Journal of Experimental Psychology: General
2015, Vol. 144, No. 5, 898–915

0096-3445/15/$12.00 http://dx.doi.org/10.1037/xge0000093

Perspective-Taking in Comprehension, Production, and Memory: An Individual Differences Approach

Rachel A. Ryskin and Aaron S. Benjamin
University of Illinois at Urbana-Champaign

Jonathan Tullis
University of Arizona

Sarah Brown-Schmidt
University of Illinois at Urbana-Champaign

The ability to take a different perspective is central to a tremendous variety of higher level cognitive skills. To communicate effectively, we must adopt the perspective of another person both while speaking and listening. To ensure the successful retrieval of critical information in the future, we must adopt the perspective of our own future self and construct cues that will survive the passage of time. Here we explore the cognitive underpinnings of perspective-taking across a set of tasks that involve communication and memory, with an eye toward evaluating the proposal that perspective-taking is domain-general (e.g., Wardlow, 2013). We measured participants' perspective-taking ability in a language production task, a language comprehension task, and a memory task in which people generated their own cues for the future. Surprisingly, there was little variance common to the 3 tasks, a result that suggests that perspective-taking is not domain-general. Performance in the language production task was predicted by a measure of working memory, whereas performance in the cue-generation memory task was predicted by a combination of working memory and long-term memory measures. These results indicate that perspective-taking relies on differing cognitive capacities in different situations.

Keywords: perspective-taking, comprehension, production, memory cue generation

Many activities of everyday life require individuals to imagine a world that differs from the one they currently inhabit—whether it is finding common ground with someone who has had different life experiences or predicting what you will remember in the future. A common setting that requires taking a different perspective into account is face-to-face conversation. When two individuals converse, some of what they might want to discuss is known to both parties and some is known only to one of them. Partners in a dialogue must model each others' knowledge states to communicate efficiently (Clark, 1992, 1996; Stalnaker, 1978). For example, imagine that a police officer is investigating a bank robbery and interrogating the prime suspect. The amount of money stolen and the bank that was robbed are known to both of them—they are jointly known and therefore "common ground" (Clark & Marshall, 1978). Only the robber knows when and how he gained access to the vault without triggering any alarms. This information is said to be the suspect's privileged ground. To avoid incriminating himself, the suspect needs to make sure not to reveal that he has any additional knowledge. Similarly, only the police officer knows that they have collected fingerprints that match the suspect's at the scene of the crime; that piece of evidence is in the officer's privileged ground. She may choose to reveal this privileged information at some strategic point in the interrogation. For the officer and the suspect to achieve their respective conversational goals, they must both consider, and keep separate, what is common information and what is privileged information. This example illustrates, in an extreme way, how information states differ in a conversational dyad and how interlocutors may benefit from modeling each other's knowledge states.

Similarly, when writing a reminder note for oneself to do something at a later time, such as to call a friend to ask for their new home address after they have moved, some notes will be more effective than others. For example, when you read a note that says "Call Rob" you might wonder what it is you were supposed to call him about. On the other hand, "Call Rob about address" may provide the appropriate cue to bring to mind all the relevant details of what action needs to be taken. When writing a reminder note for yourself, it is necessary to accurately assess aspects of your future state at the time of retrieval (Tullis & Benjamin, 2015a, 2015b).

Rachel A. Ryskin, Department of Psychology, University of Illinois at Urbana-Champaign; Aaron S. Benjamin, Department of Psychology, and Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign; Jonathan Tullis, Department of Educational Psychology, University of Arizona; Sarah Brown-Schmidt, Department of Psychology, and Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign.

This work was supported by National Science Foundation Grant BCS 12-57029 to S.B.S. We thank the members of the Conversation and Human Memory and Cognition Labs at University of Illinois at Urbana-Champaign for many helpful discussions regarding this work. We would also like to thank Cassandra Barnas, Cecille Kalafut, R. Mia Kaim, Randy Knebel, Daniel Pak, and Tatsu Shigeta for help with data collection and coding.

Correspondence concerning this article should be addressed to Rachel A. Ryskin, Department of Psychology, University of Illinois at Urbana-Champaign, 603 E. Daniel St., Champaign, IL 61801. E-mail: rryskin@gmail.com

This assessment is analogous to perspective-taking of your future self—you must determine what will be common between your current self and your future self and what is privileged information in your current state.

In the present article, we address the cognitive capacities that underlie perspective-taking, and also whether there is a core cognitive process that serves them all. We briefly review what is known about the role of perspective-taking in three domains: language comprehension, language production, and memory cue generation. We then examine the prior evidence that is thought to support a domain-general view of perspective-taking, including the cognitive individual differences that have been implicated. We conclude that, despite surface similarity in the different types of perspective-taking activities, different cognitive processes govern perspective-taking during comprehension, production, and cue generation. Finally, we provide empirical evidence that supports our hypothesis and discuss the implications of our findings.

Perspective-Taking in Language Comprehension

One set of relevant evidence comes from the visual world eye-tracking paradigm (Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). This research suggests that listeners represent information about the perspective of the speaker, and use this information to guide processing of the speaker's utterances (Brown-Schmidt, Gunlogson, & Tanenhaus, 2008; Ferguson & Breheny, 2012; Hanna, Tanenhaus, & Trueswell, 2003; Heller, Grodner, & Tanenhaus, 2008). For example, Heller et al. (2008) asked a speaker (the experimenter) and listener (the participant) to sit on either side of a physical display with cubbyholes containing objects. Most of these objects were mutually visible from either side of the display and therefore common ground. In critical conditions, one object was placed in a cubbyhole that was occluded from the speaker's view—that object was in the listener's privileged ground. Eye-movement data indicated that listeners took into account the privileged status of that object when interpreting the speaker's instructions and did not consider it to be a potential referent. In other words, listeners inferred that the speaker could not be referring to the occluded object even when the speaker's utterance was potentially consistent with it. To do this, listeners must form a representation of the speaker's perspective on the visual display and access this perspective information when interpreting the speaker's instructions.

Perspective-Taking in Language Production

When choosing what to say to another person, speakers take into account what the addressee does and does not know. At the most basic level, Spanish-English bilingual children know from an early age to speak in English to English speakers and Spanish to Spanish speakers (Genesee, Nicoladis, & Paradis, 1995). At a more fine-grained level, speakers modulate their use of adjectives based on the listener's knowledge (Matthews, Lieven, Theakston, & Tomasello, 2006; Nadig & Sedivy, 2002; Yoon & Brown-Schmidt, 2014; Yoon, Koh, & Brown-Schmidt, 2012). For example, Nadig and Sedivy (2002) examined a situation in which participants had to instruct a listener (the experimenter) about which object in a cubbyhole display to pick up (e.g., "Pick up the cup"). The target object was sometimes in a size-contrasted pair (i.e., there was a big cup and small cup in the display). When both members of the pair were common ground, adult participants, and even 5- and 6-year-olds, always used an adjective (e.g., Pick up the big cup) to make clear which cup they were referring to. When one of the pair members was occluded from the listener's view (e.g., the small cup was in the speaker's privileged ground), such that there was only one cup visible from the listener's perspective, participants used the size adjective only half of the time. These different choices reflect the fact that the speaker has a mental representation of what is available to the listener and uses that information to guide speech.

Perspective-Taking in Memory Cue Generation

Individuals often prospectively generate cues for themselves. For example, students take notes in a class to be able to recall the class material at a later time. Recall is more likely to occur when the retrieval context and the encoding context are more similar (Tulving & Thomson, 1973). Thus, creating a successful cue requires predicting the future cognitive context at the time of retrieval—in other words, taking the perspective of one's future self. Evidence that cue generators do take the perspective of their future selves comes from Tullis and Benjamin (2015b), who asked participants to generate either one-word descriptions of target words or one-word cues that would support later cued recall of targets. Generated cues elicited higher recall than generated descriptions, suggesting that learners tailor their mnemonic cues to their future cognitive state. Learners were even able to hedge against future interference among to-be-remembered synonyms by trading off cue-target associative strength for cue distinctiveness when necessary. Further, in much the same way as in speech production, learners tailor mnemonic cues for different intended recipients; cues are more idiosyncratic when generated for themselves than for others (Tullis & Benjamin, 2015a).

Individual Differences in Perspective-Taking

Despite how widespread perspective-taking seems to be in everyday life, it is not something that individuals do effortlessly (Birch & Bloom, 2007; Keysar, Lin, & Barr, 2003; cf. Ryskin & Brown-Schmidt, 2014). Children under the age of four often fail to appreciate that other people may hold different beliefs (Wimmer & Perner, 1983) when probed in an explicit verbal task. On the other hand, implicit measures show that 15-month-olds are more surprised when an agent looks for a toy in the correct location if the agent has never seen the toy be placed in that specific location (Onishi & Baillargeon, 2005). The latter result suggests that children may form representations of others' perspectives very early, but lack the cognitive resources to select verbal answers that reflect this knowledge until much later in development. These findings are consistent with evidence that the development of theory of mind in children goes hand in hand with the development of inhibitory control (Carlson & Moses, 2001). More specifically, children with better executive function are more likely to take into account the perspective of a conversation partner when speaking or listening (Nilsen & Graham, 2009).

Similar relationships between executive function and perspective-taking during language comprehension and production have been reported in adults as well. Listeners with superior executive control and working memory are more adept at using the speaker's perspective to anticipate the correct referent of a sentence (Brown-Schmidt, 2009; Lin, Keysar, & Epley, 2010). Speakers are more likely to tailor their utterances appropriately given a listener's viewpoint when they have higher working memory (Wardlow, 2013). It is worth noting however, that these effects appear to be very small and do not consistently replicate (Brown-Schmidt, 2015; Brown-Schmidt & Fraundorf, 2015; Ryskin, Brown-Schmidt, Canseco-Gonzalez, Yiu, & Nguyen, 2014).

Executive function also plays a role in a variety of memory tasks that require prospection. Individuals with high working memory capacity and superior executive control are more likely to remember to carry out an intended action at a specified point in the future (Brewer, Knight, Marsh, & Unsworth, 2010; Kliegel, Martin, McDaniel, & Einstein, 2002; Marsh & Hicks, 1998; Unsworth, Brewer, & Spillers, 2012). They are also more likely to adopt retrieval cues that allow for an organized, efficient approach to recalling a list of heterogeneous items (Fiechter, Benjamin, & Unsworth, in press; Unsworth, Brewer, & Spillers, 2013; Unsworth & Engle, 2007; Unsworth & Spillers, 2010). Further, executive and working memory processes have been linked to the specificity and richness of representations during future simulation (D'Argembeau, Ortoleva, Jumentier, & Van der Linden, 2010; Hill & Emery, 2013).

In summary, cue generation and audience design in language production both require generating a cue for someone specific (you in the future or another person) that enables the cue recipient to access the target information without undue interference. Memory retrieval spurred by a previously generated cue and perspective-taking in comprehension both require taking the generator's (yourself in the past or another person) point of view into account to interpret the received cue. All three tasks also rely on working memory and executive control, which may prove to be mediating variables underlying any domain generality across these tasks.

A Domain-General Hypothesis

It has been proposed that perspective-taking is a domain-general construct (e.g., Wardlow, 2013) that is called upon by both the comprehension and production systems. On this account, the effects of executive functioning on each domain are primarily mediated by their influence on this underlying perspective-taking construct. Further support for this domain-general hypothesis comes from neuroimaging evidence that the ability to self-project across time and space constitutes a core process (Buckner & Carroll, 2007; Buckner, Andrews-Hanna, & Schacter, 2008) that underlies a variety of high-level cognitive functions, such as episodic memory retrieval (self-projection into the past; Wagner, Shannon, Kahn, & Buckner, 2005), theory of mind and spatial perspective-taking (self-projection into someone else's mind in the present; Saxe & Kanwisher, 2003), and prospection (self-projection into the future; Schacter, Addis, & Buckner, 2007).

An alternative hypothesis is that perspective-taking is a domain-specific ability. The observed overlap in brain activity during prospection, episodic retrieval, and theory of mind may result from shared cognitive processes that do not include perspective-taking. Similarly, the correlation of executive function with both prospection and audience design in production, for instance, does not necessitate that prospection and audience design are correlated. Apart from involving the representation of an alternate perspective, the tasks of comprehending perspective-laden language, producing optimally informative instructions, and generating successful retrieval cues share many other features. For instance, they may all require the activation of, and selection among, multiple alternative representations. It may be this, or some other, more basic process that relies on executive function across domains. Perspective-taking per se may not be what is common to all three domains. Indeed, there are key differences between how perspectives are computed between language production, language comprehension, and cue generation. For example, in a production setting, the producer must keep track of what she herself knows and what the recipient knows to create an informative utterance. In a comprehension setting, the listener must keep track of what she knows, what the speaker knows, and how likely the speaker is to accurately take the listener's perspective. In the context of memory cue generation, learners must keep track of what information they will know in the future and what they are likely to have forgotten by that time.

Present Research

The goal of the present research is to evaluate the contribution of individual-difference variables to performance across a set of perspective-taking tasks, and to test the hypothesis that perspective-taking is domain-general. To do so, we measured perspective-taking in language comprehension, language production, and cue generation. We also collected measures of executive function, recall, and cognitive failures in everyday life. If perspective-taking is indeed domain-general, we would expect the three measures of perspective-taking to covary. If perspective-taking operates instead in a domain-specific fashion, we would expect no shared variance between the three tasks.

Method

Participants

One hundred fifty-two undergraduate students at the University of Illinois at Urbana-Champaign participated in this experiment in exchange for partial course credit. This large sample size was necessary for an amply powered examination of individual differences. Participants had normal or corrected-to-normal vision and were self-reported native speakers of English.

Materials and Design

Participants were recruited in pairs and both came in for two sessions that were 48 hr apart. Members of each pair did not know each other beforehand. During the first session, which lasted approximately 1 hr, participants generated mnemonic cues for the cue generation task and completed four executive function tasks. During the second session, participants were given the cues that they had generated on the first day and attempted to recall the corresponding targets for each item. Next, they completed a cued recall task and a questionnaire. Finally, they took part in the conversation task as a pair. The second session lasted about 2 hr.

Cue generation. The cue generation task was made up of two parts—the study session and the test session—that were separated by a 48-hr delay. This delay ensures that participants' performance is not at ceiling and has been used in previous work with this task (Tullis & Benjamin, 2015a, 2015b). During the study session, participants saw 80 words one at a time, in random order. They were told to generate a one-word cue for each target word. Participants were informed that they would be tested on all the target words later and that they should pick cues that would be most helpful to them for retrieving the target. They were not allowed to use the target itself as a cue, but the selection was otherwise unconstrained.

During the test session, participants were presented with the 80 cue words they had generated in a new random order. They were asked to recall the target that corresponded to each cue. In both sessions, participants typed in responses and proceeded through the task at their own pace.

Target words were collected from the University of South Florida Free Association Norms (Nelson, McEvoy, & Schreiber, 1998). As in Tullis and Benjamin (2015a, 2015b), these to-be-remembered words were selected with the intention that college-aged subjects would have some personal experiences with the items. Sixty of the target words were identical to those used by Tullis and Benjamin (2015b). Examples of words include "dancing," "haircut," and "roommate." Learners generated cues such as "ballet," "mohawk," and "maggie." The Thorndike-Lorge written frequencies of the targets ranged from 27 to 2,218, with a mean of 536 and a *SD* of 510.

Conversation task. The conversation task was used to measure perspective-taking in both language production and comprehension. During the task, participants were seated in separate rooms but communicated with each other via wireless microphones. Partners saw identical 3 × 3 grids with eight objects and a fixation cross in the center (Figure 1a). They were told to imagine that their partner was sitting on the other side of a physical grid with cubbyholes (similar to Heller et al., 2008). The displays were mirror-reversed such that, from each participant's perspective, the items were positioned as if they were indeed looking at opposite sides of a grid of cubbies. Participants were informed that a gray background in a cell indicated that their partner could not see that object; these items were in the participant's privileged ground. Conversely, a black cell indicated that their partner could see something in that cell (i.e., it had a gray background on their partner's display and was in their partner's privileged ground). The target item was circled on the speaker's display. Participants randomly alternated telling each other which item on the screen to click on. Each room was equipped with an Eyelink-1000 desktop-mounted eye-tracker so that eye movements, as well as synchronized voice recordings, were collected for both participants throughout the conversation task.

Collecting both speech and eye-tracking data from both participants allowed us to evaluate the perspective-taking ability of both participants in both language production and comprehension. This feature of the design affords—for the first time in the substantial literature on conversation—a within-subjects comparison of perspective-taking in language production and comprehension. The conversation task consisted of 288 trials and lasted about 1 hr. One-third (96) of the trials constituted critical trials for one participant in the pair, another third of these trials constituted critical trials for the other participant in the pair, and one-third were filler trials. All trial types were randomly intermingled. Stimulus presentation for all tasks, except for the questionnaire, which was on paper, was controlled using the Psychophysics Toolbox Version 3 (PTB-3; Brainard, 1997) for MATLAB.

Language comprehension. Perspective-taking during comprehension was measured by monitoring listeners' eye movements while their partner provided an instruction. On critical comprehension trials, the speaker's display was always unambiguous with respect to the instruction they were intended to produce (i.e., critical production trials were not used as stimuli for comprehension). The rate of production errors on these trials was low (6%) and trials containing speaker errors were not included in the analyses of comprehension data.[1] In other words, we can be confident that the instruction given, though unscripted, provided an appropriate auditory stimulus.

The three critical conditions[2] (Figure 1a–c) were modeled after Heller et al. (2008), and were designed to assess the listener's sensitivity to the speaker's perspective. Across conditions, listeners always heard an instruction that included a size modifier and a target object to click on (e.g., "Click on the big banana"). In the *Two Contrasts-Shared* condition (Figure 1a), listeners saw two sets of items that formed a size contrast (e.g., a big banana, a small banana, a big balloon, and a small balloon)[3] and three other filler items on the screen. One of the filler items was in the listener's privileged ground (indicated by the gray background) and the contents of one cell were hidden from the listener (indicated by the black square). The competitor—for all conditions—was defined as the item that was also part of a contrast set and the same size as the target (e.g., the big balloon when the big banana was the target). In the *Two Contrasts-Privileged* condition (Figure 1b), listeners saw two sets of items that formed a size contrast (e.g., a big balloon, a small balloon, a big banana, and a small banana) and three other filler items on the screen. Critically, the item that contrasted in size with the competitor (i.e., the small balloon when the competitor is the big balloon) was in privileged ground (indicated by the gray background). The contents of one cell were hidden from the listener (indicated by the black square). In the *One Contrast condition* (Figure 1c), listeners saw one set of items that formed a size contrast (e.g., a big banana and a small banana), one item that was cohort and size competitor for the target (e.g., a big balloon) and four filler items. One of the filler items was in privileged ground (indicated by the gray background) and the

---

[1] All the analyses were also done with speaker error trials included and the results did not differ. The reliability of the individual measure of perspective-taking in comprehension was slightly improved when the error trials were excluded so all analyses reported here do not include the 6% of trials with errors.

[2] There were two other "comprehension" conditions during which the participant acted as a listener but these trials were not of interest in terms of the analysis of perspective-driven eye movements. On these trials, the listeners often heard an instruction with no size modifier (e.g., "Click on the banana") and there was no temporary referential ambiguity because the listener only saw one instance of the target item (e.g., there was only one banana on their screen). The purpose of these trials was simply to provide a recipient for the partner's critical production trials.

[3] Note that the two items (balloon and banana) are phonological cohort competitors. Trials were constructed in this way so that the point of disambiguation (e.g., between "big balloon" and "big banana") occurred slightly later than the onset of the noun, giving us more time to observe any effects due to perspective-taking.

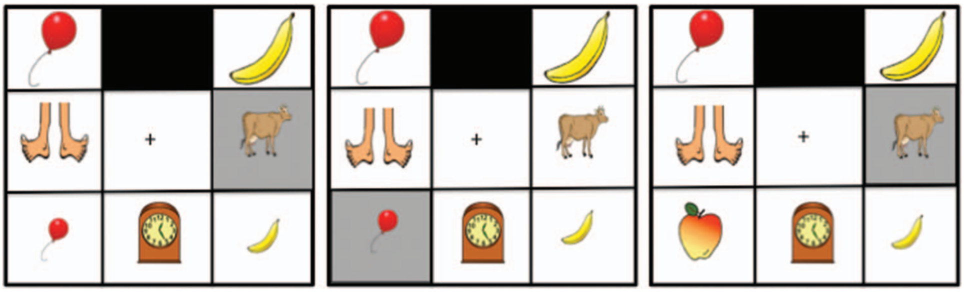

Figure 1. Sample displays from the three critical conditions during the *comprehension* trials of the conversation task. See the online article for the color version of this figure.

contents of one cell were hidden from the listener (indicated by the black square).

Language production. Perspective-taking during language production was measured using the speakers' correct use of size adjectives given the visual context. The three critical[4] conditions (see Figure 2) were modeled after Nadig and Sedivy (2002). A green circle appeared around one of the objects on the screen and speakers instructed their partner to click on that object. In the *Contrast-Shared* condition, the display contained two[5] sets of items that formed a size contrast (e.g., a big banana, a small banana, a big balloon, and a small balloon) and three other filler items on the screen. One of the filler items was in privileged ground (indicated by the gray background) and the contents of one cell were hidden from the speaker (indicated by the black square). The target item—circled in green—was part of a size contrast (e.g., the big banana). The correct instruction in this case was "Click on the big banana" because the listener sees two different-sized bananas. This production condition is equivalent to the Two Contrasts-Shared comprehension condition. In the *Contrast-Privileged* condition, the display contained two sets of items that formed a size contrast (e.g., a big banana, a small banana, a big balloon, and a small balloon) and three other filler items on the screen. Critically, the item that contrasted in size with the target (i.e., the small banana) was in privileged ground (indicated by the gray background). The contents of one cell were hidden from the participant (indicated by the black square). The correct instruction in this case was "Click on the banana" because the listener only sees one banana, despite the fact that the speaker sees two. In the *No Contrast* condition, speakers saw one target item (e.g., a big banana), one set of items that formed a size contrast (e.g., a big balloon and a small balloon), and four filler items. One of the filler items was in privileged ground (indicated by the gray background) and the contents of one cell were hidden from the speaker (indicated by the black square). The correct instruction in this case was "Click on the banana" because the listener and speaker only see one banana.

Each critical trial consisted of a random pairing of condition (comprehension: Two Contrasts-Shared, Two Contrasts-Privileged, One Contrast, or production: Contrast-Shared, Contrast-Privileged, No Contrast) and cohort set (e.g., banana-balloon, ant-anvil). The Two Contrasts-Shared/Contrast-Shared condition was critical for both comprehension and production trials. So as not to diminish the statistical power of our analyses, we made it into two conditions (one critical for production, the other for comprehension). Therefore, there are 32 trials of Two Contrasts-Shared, 16 of Two Contrasts-Privileged, and 16 of One Contrast for each participant's eye-movement data. There are also 32 trials of Contrast-Shared (the same trials as the 32 Two Contrasts-Shared trials), 16 of Contrast-Privileged, and 16 of No Contrast for each participant's production data.

Individual differences. During the first session, participants completed four tasks measuring individual differences in executive function. These tasks were chosen to be similar to those used in prior work linking perspective-taking and executive function (Brown-Schmidt, 2009; Nilsen & Graham, 2009; Wardlow, 2013). In addition, two measures of individual differences in memory were collected during the second session. The six measures are listed below:

1. Stroop Task One: Participants viewed color words presented in a colored text that did not match the meaning of the word. For instance, the word "green" might appear in blue text. In the first block of trials, participants were instructed to simply read the word and say it out loud (e.g., "green"). In the second block, they were instructed to ignore the word and say the color of the text out loud (e.g., "blue"). Across the two blocks, the word and the color of the text were always incongruent. Each color word appeared on the screen for 700 ms and participants' spoken responses were recorded throughout. Recordings were later hand-coded for accuracy; the proportion of correct responses across the two blocks was used as a measure of inhibitory control.

2. Stroop Task Two: Participants viewed color words presented in a colored text in front of a patch of color. For instance, the word "green" might appear in blue text in

---

[4] There were two other production conditions during which the participant gave instructions but these trials were not of interest for the analysis of perspective-driven utterances. On these trials, speakers saw one contrast set (e.g., big balloon, small balloon) and the target was one item in that pair such that (similar to Two Contrasts–Shared) they should produce an instruction with a size modifier (e.g., "Click on the big balloon"). The purpose of these trials was simply to provide the auditory stimulus for the partner's critical comprehension trials (Two Contrasts–Privileged and One Contrast).

[5] Note that only one of the contrast sets—the one including the target—was relevant to the utterance production.

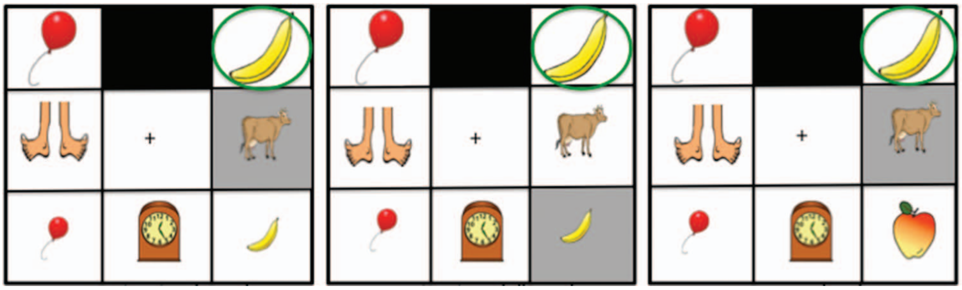

Figure 2. Sample displays from the three critical conditions during the *production* trials of the conversation task. See the online article for the color version of this figure.

front of a red rectangle. In the first block of trials, participants were instructed to ignore the word and say the color of the patch out loud (e.g., "red"). In the second block of trials, participants were instructed to ignore the word and say the color of the text out loud (e.g., "blue"). All trials were self-paced and participants' reaction times (RTs; latency of button press to advance to the next trial) and spoken responses were recorded throughout. The first block was 15% congruent (i.e., the patch color matched the word) and the second block was 51% congruent (i.e., the color of the word matched the word). Each participant received an interference score, obtained by subtracting the average RT for saying the patch color from the average RT for saying the text color. Thus, a lower score signified less interference of the incongruent color word or superior inhibitory control. These scores were recoded so that a higher score on Stroop Two (as well as Stroop One) indicated superior inhibitory control.

3. Minus-Two Span Task: In this working memory span task, participants saw a set of numbers, one by one. After each set they were asked to recall (and type in), in order, the numbers they had just viewed, subtracting two from each number. The set size varied from two to seven. Each set size was used twice for a total of 12 blocks. The score was determined by adding together the proportions correct for each block. For instance, if the participant recalled four items from the block of six correctly, that block would receive a score of 4/6. The maximum possible score was 12.

4. Operation Span Task: This working memory task was modeled after one described by Unsworth, Heitz, Schrock, and Engle (2005). Participants saw a set of math problems (e.g., $(3 \times 2) + 4 = ?$) followed by a number (e.g., 9). They were told to respond "True" (by pressing a keyboard key) if that number was the answer to the preceding math problem and "False" if it was not (in this case, "False"). After each response they saw a letter (e.g., R). At the end of a set, they were asked to recall and type in all the letters they had seen in that set, in order. Participants were told to maintain their accuracy on the math problems above 85% and were given feedback about their accuracy throughout the task. Pacing was calibrated to each subject by averaging their speed of response to practice math problems and using a speed that was 2.5 *SD*s above that average. The set size varied from three to seven. Each set size was used three times for a total of 15 blocks. The score was determined by adding together the proportions correct for each block. For instance, if the participant recalled four letters from the block of six correctly, that block would receive a score of 4/6. The maximum possible score was 15.

5. Cued Recall Task: In this task, participants studied 40 pairs of words (e.g., bitter - lemon) on the screen for 3 s each. They were told to try to remember them for a later test. After they studied the 40 pairs, they were given the left word from each pair (e.g., bitter - ?) and were asked to type in the right-side member of the pair. Number of words recalled correctly was used as a measure of basic memory recall ability.

6. Cognitive Failures Questionnaire (CFQ): The questionnaire is intended to measure the rate of a person's memory deficits, absent-mindedness, and slips of action in daily life (Broadbent, Cooper, Fitzgerald, & Parkes, 1982), because such events presumably relate to one's ability to engage in successful perspective-taking in more ecologically relevant circumstances than those measured by the word-list procedure. It contains questions such as: "Do you find you confuse left and right when giving directions?", "Do you find you forget appointments?", "Do you daydream when you ought to be listening to something?" Answers are given on a 0 (*never*) to 4 (*very often*) scale, and each participant's score consists of the sum of their responses on all 25 questions.

Predictions

Perspective-taking effects. In the cue generation task, if participants can take the perspective of their future selves, they will create effective cues for themselves and retrieve a substantial number of targets despite the 2-day delay between study and test. We also expect that cues that lead to successful retrieval will be characterized by their idiosyncracy and distinctiveness (Tullis & Benjamin, 2015b).

If participants can take into account their partner's perspective during online language comprehension, they will be more likely to fixate the target in the Two Contrasts-Privileged condition than the

Two Contrasts-Shared condition, because the speaker's use of an adjective will only be consistent with one possible referent in the display when a member of a size contrast pair is in privileged ground. We also expect that participants will be most likely to fixate the target in the One Contrast condition because only one item will be consistent with the speaker's use of an adjective—no other size contrasts appear in the display.

If participants can tailor their utterances to the perspective of their partner, they will use an adjective to describe the target referent more frequently in the Contrast-Shared condition than the Contrast-Privileged condition, because in the Contrast-Privileged condition the referent is fully disambiguated without the use of an adjective. We also expect participants to use an adjective very infrequently in the No Contrast condition because the target is not in a size contrast with anything else in the display.

Relationships between tasks. Given prior findings of a link between executive function and perspective-taking, we predict that the four measures of executive function should predict perspective-taking ability in comprehension, production, and cue generation. If the domain-general hypothesis is correct, we would expect individual measures of perspective-taking ability to correlate strongly across the three tasks. On the other hand, if there is no relationship between the individual measures of perspective-taking derived from each task, our results would lend support to a domain-specific account of perspective-taking.

Results

Evidence for Perspective-Taking in Three Domains

Cue generation. Performance on the cue generation task was based upon strict scoring, such that only the responses that matched the target identically were counted as correct. For example, if the target was "dancer," "dance" would be counted as incorrect. On average, participants retrieved 42.13 ($SD = 13.62$) of the 80 target words they had generated cues for.

Two major characteristics of learner-generated cues were measured: (a) the association between cue and target and (b) cue overload. Cue-to-target associative strength was determined using the normative cue-to-target associative strength found in the University of South Florida Free Association Norms (Nelson et al., 1998). Cued recall performance typically increases as cue-to-target associative strength in a word pair increases (Feldman & Underwood, 1957). Second, cue overload, or the number of possible targets associated with each cue, was measured by counting the total number of targets in the Free Association Norms associated with each cue, by summing the cue-to-target associative strengths from the cue to all possible targets in the database, and by counting whether the cue was in the database at all. Cues that are associated to fewer possible targets limit the search space during recall and are typically associated with better memory performance (Anderson, 1974). The characteristics of successful versus unsuccessful cues are summarized in Appendix A. As expected and as shown in prior literature (Tullis & Benjamin, 2015a), cues that lead to later successful retrieval had higher cue-to-target associative strength, a lower number of associates from cue, lower total associative strength from cue, and were less likely to be in the database. This suggests that when learners tailored their cues to be more normatively associated, but also idiosyncratic and distinctive, they were more successful at retrieving the target.

To assess how consistent participants are in the cue generation task, we measured coefficient α (the mean of all possible split-half reliabilities) using the *psych* package in R (Revelle, 2014). The results of this analysis indicate that the cue generation task is a highly reliable measure of participants' ability to generate effective cues for later retrieval ($\alpha = .9$, 95% CI = [0.88, 0.93]).

Language comprehension. To assess perspective-taking during comprehension, we measured the eye movements that listeners made as they interpreted the speaker's instruction (e.g., *Click on the big banana*). The time-course of target and competitor fixations across conditions can be seen in Figure 3. Eye movements associated with the interpretation of the instruction were analyzed in terms of a binary measure: whether the participant fixated the target during the specified time window or not.[6] A fixation was coded as a fixation to the target referent if the x, y fixation coordinates landed within the cell containing the target object (e.g., *the big banana*).

Target fixations were measured in one time window (average duration 1,100 ms) that began at the onset of the scalar adjective (e.g., *big*) and ended 600 ms after the onset of the noun (e.g., *banana*); the average duration of the noun was 600 ms. This large analysis window allows us to examine the full interpretive process occurring as the participant learns which item is the target. The time-window was offset by 200 ms because of the time needed to program and launch an eye movement (Hallett, 1986).

Target fixations were analyzed in a multilevel logistic regression, using the *lme4* software package in R (Bates, Maechler, Bolker, & Walker, 2015). The three conditions (Two Contrasts-Shared, Two Contrasts-Privileged, and One Contrast) were entered in the fixed effects as weighted orthogonal contrast codes. The first contrast compares Two Contrasts-Shared and Two Contrasts-Privileged with the One Contrast condition, and was used to test for the presence of a competition effect (fewer target fixations) in the presence of a second size-contrast set. The second contrast compares Two Contrasts-Shared to Two Contrasts-Privileged. The second contrast was our critical comparison, and tested the ability of listeners to use perspective information to eliminate the second size-contrast set from consideration. Subjects and cohort pairs (e.g., banana-balloon) were entered as crossed random effects with by-subject and by-cohort pair random slopes for condition.

Within the selected time-window, listeners were less likely to fixate the target (e.g., the big banana) when they were in one of the Two Contrast conditions than in the One Contrast condition (see Table 1). Critically, listeners were less likely to fixate the target when they were in the Two Contrast-Shared condition than when they were in the Two Contrast-Privileged condition. This latter finding shows that listeners successfully took into account the

---

[6] Note that eye-gaze analyses in the visual world paradigm sometimes use a proportion measure (proportion of target fixations in a time-window). We used a different approach because inspection of the trial-by-trial proportion-based measure revealed that the data distribution was highly zero-inflated and would violate the linear model assumption of normally distributed residuals. Further, a binary measure of comprehension is more analogous to the production measure used in this article, which is also binary (whether speakers use an adjective or not). Because our goal here is to compare these two measures, the binary comprehension measure is preferred.

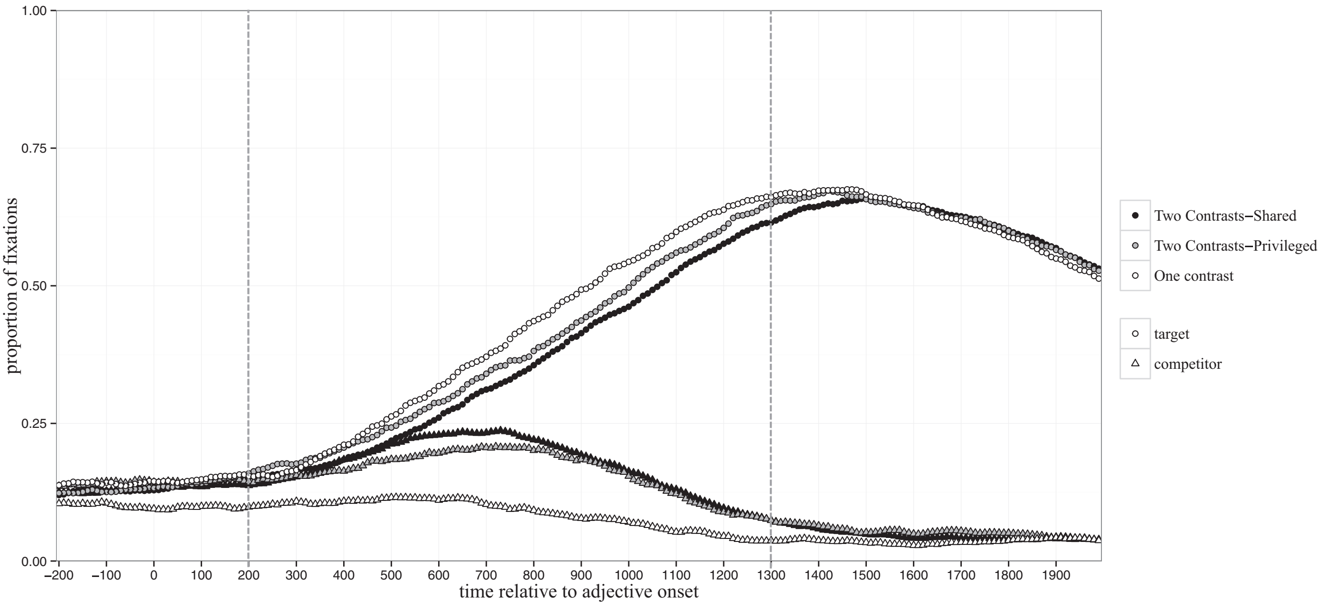

Figure 3. Proportion of fixations to targets and competitors over time, during interpretation of an instruction (e.g., "Click on the big banana"). The gray dashed lines indicate the approximate time-window used for analysis, beginning 200 ms after the onset of the adjective (e.g., "big").

speaker's perspective, reducing consideration of the contrast object (big balloon) when the small balloon was privileged ground (as there would be no reason for the speaker to say the size of the balloon in that case).

To examine individual differences related to this comprehension measure it is important to first assess its reliability as a measure of perspective-taking. To determine how consistent these effects are within a subject, we split the data in two halves based on odd and even trial numbers and extracted each subject's random slopes for both contrasts. The Spearman-Brown adjusted correlation between the Two Contrasts versus One Contrast random slopes of each half was $\rho^* = 0.29$. The Spearman-Brown adjusted correlation between the Shared versus Privileged random slopes of each half was $\rho^* = 0.49$[7]. Note that a measure is typically considered internally reliable if its reliability is larger than 0.7 (Nunnally, 1978); this convention is derived primarily from psychometric testing and may not offer a useful benchmark for experimental measures of language processing. To our knowledge, there has been very little prior investigation of the reliability of experimental effects in the language domain. This level of reliability does, however, place an upper bound on the magnitude of the relationship that this measure can bear with other measures.

Language production. The dependent measure used to index speakers' appreciation of their partner's perspective was the presence or absence of a size adjective in their production of an instruction (e.g., Click on the big banana vs. Click on the banana), in response to a target item being circled on the screen. On average, participants produced a size adjective 97% of the time in the Contrast-Shared condition, 65% of the time in the Contrast-Privileged condition, and 8% of the time in the No Contrast condition. Recall that successful perspective-taking would be indicated by *not* using a size-adjective in the Contrast-Privileged condition (as the listener only saw one, e.g., banana).

Adjective use was analyzed in a multilevel logistic regression. The three conditions were entered in the fixed effects as weighted orthogonal contrast codes. The first contrast compared Contrast-Shared and Contrast-Privileged with the No Contrast condition, and examines sensitivity of adjective use to the presence of a second size-contrast in the scene. The second contrast was our critical comparison, and compared Contrast-Shared with Contrast-Privileged. Successful perspective-taking would be indicated by fewer size-adjectives in the Contrast-Privileged condition. Subjects and cohort pairs (e.g., banana-balloon) were entered as crossed random effects with by-subject and by-cohort pair random slopes for condition.

Participants were more likely to use a size adjective when they were in one of the Contrast conditions than in the No Contrast condition (see Table 2). Critically, speakers took perspective: they were significantly less likely to mention target object size in the Contrast-Privileged condition versus the Contrast-Shared condition.

To determine how consistent these effects are within a subject, we split the data in two halves based on odd and even trial numbers and extracted each subject's random slopes for both contrasts. The Spearman-Brown adjusted correlation between the Contrast versus No Contrast random slopes of each half was $\rho^* = 0.68$. The Spearman-Brown adjusted correlation between the Shared versus Privileged random slopes of each half was $\rho^* = 0.89$. As opposed to the comprehension measure, the production task provided a highly reliable measure of individual perspective-taking ability.

---

[7] Reliabilities for the individual conditions were higher (Two Contrasts-Shared: $\rho^* = 0.6$, Two Contrasts-Privileged: $\rho^* = 0.6$, One Contrast: $\rho^* = 0.7$).

Table 1
Effect of Condition on Binary Measure of Target Fixations

Fixed effects	β	*SE*	*z*-value	*p*-value
(Intercept)	1.39	0.06	21.72	$<2e^{-16}$
Two vs. One Contrast(s)	−0.41	0.12	−3.49	<.01
Shared vs. Privileged	−0.25	0.12	−2.15	.03*

Random effects	Variance
Subjects	
(Intercept)	0.23
Two vs. One Contrast(s)	0.28
Shared vs. Privileged	0.07
Cohort pairs	
(Intercept)	0.17
Two vs. One Contrast(s)	0.68
Shared vs. Privileged	0.77

Note. Number of observations: 8,466; groups: subjects, 152; cohort pairs, 96.
* Indicates a significant effect at an α level of 0.05.

Individual Differences and Their Link to Perspective-Taking in Three Domains

Descriptive statistics and reliabilities for the individual differences measures can be found in Appendix B. Pearson product–moment correlations between all measures of individual differences are summarized in the correlation matrix in Table 3.

Mixed-effects logistic regression models were used to assess the contributions of individual differences in inhibitory control, working memory, recall, and CFQ to performance in the cue generation (see Table 4), comprehension (see Table 5), and production (see Table 6) tasks. Figure 4 summarizes the relationships between the individual difference measures and the three perspective-taking tasks. The multilevel modeling approach provides conservative parameter estimates that are adjusted for multiple comparisons (Gelman, Hill, & Yajima, 2012). Nonetheless, given that three separate models were fit to the same set of predictors, we adjusted the α level for the set of three analyses to follow using the Bonferroni correction for multiple comparisons, resulting in an overall α level of 0.05/3 = 0.016.

Cue generation. Accuracy on each item (i.e., whether each target word was correctly or incorrectly recalled) was used as the dependent measure of performance on the cue generation task. A mixed-effects logistic regression was fit with random effects for subjects and items (i.e., target words). Successful cue generation was predicted by Operation Span and Cued Recall. Participants with larger working memory span and better recall with experimenter-provided cues were more successful at generating cues for their future selves.

Comprehension. For the comprehension task, as before, binary target fixations within the critical time-window were used as the dependent variable. Subjects and Cohort Pairs were entered as random effects. The maximal random effects structure justified by the data was fit. The Shared versus Privileged condition effect and its interactions with all the measures of individual differences were

Table 2
Effect of Condition on Production of Size Adjectives

Fixed effects	β	*SE*	*z*-value	*p*-value
(Intercept)	1.49	0.13	11.44	$<2e^{-16}$
Contrast vs. No Contrast	6.91	0.24	28.70	$<2e^{-16*}$
Shared vs. Privileged	3.13	0.27	11.50	$<2e^{-16*}$

Random effects	Variance
Subjects	
(Intercept)	1.04
Contrast vs. No Contrast	2.74
Shared vs. Privileged	6.57
Cohort pairs	
(Intercept)	0.51
Contrast vs. No Contrast	1.48
Shared vs. Privileged	1.14

Note. Number of observations: 9,685; groups: subjects, 152; cohort pairs, 96.
* Indicates a significant effect at an α level of 0.05.

entered as fixed effects into a single model. No measure of individual differences interacted with the Shared versus Privileged effect, indicating that our measures of executive function and memory did not predict perspective-taking ability in comprehension. The lack of significant predictors may be partly because of the low reliability of the comprehension measure; the internal reliability of a measure places an upper bound on how much of a relationship there can be between that measure and other variables.

Production. In the production task, as before, presence or absence of the size adjective in a participant's utterance was used as the dependent variable. Subjects and cohort pairs (e.g., banana-balloon) were entered as random effects. The maximal random effects structure justified by the data was fit. Shared versus Privileged condition and all the measures of individual differences were entered as fixed effects into a single model. Operation Span performance interacted with the Shared versus Privileged effect, suggesting that participants with higher working memory capacity have a larger Shared versus Privileged effect—they are less likely to use the size adjective when it is overinformative from the listener's perspective (Contrast-Privileged) and more likely to use the size adjective when it is appropriate (Contrast-Shared). Operation Span is the only measure that predicts perspective-taking ability in production.

Do Individual Differences in Perspective-Taking Correlate Across Domains?

Individual measures of perspective-taking in language comprehension were created by extracting the best linear unbiased predictor for each participant's random coefficient corresponding to the Two Contrast-Shared versus Two Contrast-Privileged contrast. In other words, the coefficient indicated how much *more* each participant fixated the target (e.g., big banana) when the competitor (e.g., big balloon) was not as plausible by virtue of its size contrast (small balloon) being in privileged ground compared with when all relevant objects were in common ground. Similarly, individual measures of perspective-taking in language production were created by extracting each participant's random coefficient corresponding to the Contrast-Shared versus Contrast-Privileged contrast. In other words, the coefficient indicated how much *less* likely each participant was to produce a size adjective (e.g., "Click on the [big] banana") when the other pair member (e.g., small banana) was in privileged ground compared to when it was in common ground. Participants' centered and standardized cue generation score was considered to be an individual measure of their perspective-taking during cue generation.

Table 3
Correlation Matrix for All Individual Difference Measures

	Stroop One	Stroop Two	Minus Two Span	Operation Span	Cued Recall	CFQ
Stroop One	—					
Stroop Two	0.07	—				
Minus Two Span	0.03	0.10	—			
Operation Span	0.21	0.09	0.30*	—		
Cued Recall	0.22	0.02	0.21	0.32*	—	
CFQ	0.01	0.02	0.02	0.02	0.18	—

Note. CFQ = Cognitive Failures Questionnaire. All scores were centered and standardized. The Stroop One scores were first logit-transformed because they were proportions. Asterisks indicate bivariate correlations significant at a Bonferroni-corrected α level of 0.05/15 = .003. See Appendix C for disattenuated correlations.

Table 4
Effects of Individual Measures of Executive Function, Memory, CFQ on Successful Cue Generation

Fixed effects	β	*SE*	*z*-value	*p*-value
(Intercept)	0.14	0.11	1.27	.21
Stroop One	0.11	0.06	1.89	.06
Stroop Two	−0.13	0.06	−2.20	.03
Operation Span	0.17	0.06	2.72	.01*
Minus Two Span	0.00	0.06	−0.04	.97
Cued Recall	0.36	0.06	5.71	$1e^{-8*}$
CFQ	0.12	0.06	2.15	.03
Random effects	**Variance**			
Subjects				
(Intercept)	0.39			
Items				
(Intercept)	0.71			

Note. CFQ = Cognitive Failures Questionnaire. Number of observations: 11,440; groups: subjects, 143; items, 80.
* Indicates a significant effect at the corrected α level of 0.016.

Pearson's product-moment correlations suggest that individual variation is not significantly correlated among the three perspective-taking tasks (see Figure 4). Note that with our sample size (N = 152), power to detect an effect size of r = .2 is 70% (at α = .05, two-tailed). Perspective-taking in production was marginally correlated with perspective-taking in cue generation, r = .16, p = .05. However, perspective-taking in production was not correlated with perspective-taking in comprehension, $r = -0.06$, p = .43, nor was perspective-taking in cue generation correlated with perspective-taking in comprehension, $r = -0.07$, p = .43.[8] In a supplemental confirmatory factor analysis, we find that a model with a single latent perspective-taking variable provides a worse fit than a model with three domain-specific latent variables (see Appendix D). Finally, to look more closely at the relationship between cue generation and production, a logistic mixed-effects model was fit with cue generation performance (i.e., whether a target was correctly or incorrectly recalled) as the dependent variable and all the individual differences measures as fixed effects (as in Table 6), as well as the individual difference measure of perspective-taking during production (see Appendix E). Perspective-taking during production was not independently related (β = 0.06, p = .30) to successful cue generation, when controlling for executive function and memory. Therefore, to the small degree that these skills appear to be related, the nature of that relationship appears to be a shared reliance on the cognitive processes assayed by operation span.

---

[8] Correcting for attenuation, there is a correlation of r^* = 0.18 between perspective-taking in production and perspective-taking in cue generation, a correlation of $r^* = -0.09$ between perspective-taking in production and perspective-taking in comprehension, and a correlation of $r^* = -0.11$ between perspective-taking in comprehension and perspective-taking in cue generation.

Table 5

Effects of Individual Measures of Executive Function and Memory on Eye-Movements During Comprehension

Fixed effects	β	*SE*	*z*-value	*p*-value
(Intercept)	1.30	0.07	18.91	$<2e^{-16}$*
Shared vs. Privileged	−0.26	0.12	−2.26	.02†
Stroop One	0.10	0.05	2.03	.04
Stroop Two	0.01	0.05	0.21	.83
Operation Span	−0.00	0.05	−0.05	.96
Minus Two Span	−0.10	0.05	−2.07	.04
Cued Recall	0.02	0.05	0.32	.75
CFQ	−0.01	0.05	−0.30	.76
Shared vs. Privileged × Stroop One	0.02	0.07	0.23	.82
Shared vs. Privileged × Stroop Two	−0.05	0.07	−0.71	.48
Shared vs. Privileged × Operation Span	0.00	0.08	0.02	.99
Shared vs. Privileged × Minus Two Span	−0.03	0.08	−0.42	.67
Shared vs. Privileged × Cued Recall	−0.12	0.08	−1.48	.14
Shared vs. Privileged × CFQ	0.06	0.08	0.81	.42

Random effects	Variance
Subjects	
(Intercept)	0.16
Shared vs. Privileged	0.08
Cohort pairs	
(Intercept)	0.24
Shared vs. Privileged	0.75

Note. CFQ = Cognitive Failures Questionnaire. Number of observations: 6,049; groups: subjects, 144; cohort pairs, 96.

† Indicates a marginal effect at the corrected α level of 0.016. * Indicates a significant effect.

Discussion

Taking perspective is a critically important ability in a wide variety of tasks. Our results extend findings from the literature on perspective-taking in conversation, as well as the literature on cue generation. Individuals successfully take the perspectives of others during production and comprehension tasks and the perspective of their future self during cue generation tasks. However, we find little evidence of a domain-general ability that underlies these three perspective-taking tasks.

This lack of a relationship is all the more surprising when viewed alongside the impressive effects of perspective-taking that each individual task revealed. In the cue-generation memory task, participants generated memory cues for themselves to use during later retrieval. Learners recalled an average of 41 targets (out of 80), even after a 2-day delay. Cues that led to successful retrieval were strongly associated to the target (e.g., cash–money) and more idiosyncratic (e.g., roommate–sophia), suggesting that participants effectively tailored their cues to their future knowledge state by creating cues that are unique to their long-term semantic memory.

Equally impressive perspective-taking was evident during the conversation task. We found that, while speaking, participants took into account the perspective of their conversation partner very effectively. Speakers produced more size adjectives (e.g., “Click on the *big* banana.”) when their partner could see a big and a small version of the target object (e.g., a big and a small banana) than when the partner could only see one version of the target object. Critically, this was true even in cases where the speaker’s own display did contain two versions of the target object (e.g., a big banana in common ground and a small banana in privileged ground), suggesting that participants successfully represent the difference between their perspective and their partner’s. Furthermore, participants were reliable in their ability to make use of perspective information across their utterances.

Participants’ eye movements during comprehension indicated that listeners took into account what information is privileged and what is shared as well. During interpretation of stimulus sentences, participants were more likely to fixate the correct target (e.g., the big banana) when there was only one item that the size adjective (e.g., big) could be referring to given the speaker’s perspective (e.g., the speaker can see a small banana and a big banana, and only one version of every other item). Critically, this was true even when the listener’s display—but not the speaker’s—contained two potential referents for the size adjective (e.g., the listener sees a small banana and a big banana, as well as a big balloon, and a small balloon in privileged ground). To succeed, listeners both formed a representation of the speaker’s differing perspective and brought that information to bear during interpretation of the speaker’s utterances. While the group level analyses demonstrated a significant perspective-taking effect, this comprehension measure did not prove to be a reliable measure of *individual* perspective-taking ability. This highlights a key difference between the language production and comprehension processes and the role of perspective information in each, a theme that we discuss in further detail below.

Perspective-Taking Across Cognitive Domains

Despite their common reliance on some form of mental self-projection, we did not find evidence that performance on the three

Table 6
Effects of Individual Measures of Executive Function and Memory on Adjective Use During Production

Fixed effects	β	*SE*	*z*-value	*p*-value
(Intercept)	3.37	0.13	25.19	$<2e^{-16}$
Shared vs. Privileged	3.25	0.27	12.06	$<2e^{-16}$*
Stroop One	−0.14	0.12	−1.17	.24
Stroop Two	−0.12	0.11	−1.04	.30
Operation Span	0.01	0.12	0.06	.96
Minus Two Span	−0.04	0.11	−0.40	.69
Cued Recall	0.06	0.12	0.48	.63
CFQ	−0.13	0.11	−1.13	.26
Shared vs. Privileged × Stroop One	−0.03	0.25	−0.12	.90
Shared vs. Privileged × Stroop Two	−0.09	0.25	−0.35	.73
Shared vs. Privileged × Operation Span	0.68	0.27	2.53	.01*
Shared vs. Privileged × Minus Two Span	0.18	0.25	0.70	.48
Shared vs. Privileged × Cued Recall	0.29	0.27	1.08	.28
Shared vs. Privileged × CFQ	−0.16	0.25	−0.64	.52

Random effects	Variance
Subjects	
(Intercept)	0.84
Shared vs. Privileged	5.81
Cohort pairs	
(Intercept)	0.43
Shared vs. Privileged	1.07

Note. CFQ = Cognitive Failures Questionnaire. Number of observations: 6,888; groups: subjects, 144; cohort pairs, 96.
* Indicates a significant effect at the corrected α level of 0.016.

main tasks was governed by an underlying perspective-taking process. There was little relationship between the three measures, particularly when other cognitive abilities, such as executive function and basic recall, were controlled for. Of the six individual difference measures used, only performance on the Operation span task was significantly predictive of perspective-taking ability across tasks.[9] Specifically, participants who performed well on the Operation Span task were more likely to engage in perspective-taking during production and cue generation.

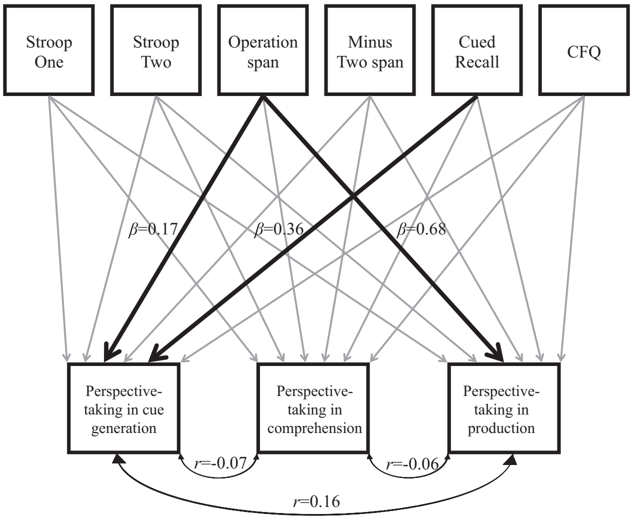

Figure 4. Relationships among perspective-taking tasks and individual difference measures. CFQ = Cognitive Failures Questionnaire. Bolded black arrows indicate significant effects.

Operation span is intended to measure the ability to maintain multiple items in memory while performing a distracting task (Unsworth et al., 2005). Perspective-taking during production may engage similar mechanisms, requiring individuals to hold active two representations of the visual environment, their own and that of their conversation partner, and to design utterances taking the partner's representation into account. Similarly, successful cue generation may rely on individuals' ability to consider their current and future cognitive contexts and ignore cues that will change before the future cognitive context. More important, the contributions of executive function are independent of the role of basic recall, consistent with the proposal that generating successful cues for later retrieval engages both long-term memory and executive function components.

Despite their mutual correlation with working memory, perspective-taking in production and perspective-taking in cue generation are very weakly related. This suggests that the two perspective-taking tasks are tapping into different aspects of the cognitive processes recruited by the operation span task; as a complex span task, operation span engages many subprocesses including attentional control, goal maintenance, and memory search (Unsworth, Brewer, & Spillers, 2009; Unsworth & Engle, 2007). Further exploration of the individual differences underlying

---

[9] Several of the individual difference measures are correlated with Operation Span and so, in the context of multiple regression, the absence of their relationship with measures of perspective-taking may be because of multicollinearity.

perspective-taking in both of these domains is needed to ascertain the extent to which production and cue generation might each place unique demands on working memory processes.

A Comprehension-Production Dissociation

It is noteworthy that we see no relationship ($r = -.06$) between perspective-taking during comprehension and perspective-taking during production. This finding, though surprising, is consistent with a previous finding with children (Nilsen & Graham, 2009). Using a similar paradigm to what was used here, Nilsen and Graham found that children's performance on executive function tasks predicted their ability to take a speaker's perspective into consideration during comprehension, as well as their ability to produce referential expressions that take the listener's view into account. However, they found that no measures of the children's performance on the comprehension task were related to their performance on the production task.

A partial explanation for this comprehension-production dissociation may come from the fact that, as an index of individual perspective-taking ability, our comprehension measure appears to be low in internal reliability (split-half reliability is $\rho^* = 0.49$). In contrast, the production measure is highly reliable ($\rho^* = 0.89$). The relatively low reliability of the comprehension measure may be due to the fact that it is derived from eye fixations, which, given their cognitively inexpensive nature, may be more variable from trial to trial than explicit verbal responses. However, gaze measures of speech perception have been shown to have high internal reliability (Farris-Trimble & McMurray, 2013), suggesting that eye-tracking can, in principle, be used fruitfully to examine individual differences (see also Henderson & Luke, 2014).

Another possibility is that the distinction in reliability between production and comprehension may point to a potential difference between the mechanisms underlying language production and language comprehension in dialogue. Both processes must involve the activation of multiple representations. During the production process, the relevant representation (e.g., the partner's perspective) can be in the focus of attention and all other representations can be temporarily ignored for the duration of utterance planning. It may seem that, in a similar fashion, during comprehension, listeners should attend only to what they know to be the speaker's viewpoint and ignore all other representations. However, this may be an inefficient strategy for the listener. If the speaker says something incorrect or inconsistent with the representation that is in the listener's attentional focus, the listener might find it challenging to recover the correct interpretation. Indeed, speakers in a conversation often produce underinformative or egocentric utterances with the expectation that a listener will provide feedback if a clarification is needed (Krauss & Weinheimer, 1966; Schober, 1993). It may be more efficient from the listener's standpoint to distribute attention more broadly to multiple representations and allow many candidate interpretations to remain partially activated throughout the producer's utterance. This may be particularly true in live conversation—which we elicited here—as compared with the comprehension of prerecorded or scripted lab stimuli. In a real conversation, listeners may infer that their partners can make mistakes when speaking. For example, listeners might accumulate direct evidence that a certain speaker is not always accurate (e.g., because the speaker is not adept at perspective-taking or simply did not pay close enough attention to the visual display), or they themselves might be error-prone speakers and attribute similar qualities to their partner, among many other options.

This view of comprehension is consistent with a wealth of findings that, as the interpretive process unfolds, many candidate interpretations remain active, from the level of the phonemes (Allopenna, Magnuson, & Tanenhaus, 1998; Dahan, Magnuson, & Tanenhaus, 2001) to semantics (Yee, Overton, & Thompson-Schill, 2009; Yee & Sedivy, 2006). These candidate representations even remain active (though to a lesser degree) once they should have been ruled out by prior information (e.g., Allopenna et al., 1998; Levy, Bicknell, Slattery, & Rayner, 2009; Nozari, Trueswell, & Thompson-Schill, 2015), potentially as a mechanism to facilitate reanalysis if the dominant interpretation turns out to be incorrect. In this way, comprehension may be a more noisy process than production, resulting in less internal reliability at the individual level (see also Yoon & Brown-Schmidt, 2013). Further work examining the relationship between production and comprehension processes is needed to tease apart process differences from measurement discrepancies.

The Role of Feedback

One key dimension on which production, comprehension, and memory cue generation differ is the potential for corrective feedback. As mentioned previously, speakers often produce ambiguous utterances with the expectation that their conversation partner will request clarification if the intended message is not transmitted successfully (Krauss & Weinheimer, 1966). Thus, perspective-taking ability is not the only determinant of a speaker's choice to use a modified expression (e.g., the *big* banana)—the speaker's willingness to say something potentially ambiguous also plays a key role. Speakers may vary in how often they place the communicative burden on the listener by resorting to underinformative or otherwise ambiguous utterances.

During comprehension, feedback is more continuous; as a sentence unfolds, some candidate interpretations are down-weighted and others are up-weighted, until one is (typically) selected. Each new "bit" of evidence provides feedback about whether the currently dominant interpretation remains consistent with the input. Listeners may vary in their willingness to maintain multiple possible interpretations active, instead of choosing one early on.

On the other hand, when generating memory cues for a future self, the opportunity for feedback is absent. When looking at a note like "call Rob" and drawing a blank on the intended meaning, one cannot query a past self about the topic of said phone call. Thus, successful cue generators must err on the side of overinformativity. This bias likely interacts with each individual's metacognitive awareness of what is likely to remain accessible in memory over time.

In conclusion, many activities of daily life require the representation of an alternate perspective, whether another person's or one's own in the future. We find that working memory, the ability to maintain multiple representations and ignore superfluous information, contributes to perspective-taking in both language production and memory, though these processes themselves are largely unrelated. Perspective-taking is not a single cognitive dimension. Instead, the ability to accurately make use of perspective information is governed by mechanisms that are specific to the representational domain.

These findings have important implications for our understanding of perspective-taking across a variety of cognitive activities. Future theories of perspective-taking must account for the effects of domain-specific goals, constraints, and memory demands. Though many processes have been thus far referred to under the "perspective-taking" umbrella, our work points to the need for more precise formulations of the underlying computations that take place in different contexts.

References

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements. *Evidence for Continuous Mapping Models, 439,* 419–439.

Anderson, J. R. (1974). Retrieval of propositional information from long-term memory. *Cognitive Psychology, 6,* 451–474. http://dx.doi.org/10.1016/0010-0285(74)90021-8

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). *lme4: Linear mixed-effects models using Eigen and S4*. R package version 1. 1–8, http://CRAN.R-project.org/package=lme4

Birch, S. A. J., & Bloom, P. (2007). The curse of knowledge in reasoning about false beliefs. *Psychological Science, 18,* 382–386. http://dx.doi.org/10.1111/j.1467-9280.2007.01909.x

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10,* 433–436. http://dx.doi.org/10.1163/156856897X00357

Brewer, G. A., Knight, J. B., Marsh, R. L., & Unsworth, N. (2010). Individual differences in event-based prospective memory: Evidence for multiple processes supporting cue detection. *Memory & Cognition, 38,* 304–311. http://dx.doi.org/10.3758/MC.38.3.304

Broadbent, D. E., Cooper, P. F., FitzGerald, P., & Parkes, K. R. (1982). The Cognitive Failures Questionnaire (CFQ) and its correlates. *British Journal of Clinical Psychology, 21,* 1–16. http://dx.doi.org/10.1111/j.2044-8260.1982.tb01421.x

Brown-Schmidt, S. (2009). The role of executive function in perspective taking during online language comprehension. *Psychonomic Bulletin & Review, 16,* 893–900. http://dx.doi.org/10.3758/PBR.16.5.893

Brown-Schmidt, S. (2012). Beyond common and privileged: Gradient representations of common ground in real-time language use. *Language and Cognitive Processes, 27,* 62–89. http://dx.doi.org/10.1080/01690965.2010.543363

Brown-Schmidt, S., & Fraundorf, S. (2015). Interpretation of informational questions modulated by joint knowledge and intonational contours. *Journal of Memory and Language, 84,* 49–74. http://dx.doi.org/10.1016/j.jml.2015.05.002

Brown-Schmidt, S., Gunlogson, C., & Tanenhaus, M. K. (2008). Addressees distinguish shared from private information when interpreting questions during interactive conversation. *Cognition, 107,* 1122–1134. http://dx.doi.org/10.1016/j.cognition.2007.11.005

Buckner, R. L., Andrews-Hanna, J. R., & Schacter, D. L. (2008). The brain's default network: Anatomy, function, and relevance to disease. *Annals of the New York Academy of Sciences, 1124,* 1–38. http://dx.doi.org/10.1196/annals.1440.011

Buckner, R. L., & Carroll, D. C. (2007). Self-projection and the brain. *Trends in Cognitive Sciences, 11,* 49–57. http://dx.doi.org/10.1016/j.tics.2006.11.004

Carlson, S. M., & Moses, L. J. (2001). Individual differences in inhibitory control and children's theory of mind. *Child Development, 72,* 1032–1053. http://dx.doi.org/10.1111/1467-8624.00333

Clark, H. H. (1992). *Arenas of language use*. Chicago, IL: University of Chicago Press.

Clark, H. H. (1996). *Using language*. New York, NY: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511620539

Clark, H. H., & Marshall, C. R. (1978). Reference diaries. In D. L. Waltz (Ed.), *TINLAP-2: Theoretical issues in natural language processing-2* (pp. 57–63). New York, NY: Association for Computing Machinery.

Dahan, D., Magnuson, J. S., & Tanenhaus, M. K. (2001). Time course of frequency effects in spoken-word recognition: Evidence from eye movements. *Cognitive Psychology, 42,* 317–367. http://dx.doi.org/10.1006/cogp.2001.0750

D'Argembeau, A., Ortoleva, C., Jumentier, S., & Van der Linden, M. (2010). Component processes underlying future thinking. *Memory & Cognition, 38,* 809–819. http://dx.doi.org/10.3758/MC.38.6.809

Farris-Trimble, A., & McMurray, B. (2013). Test-retest reliability of eye tracking in the visual world paradigm for the study of real-time spoken word recognition. *Journal of Speech, Language, and Hearing Research, 56,* 1328–1345. http://dx.doi.org/10.1044/1092-4388(2012/12-0145)

Feldman, S. M., & Underwood, B. J. (1957). Stimulus recall following paired-associate learning. *Journal of Experimental Psychology, 53,* 11–15. http://dx.doi.org/10.1037/h0048386

Ferguson, H. J., & Breheny, R. (2012). Listeners' eyes reveal spontaneous sensitivity to others' perspectives. *Journal of Experimental Social Psychology, 48,* 257–263. http://dx.doi.org/10.1016/j.jesp.2011.08.007

Fiechter, J. L., Benjamin, A. S., & Unsworth, N. (in press). The metacognitive foundations of effective remembering. In J. Dunlosky & S. K. Tauber (Eds.), *Oxford handbook of metamemory*. New York, NY: Oxford University Press.

Gelman, A., Hill, J., & Yajima, M. (2012). Why we (usually) don't have to worry about multiple comparisons. *Journal of Research on Educational Effectiveness, 5,* 189–211. http://dx.doi.org/10.1080/19345747.2011.618213

Genesee, F., Nicoladis, E., & Paradis, J. (1995). Language differentiation in early bilingual development. *Journal of Child Language, 22,* 611–631. http://dx.doi.org/10.1017/S0305000900009971

Hallett, P. (1986). Eye movements (and human visual perception). *Handbook of Perception and Human Performance, 1,* 10–11.

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,* 43–61. http://dx.doi.org/10.1016/S0749-596X(03)00022-6

Heller, D., Grodner, D., & Tanenhaus, M. K. (2008). The role of perspective in identifying domains of reference. *Cognition, 108,* 831–836. http://dx.doi.org/10.1016/j.cognition.2008.04.008

Henderson, J. M., & Luke, S. G. (2014). Stable individual differences in saccadic eye movements during reading, pseudoreading, scene viewing, and scene search. *Journal of Experimental Psychology: Human Perception and Performance, 40,* 1390–1400. http://dx.doi.org/10.1037/a0036330

Hill, P. F., & Emery, L. J. (2013). Episodic future thought: Contributions from working memory. *Consciousness and Cognition, 22,* 677–683. http://dx.doi.org/10.1016/j.concog.2013.04.002

Keysar, B., Lin, S., & Barr, D. J. (2003). Limits on theory of mind use in adults. *Cognition, 89,* 25–41. http://dx.doi.org/10.1016/S0010-0277(03)00064-7

Kliegel, M., Martin, M., McDaniel, M. A., & Einstein, G. O. (2002). Complex prospective memory and executive control of working memory: A process model. *Psychologische Beiträge, 44,* 303–318.

Krauss, R. M., & Weinheimer, S. (1966). Concurrent feedback, confirmation, and the encoding of referents in verbal communication. *Journal of Personality and Social Psychology, 4,* 343–346. http://dx.doi.org/10.1037/h0023705

Levy, R., Bicknell, K., Slattery, T., & Rayner, K. (2009). Eye movement evidence that readers maintain and act on uncertainty about past linguistic input. *Proceedings of the National Academy of Sciences of the United States of America, 106,* 21086–21090. http://dx.doi.org/10.1073/pnas.0907664106

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,* 551–556. http://dx.doi.org/10.1016/j.jesp.2009.12.019

Marsh, R. L., & Hicks, J. L. (1998). Event-based prospective memory and executive control of working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24,* 336–349. http://dx.doi.org/10.1037/0278-7393.24.2.336

Matthews, D., Lieven, E., Theakston, A., & Tomasello, M. (2006). The effect of perceptual availability and prior discourse on young children's use of referring expressions. *Applied Psycholinguistics, 27,* 403–422. http://dx.doi.org/10.1017/S0142716406060334

Nadig, A. S., & Sedivy, J. C. (2002). Evidence of perspective-taking constraints in children's on-line reference resolution. *Psychological Science, 13,* 329–336. http://dx.doi.org/10.1111/j.0956-7976.2002.00460.x

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (1998). *The University of South Florida word association, rhyme, and word fragment norms*. Retrieved from http://www.usf.edu/FreeAssociation/

Nilsen, E. S., & Graham, S. A. (2009). The relations between children's communicative perspective-taking and executive functioning. *Cognitive Psychology, 58,* 220–249. http://dx.doi.org/10.1016/j.cogpsych.2008.07.002

Nozari, N., Trueswell, J. C., & Thompson-Schill, S. L. (2015). *The interplay of local attraction, context and domain general cognitive control in activation and suppression of semantic distractors during sentence comprehension*. Manuscript under review.

Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York, NY: McGraw-Hill.

Onishi, K. H., & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science, 308,* 255–258. http://dx.doi.org/10.1126/science.1107621

Revelle, W. (2014). *Psych: Procedures for personality and psychological research, Northwestern University, Evanston, IL.* Retrieved from http://CRAN.R-project.org/package=psychVersion=1.4.8

Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling. *Journal of Statistical Software, 48,* 1–36.

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,* e92406. http://dx.doi.org/10.1371/journal.pone.0092406

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. http://dx.doi.org/10.1016/j.jml.2014.04.003

Saxe, R., & Kanwisher, N. (2003). People thinking about thinking people. The role of the temporo-parietal junction in "theory of mind." *NeuroImage, 19,* 1835–1842. http://dx.doi.org/10.1016/S1053-8119(03)00230-1

Schacter, D. L., Addis, D. R., & Buckner, R. L. (2007). Remembering the past to imagine the future: The prospective brain. *Nature Reviews Neuroscience, 8,* 657–661. http://dx.doi.org/10.1038/nrn2213

Schober, M. F. (1993). Spatial perspective-taking in conversation. *Cognition, 47,* 1–24. http://dx.doi.org/10.1016/0010-0277(93)90060-9

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

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science, 268,* 1632–1634. http://dx.doi.org/10.1126/science.7777863

Tullis, J. G., & Benjamin, A. S. (2015a). Cueing others' memories. *Memory & Cognition, 43,* 634–646. http://dx.doi.org/10.3758/s13421-014-0478-y

Tullis, J. G., & Benjamin, A. S. (2015b). Cue generation: How learners flexibly support future retrieval. [Advance online publication]. *Memory & Cognition*. http://dx.doi.org/10.3758/s13421-015-0517-3

Tulving, E., & Thomson, D. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review, 80,* 352–373. http://dx.doi.org/10.1037/h0020071

Unsworth, N., Brewer, G. A., & Spillers, G. J. (2009). There's more to the working memory capacity-fluid intelligence relationship than just secondary memory. *Psychonomic Bulletin & Review, 16,* 931–937. http://dx.doi.org/10.3758/PBR.16.5.931

Unsworth, N., Brewer, G. A., & Spillers, G. J. (2012). Variation in cognitive failures: An individual differences investigation of everyday attention and memory failures. *Journal of Memory and Language, 67,* 1–16. http://dx.doi.org/10.1016/j.jml.2011.12.005

Unsworth, N., Brewer, G. A., & Spillers, G. J. (2013). Working memory capacity and retrieval from long-term memory: The role of controlled search. *Memory & Cognition, 41,* 242–254. http://dx.doi.org/10.3758/s13421-012-0261-x

Unsworth, N., & Engle, R. W. (2007). The nature of individual differences in working memory capacity: Active maintenance in primary memory and controlled search from secondary memory. *Psychological Review, 114,* 104–132. http://dx.doi.org/10.1037/0033-295X.114.1.104

Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods, 37,* 498–505. http://dx.doi.org/10.3758/BF03192720

Unsworth, N., & Spillers, G. J. (2010). Variation in working memory capacity and episodic recall: The contributions of strategic encoding and contextual retrieval. *Psychonomic Bulletin & Review, 17,* 200–205. http://dx.doi.org/10.3758/PBR.17.2.200

Wagner, A. D., Shannon, B. J., Kahn, I., & Buckner, R. L. (2005). Parietal lobe contributions to episodic memory retrieval. *Trends in Cognitive Sciences, 9,* 445–453. http://dx.doi.org/10.1016/j.tics.2005.07.001

Wardlow, L. (2013). Individual differences in speakers' perspective taking: The roles of executive control and working memory. *Psychonomic Bulletin & Review, 20,* 766–772. http://dx.doi.org/10.3758/s13423-013-0396-1

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition, 13,* 103–128. http://dx.doi.org/10.1016/0010-0277(83)90004-5

Yee, E., Overton, E., & Thompson-Schill, S. L. (2009). Looking for meaning: Eye movements are sensitive to overlapping semantic features, not association. *Psychonomic Bulletin & Review, 16,* 869–874. http://dx.doi.org/10.3758/PBR.16.5.869

Yee, E., & Sedivy, J. C. (2006). Eye movements to pictures reveal transient semantic activation during spoken word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32,* 1–14. http://dx.doi.org/10.1037/0278-7393.32.1.1

Yoon, S., & Brown-Schmidt, S. (2013). Lexical differentiation in language production and comprehension. *Journal of Memory and Language, 69,* 397–416. http://dx.doi.org/10.1016/j.jml.2013.05.005

Yoon, S. O., & Brown-Schmidt, S. (2014). Adjusting conceptual pacts in three-party conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40,* 919–937. http://dx.doi.org/10.1037/a0036161

Yoon, S. O., Koh, S., & Brown-Schmidt, S. (2012). Influence of perspective and goals on reference production in conversation. *Psychonomic Bulletin & Review, 19,* 699–707. http://dx.doi.org/10.3758/s13423-012-0262-6

Appendix A

Analysis of Cue Characteristics by Success of the Cue

Cue characteristic	Correct cues	Incorrect cues	Welch t test
Cue to target associative strength	$M = .08$, $SD = .05$	$M = .02$, $SD = .02$	$t(183.7) = 14.42$, $p < 2\text{e}^{-16}$
Cue in database	$M = .47$, $SD = .17$	$M = .58$, $SD = .17$	$t(299.9) = -5.58$, $p = 5.5\text{e}^{-8}$
Number of associates from cue	$M = 6.11$, $SD = 2.23$	$M = 8.18$, $SD = 2.60$	$t(293.3) = -7.42$, $p = 1.3\text{e}^{-12}$
Total associative strength from cue	$M = .38$, $SD = .14$	$M = .46$, $SD = .14$	$t(300) = -5.24$, $p = 3.05\text{e}^{-7}$

Note. For more details about these measures, see Tullis and Benjamin (2015b).

Appendix B

Descriptive Statistics for Six Measures of Individual Differences

	Mean (*SD*)	Reliability
Stroop One	Proportion correct: 0.85 (0.08)	$\alpha = .81$, 95% CI [0.76, 0.86]
Stroop Two	Interference in ms: 90.7 (143.8)	$\alpha = .8$, 95% CI [0.75, 0.85]
Minus Two Span	Score out of 12: 10.26 (1.12)	$\rho^{*} = .80$
Operation Span	Score out of 15: 11.13 (2.90)	$\alpha = .87$, 95% CI [.74, 1]
Cued Recall	Number correct out of 40: 17.62 (8.47)	$\alpha = .88$, 95% CI [0.85, 0.92]
CFQ	Sum of responses (maximum 100): 43.51 (11.60)	$\alpha = .85$, 95% CI [0.81, 0.89]

Note. CFQ = Cognitive Failures Questionnaire. Minus Two Span can only be split in one way so the Spearman-Brown prediction formula is used to measure reliability rather than coefficient α.

Appendix C

Correlation Matrix for all Individual Difference Measures Corrected for Attenuation

	Stroop One	Stroop Two	Minus Two Span	Operation Span	Cued Recall	CFQ
Stroop One	—					
Stroop Two	0.09	—				
Minus Two Span	0.04	0.13	—			
Operation Span	0.25	0.11	0.36	—		
Cued Recall	0.26	0.02	0.25	0.37	—	
CFQ	0.01	0.02	0.02	0.02	0.21	—

Note. CFQ = Cognitive Failures Questionnaire. All scores were centered and standardized. The Stroop One scores were first logit-transformed because they were proportions.

(*Appendices continue*)

Appendix D

An Additional Test of the Domain-Generality of Perspective-Taking Using Factor Analysis

To provide an additional test of the domain-general hypothesis, we conducted confirmatory factor analysis, using the *lavaan* package in R (Rosseel, 2012), and compared a model with one latent perspective-taking factor (Figure D1) to a model with three latent variables (Figure D2), one for each of the domains (comprehension, production, and memory). We created two observed measures of perspective-taking for each domain by taking the odd-even split of the data. This was done because the comprehension and production measures consist of each individual's condition effect (Two Contrasts-Shared vs. Two Contrasts-Privileged and Contrast-Shared vs. Contrast-Privileged, respectively). One limitation of this analysis is that we have only two indicators per latent variable for the second model, which can lead to improper solutions. However, splitting the data further than we have done here would create very noisy measures of perspective-taking in comprehension and production because the individual effects would be computed over a very low number of trials (approximately four per condition). Thus, the results of this analysis should be interpreted with this limitation in mind.

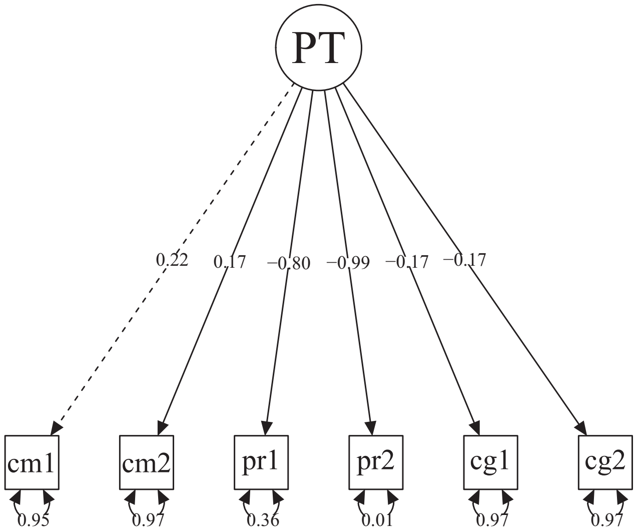

Figure D1. Path diagram of a single factor model with standardized parameter estimates. (PT = perspective-taking; cm1 = first comprehension measure; cm2 = second comprehension measure; pr1 = first production measure; pr2 = second production measure; cg1 = first cue generation measure; cg2 = second cue generation measure).

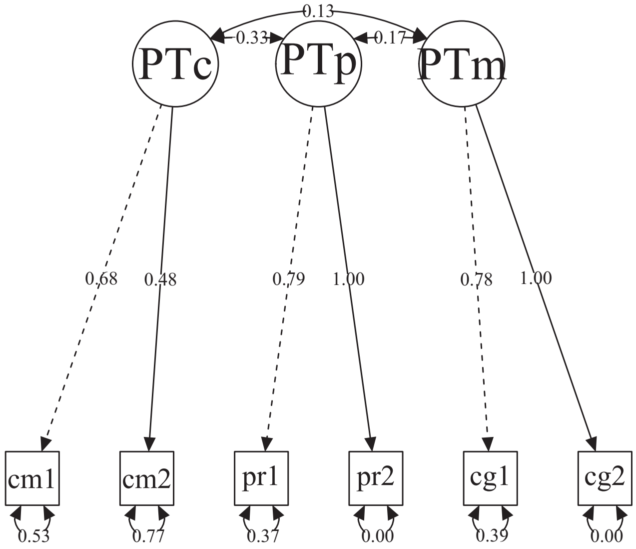

Figure D2. Path diagram of a three-factor model with standardized parameter estimates. Because of the low number of indicators, it was necessary to set inequality constraints on error variances for pr2 and cg2 to be larger than 0. (PTc = perspective-taking in comprehension; PTp = perspective-taking in production; PTm = perspective-taking in memory; cm1 = first comprehension measure; cm2 = second comprehension measure; pr1 = first production measure; pr2 = second production measure; cg1 = first cue generation measure; cg2 = second cue generation measure).

The model with a single latent perspective-taking variable did not provide a good fit to the data ($\chi^2(9, N = 151) = 159.59$, $p < .001$, standardized root mean square residual [SRMR] = 0.18, Comparative Fit Index [CFI] = 0.52, Akaike Information Criterion [AIC] = 2418.39). A model with latent variables for perspective-taking in comprehension, perspective-taking in production, and perspective-taking in memory provided an adequate fit ($\chi^2(6, N = 151) = 5.20$, $p = 0.52$, SRMR = 0.03, CFI = 1.00, AIC = 2270.01). Model comparison using a χ^2 difference test indicated that the three-factor model provided a significantly better fit to the data ($\Delta\chi^2(3, N = 151) = 154.39$, $p < 2.2e^{-16}$). This result provides further evidence that there is little covariance between the three measures of perspective-taking and fails to support a domain-general account of perspective-taking.

(*Appendices continue*)

Appendix E

Effects of Perspective-Taking During Production, Executive Function, Memory, and CFQ on Successful Cue Generation

Fixed effects	β	*SE*	*z*-value	*p*-value
(Intercept)	0.14	0.11	1.26	.21
P-T in production	0.06	0.06	1.05	.30
Stroop One	0.11	0.06	1.89	.06
Stroop Two	−0.13	0.06	−2.27	.02
Operation Span	0.16	0.06	2.51	$.01^{\dagger}$
Minus Two Span	−0.01	0.06	−0.11	.91
Cued Recall	0.35	0.06	5.63	$1.6e^{-8*}$
CFQ	0.12	0.06	2.17	.03
Random effects	**Variance**			
Subjects				
(Intercept)	0.39			
Items				
(Intercept)	0.71			

Note. CFQ = Cognitive Failures Questionnaire. Number of observations: 11,440; groups: subjects, 143; items, 80. Significance is evaluated relative to a Bonferroni-corrected α level of 0.05/7 = .007.
† indicates a marginal effect. * indicates a significant effect.

Received February 24, 2015
Revision received June 4, 2015
Accepted June 10, 2015 ■