Published in final edited form as:

Psychon Bull Rev. 2013 August ; 20(4): . doi:10.3758/s13423-013-0396-1.

Individual Differences in Speaker's Perspective Taking: The Roles of Executive Control and Working Memory

Liane Wardlow

University of California, San Diego Department of Psychology

Abstract

Both speaking and listening require taking into account the perspective of one's communicative partner. Is perspective taking a domain-specific process internal to the language production and comprehension systems? Or, is perspective taking a domain-general process regulated by the same mechanisms that are used to regulate other forms of behavior? Past research shows listeners' perspective taking is at least partially regulated by inhibitory control and working memory (WM), suggesting that it may be best thought of in domain-general terms. The current experiment uses individual differences in nonlinguistic executive functioning to assess whether domain general mechanisms help regulate speakers' perspective taking behavior. Sixty speakers participated in a referential communication task and tasks measuring WM and executive control (EC). Results revealed that WM and EC were both predictive of speakers' perspective taking, suggesting that perspective taking may be regulated by domain-general mechanisms.

Keywords

Perspective; Language Production; Common Ground; Referential Communication; Executive Function; Working Memory

1.0. Introduction

Communication allows people to interact and build meaningful relationships by providing a means to exchange thoughts, feelings and information. The processes of producing and comprehending language are fundamentally different; one involves producing ordered sequential sounds that convey an intended meaning and the other involves translating heard sounds into meaning. Yet, there are areas of overlap where both speakers and listeners engage in similar processes. One of these areas is perspective taking; both speakers and listeners use the perspective of their conversational partner to guide their language use. This is especially true when particular objects are being referred to: a process called *referential communication*. Perspective taking guides speakers' word choices when they refer to an object and it serves to narrow the possible referents listeners will consider after hearing a reference (Krauss & Fussell, 1991).

One way that speakers and listeners use perspective to shape their language use is by separating information into two broad categories: *common ground*, which refers to mutually

Please address correspondence to: Liane Wardlow, PhD Department of Psychology 0109 University of California, San Diego La Jolla, CA 92093-0109 lwardlow@ucsd.edu.

¹Nilsen and Graham also assessed reference production in children. However, they tested a type of perspective taking situation fundamentally different from that assessed here, and so this is not discussed further.

²Forward and backward span scores each correlated with modifier use on privileged ground trials and correlated with each other.

known information; and *privileged ground*, which refers to information known to only to the speaker or listener. Processing perspective requires computing or inferring perspective differences (e.g., whether information is in common or privileged ground) and then integrating that information into communicative behavior (cf., Clark & Marshall, 1981). Speakers' and listeners' abilities to make computations and inferences about perspective differences vary, resulting in occasional failures to use perspective to guide their referential behaviors. Little research has directly examined the factors that may be predictive of perspective-taking performance for speakers. Although previous research implicates a role for domain general processes in listeners' perspective taking (Lin, Keysar & Epley, 2010; Brown-Schmidt, 2009) such processes have not been examined with respect to production.

What are the general processes for producing language and how do they differ from comprehension? Producing an utterance begins with a to-be-expressed thought. That thought then activates the associated features that define the to-be-expressed information to construct a preverbal message, at a level called message encoding (Bock, 1982; Levelt, 1989). The most highly activated of those features, the ones that when expressed will enable the speaker to express the intended message, will be selected for further processing at the next stage, grammatical encoding. At the grammatical encoding stage, selected feature are mapped onto syntactic structures and lexical items. Those items are then sent to a phonological/articulatory stage at which the linguistic features are sounded out for production. Finally, before and during utterance production, a monitoring stage monitors the output to check that the to-be-conveyed message communicates what was intended at conceptual encoding.

In contrast, comprehension proceeds from the perception of sounds to the determination of meaning. Making comprehension difficult are the facts that there are generally no markers indicating where a particular word begins and ends and the acoustic signal indicating a particular word can vary by speaker and context. Even so, within a syllable or two, listeners have made decisions regarding the word's identity, syntactic class, aspects of which syntactic phrase the word belongs, and how it relates semantically to previous input (Garrett, 1990). As such, listeners must make quick decisions regarding whether information is in common or privileged ground.

In terms of perspective taking, speakers must account for perspective differences when formulating an utterance. This may occur in the initial stages of production, such as during message encoding, or as part of a later process, such as a post-grammatical-encoding monitoring process (c.f., Brennan & Hanna, 2009; Horton & Keysar, 1996; Keysar, Barr & Horton, 1998). In contrast, when listeners hear a reference, their task is to search for that referent, a process that may include restricting the searched space based upon perspective differences or only consulting perspective as part of a monitoring process (c.f., Brown-Schmidt, 2009; Keysar, Barr, Balin & Paek, 1998). That differences exist between the processes involved in accounting for perspective differences in production and comprehension suggest that one cannot directly infer that the same cognitive mechanisms that support perspective taking in comprehension will also support perspective taking in production.

The current experiment asks whether domain general mechanisms, in this case working memory (WM) and executive control (EC), are predictive of speakers' perspective taking. WM and EC are *executive functions*, which are a group of higher-order cognitive processes used to plan and execute complex tasks (Miyake et al., 2000; Pennington & Ozonoff, 1996). Executive functions include processes such as WM, EC, attention, planning, reasoning, and mental flexibility.

WM refers to a brain system that allows for temporary storage and manipulation of information. According to Baddeley and Hitch's (1974) influential model, WM is comprised of a central executive and two subsidiary systems, the phonological loop and visuospatial sketchpad. The central executive is responsible for directing attention and coordinating the activities of the subsystems. The phonological loop and visuospatial sketchpad are systems used to store and maintain verbal, visual and spatial information. Higher WM capacity means greater resources to put towards attention- and cognitively-demanding tasks (Just & Carpenter, 1992). WM may enable speakers to mentally compare to-be-referred-to objects and surrounding objects in order to design and produce an appropriate reference.

Previous research suggests that the human attentional system can be divided into three separable functional and neuroanatomical regions, including alerting, orienting and EC (cf. Fan et al, 2005). EC refers to processes that modify behavior such as strategy selection, conflict resolution, error correction and inhibitory control (Reder & Schunn, 1996). Measures of EC typically require subjects to inhibit a prepotent response in order to react to a stimulus. The current proposal is that EC may enable speakers to limit the selection of conceptual features included in a particular referential expression to those that are intended, rather than to those that are highly activated. Additionally, EC and WM were chosen for investigation because past research suggests a role for these executive functions in listeners' perspective taking. If these executive functions regulate speakers' perspective taking as well, this would suggest that the perspective-taking process is controlled by domain-general mechanisms irrespective of performance modality.

Recent work on *listeners'* perspective taking suggests domain generality. Brown-Schmidt (2009) and Nilsen and Graham (2009) both showed that listeners' verbal inhibitory control performance predicted their perspective-taking performance. Brown-Schmidt measured listeners' eye movements during a referential communication task and had participants complete a Stroop-like task (Stroop, 1935). Participants' performance on the Stroop task significantly predicted their ability to use perspective to guide their reference resolution. Similarly, Nilsen and Graham (2009) gave 4- and 5-year-old children an adapted Stroop task and a tapping task (Luria, 1966; & Diamond & Taylor, 1996) to measure inhibitory control, a backwards digit span and memory for objects task to test WM and the Flexible Item Selection Task (Jacques & Zelazo, 2001) to measure cognitive flexibility. Children also completed referential communication tasks as speakers and as listeners. Their findings indicated that children's reference resolution, but not reference production, was related to their performance on inhibitory control measures, but not to measures of WM or cognitive flexibility.¹

Research has also shown WM is predictive of adult listeners' perspective taking. Lin, Keysar and Epley (2010) measured listeners' eye movements as they participated in a referential communication task and had participants perform a task of WM. Results indicated that WM performance predicted the degree to which listeners used perspective to guide their reference resolution. This finding suggests that at least some part of the perspective-taking process for listeners is regulated by WM.

The current study uses individual differences in executive functioning to assess the domain-generality of speakers' perspective taking. Speakers were run on a referential communication task and measures of WM and EC. If perspective taking in production is regulated by these mechanisms that more generally regulate human behavior, as has been shown in language comprehension, performance on the executive function measures should predict performance on the referential communication task.

2.0 Method

2.1 Participants

Sixty undergraduates from the University of California, San Diego participated. Two participants were eliminated due to failure to follow instructions.

2.2 Perspective-taking Task

On each trial, participants viewed four pictures of objects on $8\frac{1}{2} \times 11$ inch paper. Three objects were mutually visible and one was visible to the speaker only. Speakers named a particular target object for the listener so that the listener could pick it out of the display. Listeners were told that the speaker would name a hidden object on some trials, so if the speaker named an object that the listener did not see, the listener could guess that the target was the hidden object. Objects varied in size such that the actual size of a large (or small) object (relative to the size of the other objects in that set) on one trial might be small (or large) on another trial. Critical targets were medium-sized and in common ground. The task consisted of 16 filler trials, 10 baseline trials and 10 privileged ground trials. On baseline trials, targets were unique to the set. On privileged ground trials, targets had size-contrasting pair mates in privileged ground. Each critical target was presented in both the baseline and privileged ground conditions across participants. Four filler trials required participants to name a privileged object: three of these were singlets and one had a pair mate in common ground. Three filler trials required participants to name a common ground singlet and nine required participants to name a common ground target that had a size-contrasting common ground pair mate.

Each shape was used on only one trial per subject, and never occurred with more than one other object of the same shape. Speakers' use of size-contrasting modifiers was measured: more modifiers on privileged ground trials than on baseline trials serves as evidence of failure to account for perspective differences.

To begin, an experimenter read instructions and then administered two practice trials. After 36 trials, participants were administered the digit span and flanker tasks and then run on a second round of the perspective-taking task. For the second round, participants switched speaker/addressee roles and were instructed that although they would be viewing the same shapes as in the previous round, the shapes were paired differently so using one's memory of pairings from the initial round would not be helpful.

At the beginning of each trial, addressees closed their eyes and speakers turned to a page in a stimulus binder, revealing a set of four objects. Speakers looked at those objects for two seconds, after which the experimenter pointed to one of the objects. Speakers then blocked that object using an occluder, making the object privileged. After another two-second delay, the experimenter indicated a target by pointing.

2.3 Digit Span Task

To measure WM, participants completed the forward and backward subtests of the WAIS-III (Wechsler, 1997). Forward and backward scores were combined to create one digit span score. The forward span task measures participants' ability to recall increasingly longer number strings in the same order as presented. This task measures short-term auditory memory and language-related attention resources (Hale, Hoepfner & Fiorello, 2002, Lezak, 1983). The backward span task measures participants' ability to recall number strings in the reversed order from presentation. This task measures participants' ability to store, maintain and manipulate information. Because of the linguistic component of these tasks, they likely tap into an articulatory loop to a greater degree than a visuo-spatial sketchpad. However,

research supports the view that WM capacity is a unitary construct across both verbal and visuospatial span tasks (Kane et al, 2004).

2.4 Flanker Task

Participants completed a flanker task to measure EC (Fan, et al, 2002). In this task, participants respond to the direction of a center arrow, which is flanked by horizontal lines or arrows. Participants press a key with their left hand if an arrow points left, or with their right hand if an arrow points right. Targets were 32 congruent (five arrows pointing in the same direction), 32 neutral (a single arrow flanked by lines without arrow-heads), and 32 incongruent (two arrows on each side pointing in the opposite direction of the center arrow) displays. Displays were evenly divided between left- and right-pointing center arrows and randomly presented. Trials included a 500 ms central fixation point, followed by the target stimulus. The inter-trial interval was 500 ms. Practice blocks began with six neutral trials, followed by six congruent trials, then six incongruent trials, and then six trials with equal numbers of the different conditions in random order.

Performance on the flanker task measures participants' ability to avoid inappropriate action and select appropriate action both of which are implicated in executive control (cf., Kopp, Rist & Mattler, 1996).

3.0 Results

3.1 Perspective-taking task

The dependent variable, computed for each subject by contrast-type (privileged ground versus baseline trials), was the percentage of targets speakers described with direction-appropriate modifiers. Speakers modified more often when targets had a size-contrasting match in privileged ground (29.0%) as compared to when targets had no match in the set (1.4%), $t(1,57) = 5.9$, $p < .001$. Thus, speakers produced non-perspective adjusted references on 29% of privileged ground trials.

3.2 Digit Span Task

Mean forward digit span score was 10.9 (SD 1.9). Mean backward digit span score was 7.8 (SD 2.1). Mean combined digit span score was 18.8 (SD 4.4). Combined digit span scores significantly and negatively correlated with modifier use on privileged ground trials ($r = -0.47$, $p < .005$)². The lower speakers' scores on the digit span tasks, the more often they produced non-perspective adjusted references.

3.3 Flanker Task

Table 1 displays mean RTs and error rates for each congruency condition. Participants were significantly faster on congruent than on neutral or incongruent trials (both $ps < .01$). RTs on incongruent trials, and flanker interference scores (incongruent minus neutral RTs), were significantly and positively correlated with modifier use on privileged ground trials, such that speakers with faster RTs on incongruent trials and smaller interference scores used size-contrasting modifiers less often on privileged ground trials ($r = 0.42$, $p < .01$; $r = 0.33$, $p < .01$, respectively).

4.0 Discussion

Participants' performance on perspective taking, WM and EC tasks was measured. Of interest was whether domain-general executive function measures were predictive of speakers' perspective-taking behavior. Speakers regularly made errors in perspective taking by including information in their references that contrasted the referent from an object that

was unknown to their listener. Furthermore, this performance was predicted by performance on WM and EC tasks suggesting a role for domain-general mechanisms in speakers' perspective taking.

How might WM be relevant to speakers' perspective taking? Figure 2 illustrates a possible mechanism. Scanning a display to identify each object requires that the speaker store the identity of each of the other relevant objects in WM. Storing these objects allows the speaker to compare them and enables the relevant distinguishing features to become activated at appropriate levels. Imagine a real-world situation, such as referring to a particular car in a crowded parking lot. In this situation, there would likely be many other cars from which to distinguish the target, and the target will differ from other cars in more than one manner (colors, shapes and sizes). To refer to one particular car, the speaker must determine which conceptual features are necessary to distinguish the target from the other mutually visible cars. WM is a resource that speakers could use to hold in memory the features of the target and compare those to the relevant surrounding objects, which are potential targets to a listener. This is not dissimilar to the processes necessary for performance on the digit span tasks where participants must hold the identity of numbers in memory while comparing and manipulating their order.

How might EC be relevant to perspective taking in this type of task? The level of activation of conceptual features relating to differentiating features of common and privileged ground objects must be activated and maintained appropriately and dynamically as relevant contextual conditions change. Using Figure 4, the concepts for “big” and “small” related to the triangles are relatively highly activated from Time Point 2 onward. This is necessary in case the speaker must identify one of those triangles in the context of both being visible to the listener. In this situation, there is no prediction of a role for EC. However, if one of the triangles becomes occluded, then the size contrasts should lose activation because the two triangles need not be distinguished from one another. In this situation, a role for EC in the management/inhibition of activation and selection of size-differentiating features is proposed, similarly to the management and inhibition of prepotent responses as measured in the flanker task. Outside of the laboratory, situational features that serve to increase the activation of the now non-relevant size information, or to delay its loss of activation, should also serve to make it more likely that that non-relevant size relevant information is included in the speakers' utterance.

Using the example of a parking lot, a speaker must be able to make reference to a particular car without selecting and producing inappropriate conceptual features, even if they are highly activated. Why might particular conceptual features be highly activated even when speakers do not intend to include those features in their reference? Drawing a speaker's attention to particular item features could cause the concepts associated with those features to become highly activated. Past research suggests that when this occurs, speakers are more likely to include those conceptual features in their references, even when doing so is counterproductive to their own goals (Wardlow Lane et al, 2006; Wardlow Lane & Liersch, 2012). Outside of the confines of a laboratory, there are countless ways speakers' attention can be drawn to conceptual features of a to-be-referred-to object as well as to surrounding objects. For example, a particular car might grab a speaker's attention because it is unusual (rare, odd, driving by quickly). The current proposal is that the extent to which a specific speaker can avoid producing a reference that is altered by that attention-grabbing should be related to that speaker's EC capacity.

Two prominent theories propose explanations for how interlocutors take others' perspectives into account. The models make different predictions about how executive control could influence perspective taking. First, the constraint satisfaction model proposes that the

distinction between common and privileged knowledge is one of many contextual constraints speakers use and therefore its effect is only partial (Hanna, Tanenhaus & Trueswell, 2003; Nadig & Sedivy, 2002). These constraints are weighted according to their salience and reliability and their effects are related to the probability that each cue will be important for communication (Hanna, Tanenhaus & Trueswell, 2003).

In contrast, the dual process model proposes that interlocutors take perspective into account through a two-stage process. The first stage is a quick and automatic stage during which interlocutors begin formulating their message or interpreting a message using only their own perspective. The second stage is an effortful and controlled stage during which the production and comprehension systems monitor for perspective. Given adequate resources, the system(s) will adjust to correct for the original egocentric perspective Keysar, Barr, Baliln & Paek, 1998; Keysar, Lin & Barr, 2003).

A contrast satisfaction approach might predict that EC mechanisms reduce or “shut off” the size feature representation when inclusion of that representation violates perspective differences. In contrast, a dual process approach might predict that EC mechanisms influence perspective taking via a monitoring process, recognizing at some point in the production process that a formulation with the size feature representation violates perspective differences. A third possibility is that EC mechanisms are able to affect perspective taking in both of these ways: through both a reduction of activation of the size feature representation and through a later monitoring process. Further specification of process models of speakers' perspective taking should take into account domain-general mechanisms that influence individual speakers' abilities to use perspective to guide their referential behaviors.

This proposal does not assume that inclusion of a modifier is a prepotent response, only that the concepts related to the modifier will sometimes be highly activated due to the context within which a reference is made, and that that in turn will influence a particular speakers' tendency to include that information as a function of their EC capacity. In this way, the current proposal differs from proposals made regarding listeners' inhibitory control and perspective taking. Those proposals suggested that listeners inhibit their own perspective (e.g., prepotent response) to correctly identify a referential target (Brown-Schmidt, 2009; Nilsen & Graham, 2009). The current data do not speak specifically to whether speakers must inhibit a prepotent response in the form of their own perspective before composing their utterance. These results reveal that domain-general mechanisms exert influence over speakers' perspective taking.

References

- Baddeley AD, Hitch GJ. Working memory. *The psychology of learning and motivation*. 1974; 8:47–89.
- Bock JK. Toward a cognitive psychology of syntax: Information processing contributions to sentence formulation. *Psychological Review*. 1982; 89:1–47.
- Brennan SE, Hanna JE. Partner-Specific Adaptation in Dialog. *Topics in Cognitive Science*. 2009; 1(2):274–291.
- Brown-Schmidt S. The role of executive function in perspective taking during online language comprehension. *Psychonomic Bulletin & Review*. 2009; 16:893–900. [PubMed: 19815795]
- Fan J, McCandliss BD, Sommer T, Raz A, Posner MI. Testing the efficiency and independence of attentional networks. *Journal of Cognitive Neuroscience*. 2002; 14:340–347. [PubMed: 11970796]
- Fan J, McCandliss BD, Fossella J, Flombaum JI, Posner MI. The activation of attentional networks. *NeuroImage*. 2005; 26(2):471–479. [PubMed: 15907304]
- Garrett, MF. *Sentence processing*. The MIT Press; Cambridge, MA, US: 1990.

- Hale JB, Hoeppepner & Fiorello CA. Analyzing digit span components for assessment of attention processes. *Journal of Psychoeducational Assessment*. 2002; 20(2):128–143.
- Hanna JE, Tanenhaus MK, Trueswell JC. The effects of common ground and perspective on domains of referential interpretation. *Journal of Memory & Language*. 2003; 49(1):43–61.
- Horton WS, Keysar B. When do speakers take into account common ground? *Cognition*. 1996; 59(1): 91–117. [PubMed: 8857472]
- Jacques S, Zelazo PD. The Flexible Item Selection Task (FIST): A measure of executive function in preschoolers. *Developmental Neuropsychology*. 2001; 20(3):573–591. [PubMed: 12002094]
- Just MA, Carpenter PA. A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*. 1992; 99:122–149. [PubMed: 1546114]
- Keysar B, Barr DJ, Balin JA, Paek TS. Definite reference and mutual knowledge: Process models of common ground and comprehension. *Journal of Memory and Language*. 1998; 39(1):1–20.
- Keysar B, Barr DJ, Horton WS. The egocentric basis of language use: Insights from a processing approach. *Current directions in psychological science*. 1998; 7(2):46–50.
- Keysar B, Lin S, Barr DJ. Limits on theory of mind use in adults. *Cognition*. 2003; 89(1):25–41. [PubMed: 12893123]
- Kopp B, Rist F, Mattler U. N200 in the flanker task as a neurobehavioral tool for investigating executive control. *Psychophysiology*. 2007; 33(3):282–294. [PubMed: 8936397]
- Krauss RM, Fussell SR. Perspective-taking in communication: Representations of others' knowledge in reference. *Social Cognition*. 1991; 9(1):2–24.
- Lin S, Keysar B, Epley N. Reflexively mindblind: Using theory of mind to interpret behavior requires effortful attention. *Journal of Experimental Social Psychology*. 2010; 46(3):551–556.
- Levelt, WJM. *Speaking: From intention to articulation*. MIT Press; Cambridge, MA: 1989.
- Lezak, M. *Neuropsychological assessment*. 2nd ed.. Oxford University Press; New York: 1983.
- Luria, AR. *Higher cortical functions in man*. Basic Books; New York: 1966.
- Miyake A, Friedman NP, Emerson MJ, Witzki AH, Howerter A, Wager TD. The unity and diversity of executive functions and their contributions to complex “frontal lobe” tasks: A latency variable analysis. *Cognitive Psychology*. 2000; 41(1):49–100. [PubMed: 10945922]
- Nadig AS, Sedivy JC. Evidence of perspective-taking constraints in children's on-line reference resolution. *Psychological Science*. 2002; 13(4):329–336. [PubMed: 12137135]
- Nilsen E, Graham S. The relations between children's communicative perspective-taking and executive functioning. *Cognitive Psychology*. 2009; 58:220–249. [PubMed: 18809176]
- Pennington BF, Ozonoff S. Executive functions and developmental psychopathology. *Journal of Child Psychology and Psychiatry and Allied Disciplines*. 1996; 37(1):51–87.
- Reder, LM.; Schunn, CD. Metacognition does not imply awareness: Strategy choice is governed by implicit learning and memory. In: Reder, LM., editor. *Implicit Memory and Metacognition*. Erlbaum; Mahwah, NJ: 1996. p. 45-77.
- Stroop JR. Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*. 1935; 18:643–662.
- Wardlow Lane L, Groisman M, Ferreira VS. Don't talk about pink elephants: Speakers' control over leaking private information during language production. *Psychological Science*. 2006; 17:273–277. [PubMed: 16623681]
- Wardlow Lane L, Liersch MJ. Can you keep a secret? Increasing speakers' motivation to keep information confidential yields poorer outcomes. *Language and Cognitive Processes*. 2012; 27(3): 462–473.
- Wardlow Lane L, Ferreira VS. Speaker-external versus speaker-internal forces on utterance form: Do cognitive demands override threats to referential success? *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 2008; 34(6):1466–1481.
- Wechsler, D. *Wechsler Adult Intelligence Scale*. Third Edition. The Psychological Corporation; San Antonio, TX: 1997.

Figure 1.
Example Experimental Display.

Figure 2.
Task Analysis

Table 1

Flanker Task and Digit Span Means and SDs

		Mean	Standard Deviation
Flanker Task	Congruent RT	495 ms	91 ms
	Incongruent RT	571 ms	147 ms
	Neutral RT	512 ms	105 ms
	Incongruent – Neutral RT	59 ms	71 ms
	Congruent Error Rate	0.4%	1.6%
	Incongruent Error Rate	2.0%	4.2%
	Neutral Error Rate	0.6%	1.9%
Digit Span Task	Forward Digit Span Score	11	2
	Backward Digit Span Score	8	2
	Combined Digit Span	18.5	4.4

Table 2Correlation Matrix *Pearson bivariate correlations between task scores*

	%Modifiers	Digit Span	Flanker Incongruent RT
Digit Span Combined	−0.37 [*]	-	
Flanker Incongruent RT	0.42 [*]	−0.1	-
Flanker Difference Score	0.33 [*]	−0.04	0.76 [*]

^{*}
p < .05