S. (2016). Perspective-taking behavior as the probabilistic weighing of multiple domains. Cognition, 149, 104-120.

Horton, W. S., & Gerrig, R. J. (2005a). Conversational common ground and memory processes in language production. *Discourse Processes, 40(1)*, 1-35. doi:10.1207/s15326950dp4001_1

Horton, W. S., & Gerrig, R. J. (2016). Revisiting the memory-based processing approach to common ground. *Topics in Cognitive Science, 8*, 780-795. 10.1111/tops.12216

Horton, W. S., & Keysar, B. (1996). When do speakers take into account common ground?. Cognition, 59(1), 91-117.

Huttenlocher, J., & Presson, C. C. (1973). Mental rotation and the perspective problem. Cognitive Psychology, 4(2), 277-299.

Huttenlocher, J., & Presson, C. C. (1979). The coding and transformation of spatial information. Cognitive psychology, 11(3), 375-394.

Kessler, K., & Thomson, L. A. (2010). The embodied nature of spatial perspective taking: Embodied transformation versus sensorimotor interference. Cognition, 114(1), 72-88.

Keysar, B., Barr, D. J., Balin, J. A., & Brauner, J. S. (2000). Taking perspective in conversation: The role of mutual knowledge in comprehension. Psychological Science, 11(1), 32-38.

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

Kirchner, W. K. (1958). Age differences in short-term retention of rapidly changing information. *Journal of experimental psychology, 55*(4), 352.

Lin, S., Keysar, B., & Epley, N. (2010). Reflexively mindblind: Using theory of mind to interpret behavior requires effortful attention. Journal of Experimental Social Psychology, 46(3), 551-556.

Luce, K., & Almor, A. (2023) Inconsistency in Perspective-Taking during Comprehension. *Psychonomic Bulletin & Review*. https://doi.org/10.3758/s13423-023-02315-0

Michelon, P., & Zacks, J. M. (2006). Two kinds of visual perspective taking. Perception & psychophysics, 68(2), 327-337.

Mozuraitis, M., Stevenson, S. and Heller, D. (2018), Modeling Reference Production as the Probabilistic Combination of Multiple Perspectives. Cognitive Science, 42, 974-1008. doi:10.1111/cogs.12582

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, *27(2)*, 169-190. doi:10.1017/S0140525X04000056

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(2), 230-236.

R Core Team (2023). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rouder, J. N., & Haaf, J. M. (2019). A psychometrics of individual differences in experimental tasks. *Psychonomic bulletin & review*, *26*(2), 452-467.

Ryskin, R. A., & Brown-Schmidt, S. (2014). Do adults show a curse of knowledge in false-belief reasoning? A robust estimate of the true effect size. PloS one, 9(3), e92406.

Ryskin, R. A., Brown-Schmidt, S., Canseco-Gonzalez, E., Yiu, L. K., & Nguyen, E. T. (2014). Visuospatial perspective-taking in conversation and the role of bilingual experience. Journal of Memory and Language, 74, 46-76.

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.

Ryskin, R. A., Wang, R. F., & Brown-Schmidt, S. (2016). Listeners use speaker identity to access representations of spatial perspective during online language comprehension. Cognition, 147, 75-84.

Samson, D., Apperly, I. A., Braithwaite, J. J., Andrews, B. J., & Bodley Scott, S. E. (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.

Schober, M. F. (1995). Speakers, addressees, and frames of reference: Whose effort is minimized in conversations about locations?. Discourse Processes, 20(2), 219-247.

Stalnaker, R. C. (1978). Assertion. In P. Cole (Ed.), *Syntax and semantics, Vol. 9: Pragmatics* (pp. 315–332). New York: Academic..

Staub, A. (2021). How reliable are individual differences in eye movements in reading?. Journal of Memory and Language, 116, 104190.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. Journal of experimental psychology, 18(6), 643.

Surtees, A., Samson, D., & Apperly, I. (2016). Unintentional perspective-taking calculates whether something is seen, but not how it is seen. *Cognition*, 148, 97-105.

Tian, M., Luo, T., Ding, J., Wang, X., & Cheung, H. (2021). Spatial ability and theory of mind: A mediating role of visual perspective taking. Child Development, 92(4), 1590- 1604.

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? Journal of Memory & Language, 28, 127-154.

Wardlow, L. (2013). Individual differences in speakers' perspective taking: The roles of executive control and working memory. Psychonomic bulletin & review, 20(4), 766-772.

Wellman, H. M., & Liu, D. (2004). Scaling of theory-of-mind tasks. Child development, 75(2), 523-541.

Wu, S., & Keysar, B. (2007). The effect of culture on perspective taking. *Psychological Science*, 18(7), 600-606.

Zehr, J., & Schwarz, F. (2018) PennController for Internet Based Experiments (IBEX). https://doi.org/10.17605/OSF.IO/MD832.

APPENDIX A

PERMISSION TO REUSE

Copyright Transfer Statement

SPRINGER NATURE

Assignee:	The Psychonomic Society, Inc.	(the 'Assignee')
Journal Name:	Psychonomic Bulletin & Review	(the 'Journal')
Manuscript Number:	PBR-BR-22-114.R2	
Proposed Title of Article:	Inconsistency in perspective-taking during comprehension	(the 'Article')
Author(s) [Please list all named Authors]:	Kanan Luce, Amit Almor	(the 'Author')
Corresponding Author Name:	Kanan Luce	

1 Grant of Rights

a) For good and valuable consideration, the Author hereby assigns the copyright in the article to the Assignee. Without limiting the foregoing, the Author grants to the Assignee the perpetual, exclusive, world-wide, assignable, sublicensable and unlimited right to: publish, reproduce, copy, distribute, communicate, display publicly, sell, rent and/ or otherwise make available the article identified above, including any supplementary information and graphic elements therein (e.g. illustrations, charts, moving images) (the "Article") in any language, in any versions or editions in any and all forms and/or media of expression (including without limitation in connection with any and all end-user devices), whether now known or developed in the future. Without limitation, the above grant includes: (i) the right to edit, alter, adapt, adjust and prepare derivative works; (ii) all commercial use, advertising, and marketing rights, including without limitation graphic elements on the cover of the journal and in relation to social media; (iii) rights for any training, educational and/or instructional purposes; (iv) the right to add and/or remove links or combinations with other media/works; and (v) the right to create, use and/or license and/or sublicense content data or metadata of any kind in relation to the Article (including abstracts and summaries) without restriction. The above rights are granted in relation to the Article as a whole or any part and with or in relation to any other works.

b) Without limiting the rights granted above, Assignee is granted the rights to use the Article for the purposes of analysis, testing, and development of publishing- and research-related workflows, systems, products, projects, and services; to confidentially share the Article with select third parties to do the same; and to retain and store the Article and any associated correspondence/files/forms to maintain the historical record, and to facilitate research integrity investigations. The grant of rights set forth in this clause (b) is irrevocable.

c) The Assignee will have the right, but not the obligation, to exercise any or all of the rights granted herein. If the Assignee elects not to publish the Article for any reason, all publishing rights under this Agreement as set forth in clause 1.a) above will revert to the Author.

2 Use of Article Versions

a) For purposes of this Agreement: (i) references to the "Article" include all versions of the Article; (ii) "Submitted Manuscript" means the version of the Article as first submitted by the Author; (iii) "Accepted Manuscript" means the version of the Article accepted for publication, but prior to copy-editing and typesetting; and (iv) "Version of Record" means the version of the Article published by the Assignee, after copy-editing and typesetting. Rights to all versions of the Manuscript are granted on an exclusive basis, except for the Submitted Manuscript, to which rights are granted on a non-exclusive basis.

b) The Author may make the Submitted Manuscript available at any time and under any terms (including, but not limited to, under a CC BY licence), at the Author's discretion. Once the Article has been published, the Author will include an acknowledgement and provide a link to the Version of Record on the publisher's website: "This preprint has not undergone peer review (when applicable) or any post-submission improvements or corrections. The Version of Record of this article is published in [insert journal title], and is available online at https://doi.org/[insert DOI]".

c) The Assignee grants to the Author (i) the right to make the Accepted Manuscript available on their own personal, self-maintained website immediately on acceptance, (ii) the right to make the Accepted Manuscript available for public release on any of the following twelve (12) months after first publication (the "Embargo Period"): their employer's internal website; their institutional and/or funder repositories. Accepted Manuscripts may be deposited in such repositories immediately upon acceptance, provided they are not made publicly available until after the Embargo Period.

The rights granted to the Author with respect to the Accepted Manuscript are subject to the conditions that (i) the Accepted Manuscript is not enhanced or substantially reformatted by the

Page 1 of 3

Author or any third party, and (ii) the Author includes on the Accepted Manuscript an acknowledgement in the following form, together with a link to the published version on the publisher's website: "This version of the article has been accepted for publication, after peer review (when applicable) but is not the Version of Record and does not reflect post-acceptance improvements, or any corrections. The Version of Record is available online at: http://dx.doi.org/[insert DOI]. Use of this Accepted Version is subject to the publisher's Accepted Manuscript terms of use https://www.springernature.com/gp/open-research/policies/accepted-manuscript-terms". Under no circumstances may an Accepted Manuscript be shared or distributed under a Creative Commons or other form of open access licence.

d) The Assignee grants to the Author the following non-exclusive rights to the Version of Record, provided that, when reproducing the Version of Record or extracts from it, the Author acknowledges and references first publication in the Journal according to current citation standards. As a minimum, the acknowledgement must state: "First published in [Journal name, volume, page number, year] by Springer Nature".

   i. to reuse graphic elements created by the Author and contained in the Article, in presentations and other works created by them;

   ii. the Author and any academic institution where they work at the time may reproduce the Article for the purpose of course teaching (but not for inclusion in course pack material for onward sale by libraries and institutions);

   iii. to reuse the Version of Record or any part in a thesis written by the same Author, and to make a copy of that thesis available in a repository of the Author(s)' awarding academic institution, or other repository required by the awarding academic institution. An acknowledgement should be included in the citation: "Reproduced with permission from Springer Nature"; and

   iv. to reproduce, or to allow a third party to reproduce the Article, in whole or in part, in any other type of work (other than thesis) written by the Author for distribution by a publisher after an embargo period of 12 months.

   For clarity, where any graphic element is used on the cover of the Journal all intellectual property rights in any compilation constituting the cover will be vested solely in the Licensee, and the Author is not entitled to make or authorise any other person to make use of any such compilation without the prior written consent of the Licensee.

Warranties & Representations

Author warrants and represents that:

a)

   i. the Author is the sole copyright owner or has been authorised by any additional copyright owner(s) to the rights defined in clause 1,

   ii. the Article does not infringe any intellectual property rights (including without limitation copyright, database rights or trade mark rights) or other third party rights and no licence from or payments to a third party are required to publish the Article,

   iii. the Article has not been previously published or licensed, nor has the Author committed to licensing any version of the Article under a licence inconsistent with the terms of this Agreement,

   iv. if the Article contains materials from other sources (e.g. illustrations, tables, text quotations), Author has obtained written permissions to the extent necessary from the copyright holder(s), to license to the Assignee the same rights as set out in clause 1 but on a non-exclusive basis and without the right to use any graphic elements on a stand-alone basis and has cited any such materials correctly;

b) all of the facts contained in the Article are according to the current body of research true and accurate;

c) nothing in the Article is obscene, defamatory, violates any right of privacy or publicity, infringes any other human, personal or other rights of any person or entity or is otherwise unlawful and that informed consent to publish has been obtained for any research participants;

d) nothing in the Article infringes any duty of confidentiality owed to any third party or violates any contract, express or implied, of the Author;

e) all institutional, governmental, and/or other approvals which may be required in connection with the research reflected in the Article have been obtained and continue in effect;

f) all statements and declarations made by the Author in connection with the Article are true and correct; and

g) the signatory who has signed this agreement has full right, power and authority to enter into this agreement on behalf of all of the Authors.

4 Cooperation

The Author will cooperate fully with the Assignee in relation to any legal action that might arise from the publication of the Article, and the Author will give the Assignee access at reasonable times to any relevant accounts, documents and records within the power or control of the Author. The Author agrees that any Assignee affiliate through which the Assignee exercises any rights or performs any obligations under this Agreement is intended to have the benefit of and will have the right to enforce the terms of this Agreement.

5 Author List

Changes of authorship, including, but not limited to, changes in the corresponding author or the sequence of authors, are not permitted after acceptance of a manuscript.

6 Post Publication Actions

The Author agrees that the Assignee may remove or retract the Article or publish a correction or other notice in relation to the Article if the Assignee determines that such actions are appropriate from an editorial, research integrity, or legal perspective.

7 Controlling Terms

The terms of this Agreement will supersede any other terms that the Author or any third party may assert apply to any version of the Article.

8 Governing Law

This Agreement will be governed by, and construed in accordance with, the laws of the State of New York. The courts of New York, N.Y. will have exclusive jurisdiction.

Signed for and on behalf of the Author(s)
Corresponding Author: Kanan Luce
Email: kbluce@email.sc.edu
IP Address: [redacted]

Time Stamp: 2023-06-11 11:17:46