Speakers' use of agency and visual context in spatial descriptions

SPEAKERS' USE OF AGENCY AND VISUAL CONTEXT IN SPATIAL DESCRIPTIONS

Alessia Tosi[1], Martin J. Pickering[1], Holly P. Branigan[1]

[1] Department of Psychology, University of Edinburgh

Corresponding author:

Alessia Tosi

Department of Psychology, University of Edinburgh, 7 George Square, Edinburgh EH8 9LW, U.K.

email: alessia.tosi@gmail.com

phone: +44 787 3167109

Word Count: 10523

Keywords: Perspective taking; Spatial language production; Self; Other; Agency

Abstract

When speakers describe the world, they typically do so from their own perspective. However, they are able to adopt a different perspective, and sometimes do so even when they are not communicating with someone who has a different perspective from their own. In three experiments, we investigated the factors that might lead speakers to adopt a non-self perspective. Participants described how objects were located in scenes that contained no other entity, a person, or a symmetrical plant. They were more likely to adopt a non-self perspective (e.g., using *to the left* to refer to an object on their own right) if the scene contained a person facing them than a person facing away or a plant, and if it contained a person who could see (and potentially act on) the object than a person who could not, even when that person showed no intention to act. Our results suggest that speakers can put themselves in the shoes of another potential agent and use a simulation of that agent's perspective as the basis for formulating their descriptions.

1 Introduction

You are writing a new post for your blog about tourist frenzy, using your last holiday in Italy as an excuse. You have uploaded the picture you took of the Leaning Tower of Pisa and describe how, while you were carefully framing the photo to make it look as though a little rosebush on the left was holding the tower up, a tourist suddenly appeared on the right pretending to push the tower over. Language allows us to describe space in a variety of ways, explicitly or implicitly locating things in relation to ourselves (e.g., *on my left*), to another object (e.g., *to the left of the tower*), or to fixed environmental features (e.g., *to the north*; Avraamides, Mello, & Greenauer, 2013; Levinson, 1996; Tversky, 1996). Speakers can always describe a scene from their *self-perspective.* In our example, you might describe the rosebush as being *on the left* because it is on the left with respect to your own left and right sides.[i] And of course, this self-perspective would also correspond to the perspective of your audience as they view the photo on your blog. But speakers sometime describe locations from a different, *non-self perspective*. For instance, you might sometimes describe the rosebush as being *on the right,* even though this does not correspond to your own or your audience's perspective. Why might people do so? In this paper, we investigate whether speakers might sometimes adopt a non-self perspective because they are influenced by features of the scene that they are describing, and explore whether such features critically involve another agent's viewpoint or alternatively involve salient reference points.

Models of language production assume that when speakers produce an utterance, they first determine the message that they wish to convey (i.e., form a particular communicative intention; *conceptualisation*), and then map this message onto appropriate linguistic representations (*formulation;* Levelt, 1989). The same state of affairs can be conceived in many different ways (i.e., is compatible with many messages), and so speakers must choose a particular way of viewing that state of affairs. In other words, they must think in a way that can serve as a basis for speaking (i.e., 'thinking for speaking'; Slobin, 1987). For example, the

same arrangement of objects in space can be conceived from many different viewpoints. In our example, the rosebush and the leaning tower can be conceptualised from you and your audience's viewpoint (in which the rosebush is to the left of the tower), or from the tourist's viewpoint (in which the rosebush is behind the tower), or from the tower's viewpoint (in which case the rosebush is in front of it).

This means that in order to produce a spatial description, speakers must choose a particular spatial perspective from which to describe the position of one entity relative to another. That is, they must decide a point of reference or *origin* (e.g., the tourist) and a way of locating objects with respect to that origin, or *reference frame* (e.g., in terms of the tourist's intrinsic axes, such as her vertical axis with her head being above her feet; Carlson, 1999). The origin and reference frame together determine how spatial terms (e.g., *above, below, front, behind, left*, and *right*) are interpreted (see also Carlson-Radvansky & Logan, 1997; Carlson-Radvansky & Irwin, 1994; Herskovits, 1986; Levelt, 1984; Talmy, 1983). The ensuing message, specifying which elements are spatially related to each other, and how they are related, then forms the input to subsequent processes of linguistic formulation (e.g., lexical retrieval, grammatical encoding).

How do speakers choose a particular perspective? They can always use themselves as the origin (i.e., adopt a self-perspective), and in fact speakers are assumed to use a self-perspective as a default (Levelt, 1989; Pick & Lockman, 1981).[ii] But it is uncontroversial that speakers can take a non-self perspective if explicitly asked to do so (Bryant, Tversky, & Franklin, 1992). Many studies have shown that speakers are also likely to spontaneously take a non-self perspective when producing a spatial description in interactive contexts that involve a joint spatial task (e.g., Mainwaring et al. 2003; Schober, 2009; see also Galati & Avraamides, 2013). In these cases, the decision to take a non-self perspective appears to be related to speakers' beliefs about their addressees' knowledge, understanding, or interests. In other words, speakers' choice of which spatial perspective to adopt is probabilistically

influenced by *audience design* (Clark & Murphy, 1982), in the same way as - for example their - decisions about referential expressions or syntactic structure (Brennan & Clark, 1996; Haywood, Pickering, & Branigan, 2005).

For instance, when speakers tell addressees where objects are located, they often produce descriptions relative to their addressees' spatial perspective in order to help them reach the correct interpretation (Levelt, 1984; see also Bürkle, Nirmaier, & Herrmann, 1986; Herrmann, Bürkle, & Nirmaier, 1987). This tendency is stronger if the addressee is not co-present and so cannot negotiate a perspective or give feedback on understanding (Mainwaring, Tversky, Ohgishi, & Schiano, 2003; Schober, 1993; see also Tenbrink, Maiseyenka, & Moratz, 2007), and weaker when the spatial descriptions are irrelevant to the addressees' task (Roche, Dale, & Kreuz, 2010; Yoon, Koh, & Brown-Schmidt, 2012).

Speakers' use of non-self perspectives is also influenced by their estimation of their addressees' abilities to process and reconstruct the speaker's perspective: Speakers with high spatial abilities are more likely to provide descriptions from their partner's perspective if they perceive their partner to be less competent in spatial cognition, or to be having difficulties carrying out the task successfully (Schober, 2009). These findings demonstrate that speakers sometimes choose a non-self perspective when addressing someone who has an alternative perspective and who is involved in a joint spatial task (whether the addressee is co-present or not). Such audience design effects may involve quite explicit reasoning about addressees' likely understanding and explicit simulation of their self-perspective.

But other factors may also affect choice of perspective. We know that speakers' choices of referring expressions and syntactic structure do not depend entirely on considerations of audience design, but also depend probabilistically on factors such as visual salience to the speaker (Goudbeek & Krahmer, 2012) or ease of processing for the speaker (Bock, 1986; Ferreira & Dell, 2000); such factors may influence speakers' decisions alongside considerations of audience design (Haywood et al., 2005). In the same way, recent studies

suggest that audience design is not the only determinant of speakers' choice of spatial perspective.

For example, Tversky and Hard (2009) found that that speakers were more likely to adopt a non-self perspective when they described the location of objects in a scene that contained an agent with a different spatial perspective than when they described the location of the same objects in a scene that did not contain such an agent. These effects occurred even though participants were not interacting with, or had any communicative intent towards, that agent. Thus when participants viewed a photographed scene and answered a question such as *In relation to the bottle, where is the book?* for an unspecified audience, they took a non-self perspective about a quarter of the time if the scene contained a person who had an opposite spatial perspective on the object (i.e., that person's self-perspective that did not match the participant's self-perspective) and who was either looking at or reaching for the object (compared to less than 5% of the time if the scene did not contain a person).

Our concern in this paper is with such cases, where the adoption of a non-self perspective does not appear to have a basis in audience design: What features of the scene could have caused Tversky and Hard's (2009) participants to adopt a non-self perspective? More generally, what factors other than audience design can affect speakers' choice of perspective during the production of spatial descriptions? We contrast two types of account. According to an *intentional account*, speakers may take another (non-self) perspective when they detect the alternative viewpoint of another agent. In one version of this account, speakers are sensitive to the viewpoint of any agent that might have intentions (*perceived intentional* version); under another version, they are sensitive to the viewpoint of another agent that is currently acting or showing an intention to act (*actional intentional* version). But according to a *non-intentional account*, speakers' choice of spatial perspective may be affected by featural or functional aspects of the environment (e.g., a salient environmental feature that attracts speakers' attention away from themselves and that provides a spatial reference point).

Consistent with an intentional account, some evidence suggests that speakers' tendency to adopt a non-self perspective is linked to a need to understand other agents' intentions or actions, even when the agent displaying these intentions and actions is not the addressee of the speaker's utterance. In Tversky and Hard (2009), participants' tendency to adopt a non-self perspective for scenes involving another person was enhanced if the question referred to an action (e.g., *In relation to the bottle, where is the book placed?*) than if it just referred to the location (*In relation to the bottle, where is the book?*). Other studies using similar paradigms showed that the tendency to take a non-self perspective was modulated by how explicitly the observed person was acting on the object (i.e., gazing vs. grasping; Mazzarella, Hamilton, Trojano, Matromauro & Conson, 2012). For instance, Furlanetto, Cavallo, Manera, Tversky, and Becchio (2013) presented participants with videos of a person who displayed a range from less to more explicit intentions to act (i.e., neutral, gazing, grasping). Participants' tendency to produce non-self perspective descriptions increased as the agent's discernible intentionality increased.

However, people are not always sensitive to the presence of another agent. Surtees, Noordzij, and Apperly (2012) found that effects of agency depended on the context in which participants were exposed to scenes that included potential agents. In their study, participants judged the appropriateness of statements that described the position of a ball with respect to an agent-like (doll) or a non-agent-like (chair) referent. Some statements accurately described the position of the ball only if the participants considered their own perspective but not the referent's perspective (i.e., they were appropriate only under the self-perspective); other statements accurately described the position of the ball only if the participants took the referent's perspective but not their own (i.e., they were accurate only under the non-self perspective). When participants experienced both agent-like and non-agent-like referents in the same block of stimuli, they showed an effect of agency: They gave higher ratings for statements that were appropriate under the non-self perspective than to statements that were

inappropriate under the non-self perspective but appropriate under the self-perspective (thus displaying sensitivity to the non-self perspective), and this tendency was stronger when the referent was agent-like. But importantly, when participants experienced the two types of referent in separate blocks, the effect of agency disappeared (see also Clements-Stephens, Vasiljevic, Murray, & Shelton, 2013).

In summary, some evidence suggests that speakers may implicitly adopt a non-self perspective as a result of perceiving the spatial perspective of another agent who demonstrates the potential for action, is currently engaged in action, or who evinces an intention to act. These results are compatible with an intentional account. To account for them, researchers have suggested that people engage in mental simulation when they detect another agent, engaging in a shift from themselves, the observer, to the observed person and simulating what the scene would look like from that position (e.g., Johnson & Demiris, 2005).

However, other evidence supports a non-intentional account. For instance, some studies show that speakers' use of a non-self perspective in their spatial descriptions is influenced by functional relationships between objects, even in the absence of an alternative viewpoint or any evidence of agency. Carlson-Radvansky and Radvansky (1996) found that when people were presented with isolated pairs of functionally-related objects (e.g., hammer and nail), they tended to use a non-self perspective that was based on the directional features of one object to describe the location of the other object (e.g., "the nail is in front of the hammer" rather than "the nail is to the right of the hammer"; see also Johannsen & De Ruiter, 2013a).

Studies using non-linguistic spatial perspective-taking tasks also support a non-intentional account. These studies suggest that the properties of inanimate entities can affect how people adopt a spatial perspective. For example, participants showed interference when locating dots on a computer display if there was a conflict between their perspective and the 'perspective' of a moving symmetrical triangle to which they had attributed left/right sides

(based on its directionality of motion; Zwickel, 2009).

More critically, findings from non-linguistic tasks suggest that attention-orienting mechanisms can affect spatial perspective-taking independently of perceptions of agency, in keeping with Heyes' (2014) proposal that many entities have attention-eliciting features that can lead people to code spatial responses according to their intrinsic axes; for example, people's attention may be drawn to an entity's front by its eyes or mouth (in the case of a person) or its control panel (in the case of a machine), so that they tend to use this axis in a subsequent spatial response. Santiesteban, Catmur, Hopkins, Bird, and Heyes (2014) had participants count the number of dots on a screen. On some trials, a human-like avatar or a pointing arrow was present on the screen. Participants' response latency and accuracy were affected to the same extent when the arrow's directional features (i.e., the arrow's tip) pointed to a subset of the dots as when the avatar's directional features (i.e., the avatar's head) did so, suggesting no effect of agency. Although these results do not involve language, they challenge the intentional account – in other words, that it is the perception of another agent's viewpoint that can trigger speakers to adopt a non-self perspective.

In sum, spatial language use, and specifically speakers' choice of perspective, does not depend solely on consideration of an addressees' immediate comprehension needs: Speakers sometimes adopt a non-self perspective when producing spatial descriptions of scenes that contain another person who is not the addressee. But findings are contradictory about the factors that trigger speakers to adopt a non-self perspective. Some findings support an intentional account that involves perceptions of agency, whereas other findings (largely from non-linguistic spatial tasks) support a non-intentional account that involves the presence of salient reference points.

The current study

Our study set out to examine what might cause people to adopt a non-self perspective when producing spatial descriptions, other than attempting to meet the needs of their addressees.

Specifically, we investigated what is it about an entity that affects whether speakers describe its location using a non-self perspective: Are speakers influenced by the presence of an entity with an alternative viewpoint and by the actual or potential agency of that entity? Or are speakers influenced instead by whether an entity has a good source of spatial coordinates?

Building on the task used by Tversky and Hard (2009), we had participants describe the spatial relation between two objects (e.g., a candle and a pineapple) in a photographed scene in response to the question *On which side of the X is the Y?* (see Figure 1). The objects were symmetrical and had no functional relationship to each other. We manipulated whether the scene included a non-agentive entity (e.g., a plant), an agentive entity (i.e., a person) that was not acting and showed no intention to act, or no additional entity. We also manipulated where that entity was positioned with respect to the two objects, so that the agentive entity either could or could not perceive the objects. We measured the effect of these manipulations on whether participants used a self-perspective in their spatial descriptions (e.g., *The candle is to the left of the pineapple* to describe a scene such as Figure 1), or whether they instead adopted a non-self perspective (e.g., *The candle is to the right of the pineapple*).

In our task, the questions referred only to the two objects on the table. Hence our task did not ask participants to take into account the spatial relation between the objects and the additional entity. This is important as previous studies displaying only one object alongside the additional key entity (e.g., a person) may have induced participants to base their spatial descriptions on this additional entity, making the task inadvertently biased towards a non-self perspective (Mazzarella et al., 2012). As in many previous studies (e.g., Carlson-Radvansky & Radvansky, 1996; Tversky & Hard, 2009), the task required participants to produce descriptions for an unspecified audience; participants had no reason to believe that the photographed person who was present in the scene should be considered the addressee of these descriptions (i.e., the person who needed to understand them).

In Experiment 1, we tested whether participants produced more non-self

perspectives when describing the location of one object relative to another object in a scene containing a human who was in a neutral pose and whose orientation differed from the participant's orientation (agentive entity condition) than a scene that contained a plant (non-agentive entity condition) or no additional entity (no entity condition). We further tested whether any such effect was affected by whether the human could see and potentially act on the object.

In Experiment 2, we manipulated the order of presentation of the three entity conditions (agentive vs. non-agentive vs. no entity) using a blocked design, in order to test whether non-self perspective-taking was influenced by the experimental context. Specifically, we examined whether scenes involving a non-agentive entity could always induce a non-self perspective, or whether participants adopted a non-self perspective in non-agentive trials only if they had previously used that perspective in agentive trials.

In Experiment 3, we manipulated the orientation of the human so that it faced in the same or the opposite direction to the participant, to examine whether non-self perspective-taking was affected by another agent's perspective when this perspective corresponded to the participant's own perspective as well as when it constituted an alternative perspective. Differences in patterns of non-self perspective descriptions between conditions would be informative about the extent to which speakers producing spatial descriptions are triggered to take a non-self perspective by the detection of an agent that might have intentions (i.e., consistent with the perceived version of the intentional account), the detection of an agent evincing a current action or intention to act (i.e., consistent with the actional version of the intentional account), or an entity with salient directional intrinsic features (i.e., consistent with the non-intentional account).

2 Experiment 1

In this experiment, participants produced spatial descriptions for scenes containing an agentive entity, a non-agentive entity, or no additional entity (ENTITY manipulation). The

agentive entity was a human in a stationary pose. The human was facing forward and had crossed arms; he or she was not engaged in action and did not evince any clear intention to act (e.g., reaching for or looking at an object). This pose also ensured that the human displayed no attention-orienting cues caused by specific asymmetric postures (e.g., head turned, one arm stretched out). The non-agentive entity was a plant, which was closely matched in size to the human and had no directional features. We assume that people do not regard this plant (or, most likely, plants in general) as intentional or as having a perspective or distinct front/back or left/right sides.

We also manipulated whether the agentive or non-agentive entity appeared in front of or behind the objects (LOCATION manipulation). Crucially, the agentive entity always faced the participant so that its orientation and viewpoint were the opposite of the participant's (see Figure 1). As a consequence, the entity could potentially see and act on the objects (i.e., participants could perceive a non-self relative perspective) only in the Agent-Behind condition (Figure 1d).

Under the perceived intentional account, participants should be more likely to produce a non-self description (e.g., *the candle is to the right of the pineapple*) when they perceive the entity to be intentional and to have a perspective corresponding to a non-self relative perspective; moreover this tendency should not depend on the entity carrying out an observable action or showing intention to act. This should be the case in the Agent-Behind condition (Figure 1d). It should not be the case in the Agent-Front condition (Figure 1c) because the entity would not be able to see or act on the objects (i.e., the entity does not provide an alternative relative perspective). More obviously, it should not occur in either of the non-agent conditions; in the Non-agent-Behind condition, even if participants were to simulate themselves in the position of the plant, there is no reason why they should assume it to be facing them.

Under the actional intentional account, participants' tendency to produce non-self

descriptions should not be affected by the entity (agent vs. non-agent) or its position (agent-front or agent-behind), because the agent (and, obviously, the non-agent) displayed no action or intention to act. Participants should therefore be equally likely to produce non-self descriptions in all conditions.

Under the non-intentional account, in contrast, both Agent conditions should trigger more non-self perspectives than the Non-agent conditions and the No-entity condition, because people have salient directional intrinsic features that can be used as a basis for coding a spatial response, whereas plants do not. In addition, there should be no difference between the Agent-Behind and the Agent-Front conditions because the directional features of the human body do not change between the two conditions.

2.1 Method

2.1.1 Participants

We recruited 144 participants on CrowdFlower, a crowdsourcing self-service platform that allows clients to design and execute their own workflow (and allows full anonymity). All participants were English native speakers and were paid $0.40.

2.1.2 Materials

We used twelve experimental items. Each item involved a pair of experimental objects that had no clear front/behind or left/right (e.g., candle/pineapple). We constructed two sets of photographed scenes for each item, one set for each possible position of the two objects on the table (i.e., candle on the right and pineapple on the left vs. candle on the left and pineapple on the right). Each set comprised five photographs, corresponding to the four combinations of ENTITY (Agent vs. Non-agent) and LOCATION (Front vs. Behind), and the NO ENTITY control. See Figure 1. In the ENTITY conditions, we used four plants and four people (two female, two male). Each scene was paired with a question of the form *On which side of X is Y?* (e.g., *On which side of the candle is the pineapple?*). Across the item set, each object appeared on the left and right an equal number of times, and was asked about an equal number of times;

each person/plant appeared front/behind an equal number of times.

We also constructed 16 filler items, each displaying two further objects on a table. In half the fillers, objects were different colours (e.g., a red baking tin and a green salad bowl) and participants were asked *What colour is the X?* (e.g., *What colour is the salad bowl?*) In the other half, objects were of different size (e.g., a tube of toothpaste and a sewing machine) and participants were asked *With respect to the X, how big is the Y?* (e.g., *With respect to the toothpaste, how big is the sewing machine?*)

We created 12 lists, each comprising 24 experimental and 16 filler trials. Each list comprised eight ENTITY/Agent trials (four Front, four Behind), 8 ENTITY/Non-agent trials (four Front, four Behind) and eight NO ENTITY trials. In each list, each of the twelve experimental items was assigned to two of the conditions, in such a way that it never appeared twice in the same ENTITY/LOCATION condition. The sequencing of experimental and filler trials was fixed across all lists so that there were never more than two consecutive experimental trials. We created six versions of each list by randomizing the order of experimental items and filler items, with the constraint that across lists, the first trial was an Agent, Non-agent, or No Entity trial an equal number of times (i.e., six), yielding 72 lists. Participants were randomly assigned to one of the lists (two participants per list).

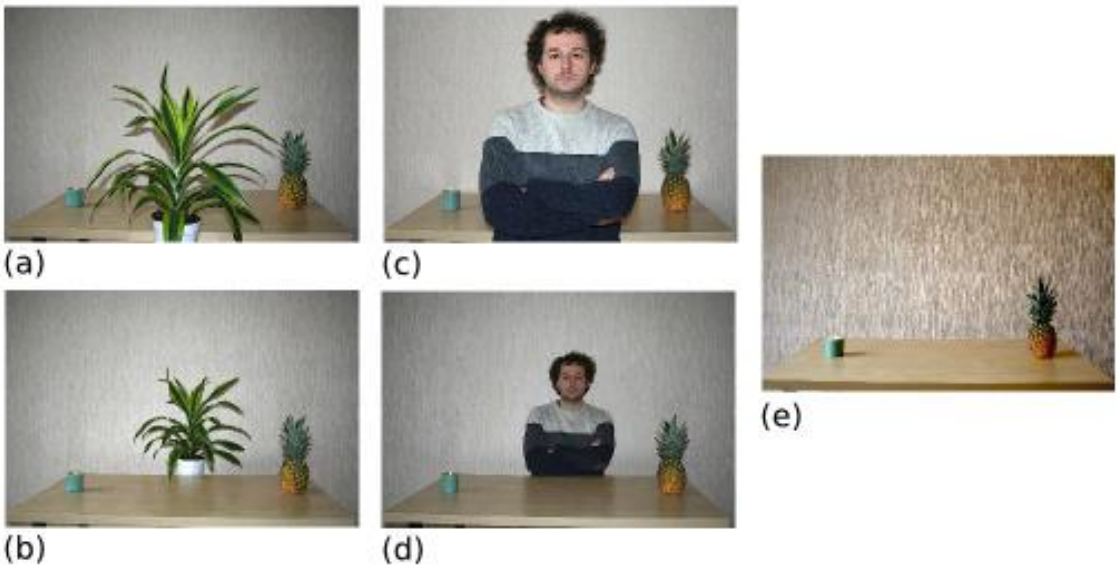

Figure 1 *Example of a set of photograph scenes (item candle/pineapple) used in Experiment 1 and in Experiment 2: (a) Non-agent Front; (b) Non-agent Behind; (c) Agent Front; (d) Agent Behind; (e) No Entity.*

2.1.3 Design

We manipulated: whether an Agent (person) or a Non-agent (plant) was present in the scene between the two objects (additional ENTITY); whether the entity was located in front or behind the objects (LOCATION). The study used an incomplete within-subjects factorial design with NO ENTITY as baseline and a 2×2 design with ENTITY (Agent vs. Non-agent) and LOCATION (Front vs. Behind) as factors.

2.1.4 Procedure

On the CrowdFlower platform, participants read the information introducing the study and gave their informed consent to take part. They were then asked to follow a link to an external website, hosted on the university's server, where the experiment was run. By clicking on the link, they were randomly assigned to one of the 72 experimental lists. The first page described the task, informing participants that they would see a series of photographs of scenes and would be asked a question about their content. Scenes were 600×400px and were displayed in the centre of the screen. The question (e.g., *On which side of the candle is the pineapple?*) and appeared on the screen below the scene at the same time as the scene itself. Participants typed in their answer and pressed Enter to continue on to the next trial.

2.2 Data analysis

2.2.1 Coding and exclusion criteria

Twelve participants provided no responses. We eliminated 72 of the experimental trials because the photo did not upload, participants typed in nonsense strings of characters, or participants did not answer the question asked. We ended up with 3096 data points from 132 subjects (1017 in the Agent, 1043 in the Non-Agent, and 1035 in the No Entity condition). Of the 3096 responses, 3095 were categorized as mentioning that the object was on the right (e.g., *right, right side, to the right*) or on the left (e.g., *left, left side, to the left*) with respect to the other object on the table and were included for analysis. One response (in the Agent-behind condition) mentioned both left and right (i.e., *my left, his right*) and was excluded. Responses were coded as 'self' if the position of the object was described from the participant's perspective and 'non-self' otherwise.

2.2.2 Models

We used mixed logit regression analysis with crossed random effects (Jaeger, 2008) to predict participants' likelihood to describe the spatial location of objects from a non-self perspective (DV: 'non-self' response = 1; 'self' response = 0) based on ENTITY and LOCATION. The crossed hierarchical nature of the model allows explicit estimation of by-participant and by-item individual differences. Predictors were coded with deviation contrasts (levels: -.5/.5) to reduce collinearity and obtain ANOVA-like interpretation of main effects. Their regression coefficient represents the difference on the log-odds scale in producing a non-self description between the two levels of the predictor. A significant positive log-odds corresponds to a higher log-odds (and so a higher probability) to produce a non-self description for the predictor level coded positively (i.e., .5). We also report odds-ratio (i.e., $\exp^{\text{log-odds}}$).

The maximal by-participants and by-items random structure justified by the design was included to avoid anti-conservatism (Barr, Levy, Scheepers, & Tily, 2013). All models were fitted using the lme4 package (version 1.1-21) (Bates, Maechler, Bolker, & Walker 2015) in R

(version 3.5.2) (R Code Team, 2018). In the event of failed convergence or singular fit, a sign of over-parameterization of the model's random structure, we followed Barr et al. (2013) and re-fitted the model after removing the random-effect terms required to allow a non-singular fit. The significance of the predictors (i.e., the model fixed-effect parameter estimates) was evaluated via Wald's Z statistic, which assesses whether the fixed-effects coefficients are significantly different from zero (given the estimated S.E.).

2.3 Results

Overall, participants took a non-self perspective on 11% of trials. 44 participants (i.e., 33% of participants) never produced descriptions from a non-self perspective; in contrast, only 1 participant (i.e., 0.75% of participants) always produced descriptions from a non-self perspective. Table 1 and Figure 2 show frequency and proportion of non-self descriptions by ENTITY and LOCATION condition.

Table 1

Frequency of *self* and *non-self* perspectives by ENTITY and LOCATION (N = 3059) in Experiment 1

		Perspective Responses	
Entity	Location	Self	Non-self
Agent	Front	448	50
	Behind	446	73
Non-agent	Front	474	49
	Behind	455	65
No Entity		932	103

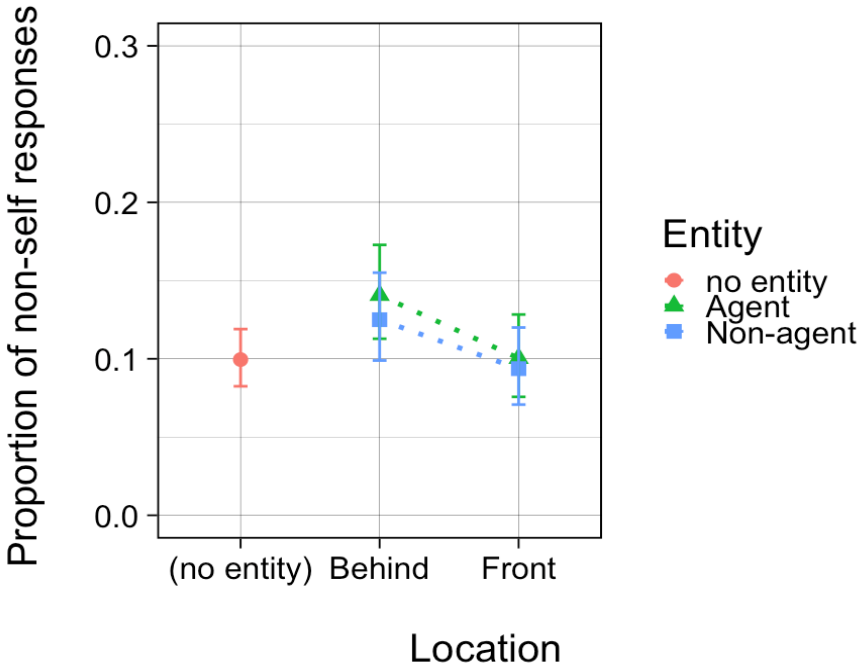

Figure 2 *Observed proportions of non-self descriptions in the five experimental conditions (Experiment 1). Error bars represent non-parametrically bootstrapped (BCa) 95% confidence intervals on the participant-wise conditional means.*

In order to assess whether the presence in the scene of an agentive vs. non-agentive additional entity influenced participants' perspective, and whether this was further mediated by the location of the entity with respect to the objects, we ran a mixed-effects logistic regression with ENTITY (Agent = .5/Non-agent = -.5), LOCATION (Behind = .5/ Front =-.5) and their interaction as fixed effects, and maximal random structure. The results showed that participants were more likely to take a non-self perspective when an additional entity appeared in the scene, located behind the table, as shown by the significant positive coefficient for LOCATION ($p < .01$). LOCATION did not interact with ENTITY ($p > .5$), indicating that the effect was not mediated by whether or not the additional entity was agentive ($p > .5$). The summary of the model is given in Table 2.

To confirm that our data supported the null hypothesis of no interaction between ENTITY and LOCATION, we performed a Bayes Factor analysis, which quantifies the likelihood of observing the data if agentive and non-agentive entities lead to different patterns of non-self descriptions in the behind and front conditions, compared to if they do not (Wagenmakers, 2007). We constructed the null-model, a GLMM with only the main fixed effects of ENTITY and

LOCATION and the same random structure as the original model; this model assumes that agents and non-agents do not differently affect perspective in the behind and front locations. The alternative model is our original model that includes the ENTITY by LOCATION interaction. We then used the two models' Bayesian Information Criterion (BIC) values to estimate the Bayes Factor as $e^{(BIC\_alternative - BIC\_null)/2}$ (Wagenmakers, 2007; Masson, 2011). The null model (i.e., without the LOCATION by ENTITY interaction) fit the data better by a Bayes Factor of $e^{(1247.9 - 1240.5)/2} = 40.5$, providing strong evidence against the hypothesis that the agentive entity led to a different patterns of spatial descriptions in the behind versus front location compared to the non-agentive entity (posterior probability in favour of the null-model BF/(BF + 1) = .98, which represents strong evidence according to Raftery's, 1995, categorization).

Table 2

Summary of the mixed logit model with Entity (Agent = .5 / Non-agent = -.5) and Location (Behind = .5 / Front = -.5) as predictors for Experiment 1 (N=2060)[1]

Predictors (fixed effects)	Parameter estimates			Wald's test	
	Log-odds (β)	S.E.	Odds Ratio [95% CI]	Z	p(β = 0)
Intercept	-3.56	.30	0.03 [0.02; 0.05]	-11.95	
Entity	-0.14	.51	0.87 [0.32; 2.37]	-0.27	.790
Location	1.60	.56	4.94 [1.66; 14.74]	2.86	.004
Entity: Location	0.40	1.01	1.49 [0.21; 10.82]	0.35	.693
Random effects	Explained standard deviation estimates				
Subjects:(intercept)	2.05				
Subjects: *Entity*	1.62	.25			
Subject: *Location*	1.03	-.95	-.36		
Subject: *Ent:Loc*	1.89	-.24	-.81 .50		
Items: *Location*	0.99				
Items: *Ent:Loc*	0.69				

[1] Because of singularity issues, random-effect terms estimated to -1/1 (if correlations) or to zero (if variance) were iteratively removed until this led to a non-singular fit of the model. All correlations between fixed-effects < |.7|

To verify whether participants produced more non-self perspectives when any additional entity was present in the scene than when no additional entity was present (cf. Tversky & Hard, 2009), we performed an additional mixed logit analysis comparing the NO ENTITY baseline with the individual four levels of the 2×2 design. To do this, we created one factor representing each combination of ENTITY and LOCATION, and then converted this factor to a deviation-coding numeric representation. The model summary is presented in Table 3. No difference was observed for either type of entity in the Front (ps > .2) or in the Behind condition (ps > .2) compared to the NO ENTITY baseline.

Table 3

Summary of the mixed logit model comparing the No Entity baseline and the four unique combinations of Entity and Location via deviation coding in Experiment 1 (N=3095)[1]. Regression coefficients represent the difference to produce a non-self description (on the logit scale) between the specified Entity-Location and the No Entity.

Predictors (fixed effects)	Parameter estimates			Wald's test	
	Log-odds (β)	S.E.	Odds Ratio [95% CI]	Z	p(β = 0)
Intercept	-3.50	0.26	0.03 [0.02; 0.05]	-13.48	
Agent/Front (d)	-0.30	0.36	0.74 [0.36; 1.50]	-0.84	0.404
Agent/Behind (d)	0.41	0.36	1.51 [0.75; 3.03]	1.15	0.251
Non-agent/Front (d)	-0.39	0.34	0.68 [0.35; 1.33]	-1.13	0.257
Non-agent/Behind (d)	0.24	0.34	1.27 [0.66; 2.46]	0.71	0.479

Random effects	Explained standard deviation estimates
Subjects: (intercept)	2.02
Subjects: *Ag/Fro*	1.43
Subject: *Ag/Beh*	1.00
Subject: *NnAg/Fro*	0.94
Subject: *NnAg/Fro*	0.95
Items: (intercept)	0.18
Items: *Ag/Beh*	0.84
Items: *NnAg/Fro*	0.36
Items: *NnAg/Beh*	0.74

[1] Because of singularity issues, random-effect terms estimated to -1/1 (if correlations) or to zero (if variance) were iteratively removed until this led to a non-singular fit of the model. Highest correlation between fixed effects, r = 0.19.

2.4 Discussion

When describing the location of an object relative to another object, participants took a self-perspective more frequently than a non-self perspective. These results are consistent with the assumption that speakers tend to adopt a self-perspective by default (Levelt, 1989). However, they were more likely to adopt a non-self perspective when the scene contained an additional entity that appeared behind the table (i.e., on the other side of the table, relative to the participant) than when it contained an additional entity that appeared in front of the table (on the same side of the table, relative to the participant). This tendency was not affected by whether the additional entity was agentive (a human) or non-agentive (a plant).

These results support previous findings that speakers' choice of spatial perspective may be influenced by non-communicative factors (e.g., Tversky & Hard, 2009). They are incompatible with the non-intentional account, as the plants did not have directional features that could attract attention and the humans' directional features were the same in the agent-front and agent-behind conditions. They are also incompatible with an actional intentional account, as the humans (and obviously the plants) were not acting on, or showing an intention to act on, the objects. In contrast, they are compatible with the perceived intentional account – but only if participants treated the non-agent as intentional and forward-facing.

If they did so, they may have transferred its intentionality and direction from the Agent condition. In accord with this possibility, some research suggests that spatial perspective-taking can be affected by experimental experience. Clements-Stephans et al. (2012) found that the orientation of featureless triangles affected how well participants determined from which triangle's position a series of pictures had been taken. But this only occurred when participants had previously been exposed to triangles depicted with eyes (hence, encouraging them to attribute intentionality to the triangles). In a similar way, participants might have adopted a non-self perspective for trials containing a person, and then transferred this perspective to scenes that included plants appearing in the same location as

the person (see also Ryskin, Brown-Schmidt, Canseco-Gonzalez, Yiu, & Nguyen, 2014; Ryskin, Wang, & Brown-Schmidt, 2016, for evidence of persistence in spatial perspective-taking).

In addition, we did not find that participants produced more non-self descriptions when an additional entity was in the scene compared to the baseline with no additional entity. This was the case for all four Entity-Location conditions. This finding reinforces the assumption that self-perspective is a strong default option for speakers. However, it may also be the result of the experimental setting, with participants being exposed to the different experimental conditions in a highly alternating way; similarly to what may have happened for the agent and no-agent conditions in the behind Location, the experimental context may have led participants to sometimes produce a non-self perspective even when no additional entity was in the scene.

To test these possibilities, we controlled for possible carry-over effects between trials in Experiment 2 by presenting the agentive (person) stimuli, the non-agentive (plant) stimuli and the no-entity stimuli in blocks.

3 Experiment 2

Experiment 2 manipulated the same factors as Experiment 1, but the three Entity conditions were presented in different blocks, with blocks counterbalanced for order of presentation across participants. If the perceived intentional account is correct but the results of the non-agentive conditions in Experiment 1 occurred because participants transferred intentionality/orientation from the Agent stimuli to the Non-agent stimuli, then participants should produce more non-self descriptions for the Agent-Behind conditions compared to the Agent-Front conditions, but no such tendency for them to produce more non-self descriptions for the Non-agent-Behind conditions. We should also find that participants should produce fewer non-self perspectives in the no-entity baseline compared to the Agent-Behind condition.

3.1 Methods

The experiment was identical to Experiment 1 except as specified in sections 3.1.1-3.1.3

below.

3.1.1 Participants

We recruited 288 participants from the same population as Experiment 1 to take part in the online experiment, under the same conditions.

3.1.2 Materials

We used the same stimuli and the same ratio between experimental and filler trials as in Experiment 1, with 8 experimental and 5 filler trials per block. We constructed six versions of each of the 12 lists of trials, one version for each order of the blocks (Non-agent –Agent – No Entity; Non-agent – No Entity – Agent; Agent – Non-agent – No Entity; Agent – No Entity – Non-agent; No Entity – Nonagent – Agent; No Entity – Agent – Non-agent), for a total of 72 lists. Four participants were randomly assigned to each list. To avoid carryover effects, a filler trial always occurred between the end of one block and the beginning of the next block. We also created three sequences of experimental and filler trials making sure that there were no more than two consecutive experimental trials, and that each block order was combined with each sequence order.

3.1.3 Procedure

At the end of the experiment, participants reported their gender, first language, and age.

3.2 Data analysis

3.2.1 Coding and exclusion criteria

We excluded 52 participants (18%) who had taken part in Experiment 1, 23 participants (8%) for whom English was not their first language, and 3 participants (1%) who provided no responses. We further excluded 1 participant who acknowledged that more than one perspective was possible but did not provide any (e.g., *It all depends on perspective*), 1 participant who described the location of the objects without taking any perspective (e.g., *opposite*), and 1 participant who explicitly took both perspectives (e.g., *my right, the cup's left*) for all experimental trials. Fifteen experimental trials were excluded from analysis because of

technical issues with photo uploading or the participants provided no response. This led to 4947 observations from 207 participants which were used for analysis (1639 in the Agent, 1653 in the Non-agent, and 1655 in the No Entity condition).

3.3 Results

On average, participants took a non-self perspective on 11.2% of trials. 88 participants (i.e., 43% of participants) never produced descriptions from a non-self perspective; in contrast, only 4 participants (i.e., 1.9% of participants) always produced descriptions from a non-self perspective. Table 4 and Figure 4 show the frequency and proportion of non-self descriptions by ENTITY and LOCATION condition.

Table 4

Frequency of *self* and *non-self* descriptions by ENTITY and LOCATION (N = 4947) in Experiment 2

		Perspective Responses	
Entity	Location	Self	Non-self
Agent	Front	727	101
	Behind	684	127
Non-agent	Front	750	78
	Behind	748	77
No Entity		1486	169

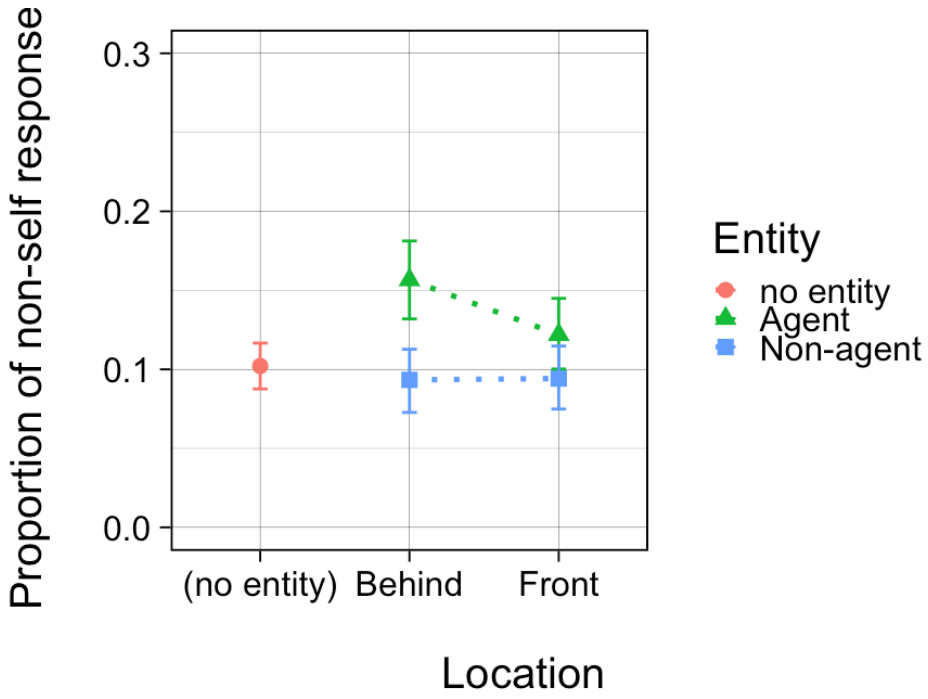

Figure 3 *Observed proportions of non-self descriptions in the five experimental conditions (Experiment 2). Error bars represent non-parametrically bootstrapped (BCa) 95% confidence intervals on the participant-wise conditional means.*

A logit mixed model analysis was used to model outcomes. No effect of ENTITY ($p > .9$) or LOCATION ($p > .1$) was found. Crucially, there was a significant interaction between these two factors ($p < .05$); participants were more likely to take a non-self perspective to describe the spatial location of objects when an agentive entity but not a non-agentive entity was in the scene, and was located behind the objects. A summary of the model is given in Table 5.

Table 5

Summary of the mixed logit model with Entity (Agent = .5, / Non-agent = -.5) and Location (Behind = .5 / Front = -.5) as predictors for Experiment 2 (N=3292)[1]

Predictors (fixed effects)	Parameter estimates			Wald's test	
	Log-odds (β)	S.E.	Odds Ratio [95% CI]	Z	p(β = 0)
Intercept	-4.41	0.37	0.01 [0.00; 0.02]	-12.06	
Entity	0.02	0.62	1.02 [0.30; 3.42]	0.03	.976
Location	0.59	0.58	1.75 [0.56; 5.43]	0.97	.334
Entity:Location	2.54	1.13	12.68 [1.39; 115.59]	2.25	.024
Random effects	Explained standard deviation estimates				

Subjects: (intercept)	2.81			
Subjects: *Entity*	2.24	0.33		
Subject: *Location*	1.28	-0.28	-0.27	
Subject: *Ent:Loc*	2.22	-0.85	-0.30	0.45
Items: *Entity*	0.70			
Items: *Location*	0.37			
Items: *Ent:Loc*	0.87			

[1] Because of singularity issues, random-effect terms estimated to -1/1 (if correlations) or to zero (if variance) were iteratively removed until this led to a non-singular fit of the model. Highest correlation between fixed effects < |.36|.

As with Experiment 1, we then tested whether participants' likelihood of describing the scenes from a non-self perspective differed in the four entity conditions compared to when no additional entity was in the scene. We ran a mixed logit analysis comparing the NO ENTITY baseline with the individual four levels of the 2×2 design (see section 2.3 for details). The model summary is presented in Table 6. No difference was observed for any conditions (ps > .1).

Table 6

Summary of the mixed logit model comparing the No Entity baseline and the four unique combinations of Entity and Location via deviation coding in Experiment 2 (N=4947)[1]. Regression coefficients represent the difference to produce a non-self description (on the logit scale) between the specified Entity-Location and the No Entity.

Predictors (fixed effects)	Parameter estimates			Wald's test	
	Log-odds (β)	S.E.	Odds Ratio [95% CI]	Z	p(β = 0)
Intercept	-4.14	0.28	0.01 [0.06; 0.03]	-14.64	
Agent/Front (d)	-1.11	0.75	0.33 [0.07; 1.43]	-1.47	0.137
Agent/Behind (d)	0.40	0.54	1.48 [0.51; 4.30]	0.73	0.467
Non-agent/Front (d)	0.22	0.47	1.25 [0.50; 3.12]	0.48	0.632
Non-agent/Behind (d)	-0.33	0.56	0.72 [0.24; 2.14]	-0.59	0.552

Random effects	Explained standard deviation estimates		
Subjects: (intercept)	2.64		
Subjects: *Ag/Fro*	2.82	0.42	
Subject: *Ag/Beh*	1.82	0.18	0.76

Subject: *NnAg/Fro*	0.93	-0.42	0.30	-0.13	
Subject: *NnAg/Fro*	1.09	0.11	0.35	0.20	0.43
Items: *Ag/Fro*	0.07				
Items: *Ag/Beh*	0.83				
Items: *NnAg/Fro*	0.20				

[1] Because of singularity issues, random-effect terms estimated to -1/1 (if correlations) or to zero (if variance) were iteratively removed until this led to a non-singular fit of the model. Highest correlation between fixed effects, r = 0.40.

To investigate whether exposure to other experimental conditions may have affected how participants responded to the no-entity baseline (i.e., no significant difference in non-self perspective use between the baseline and any of the entity conditions) despite the blocked design, we ran an additional exploratory analysis on Block 1 only, in order to isolate the effect of first exposure to each ENTITY condition. Figure 4 shows the proportion of non-self descriptions in Block 1 only. In a mixed logit model, we included the deviation-coding numeric predictor comparing each individual ENTITY - LOCATION condition against the NO ENTITY baseline as fixed effects, and the full random structure justified by the design (by-participants, this only included random intercept). Only the Agent-Behind condition triggered significantly more non-self perspective descriptions compared to the NO ENTITY baseline (β = 2.04, S.E. = .68, odds-ratio= 7.70, 95% CI [2.02; 29.38], p = 0.003); all other comparisons were not significant (p*s* > .05).

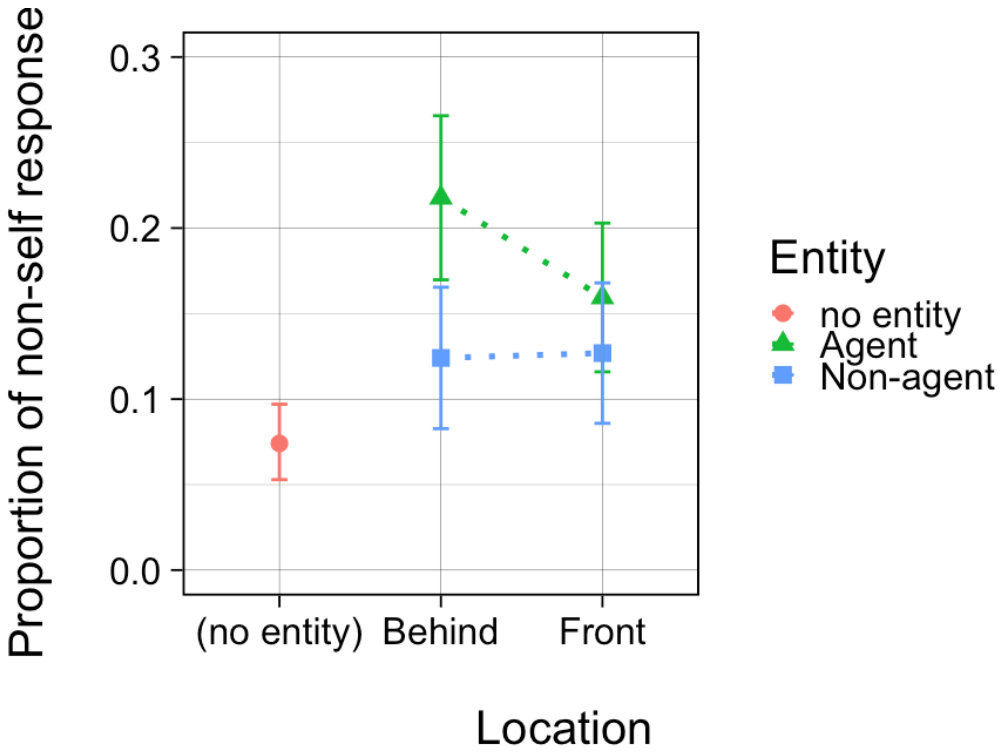

Figure 4 *Observed proportions of non-self descriptions in Block 1 of Experiment 2. Error bars represent non-parametrically bootstrapped (BCa) 95% confidence intervals on the participant-wise conditional means.*

3.4 Discussion

When participants described Agent, Non-agent, or No Entity scenes in a blocked presentation, they were consistently more likely to adopt the non-self perspective in the Agent-Behind condition compared to the Agent-Front condition (as in Experiment 1). However, they showed no such tendency in the Non-Agent conditions. These findings therefore support the perceived intentional account, and suggest that in Experiment 1, participants transferred intentionality and orientation from agents (humans) to non-agents (plants). Additionally, in Experiment 2 participants were more likely to describe the scene from a non-self perspective in the Agent-Behind condition compared to the No Entity baseline, but in Block 1 only, i.e., when participants first experienced the task and were exposed to only one Entity condition. One possible explanation for why the effect of Agent-Behind vs. No Entity did not occur when the data from the three blocks were collapsed into a single analysis is the variability in the effect induced by the exposure to other experimental conditions.

4 Experiment 3

Taken together, Experiments 1 and 2 suggest that speakers' tendency to adopt a non-self perspective can be affected by the presence of an agentive entity that has a different viewpoint from themselves and that evinces no current action or intention to act. Both experiments found more non-self descriptions in the Agent-Behind conditions (where the agent could potentially see and act on the described objects) than in the Agent-Front conditions (where the agent could not). In both experiments, the agent's perspective contrasted with the speaker's. Our results are not consistent with the non-intentional account (which assigns no role to agency), or the actional intentional account (which requires the agent to evince action or an intention to act). Instead they support the perceived intentional account, in which speakers' choice of perspective is influenced by their perception of the agent's ability to see or act on objects.

Hence Experiments 1 and 2 identify one factor affecting speakers' choice of spatial perspective that is not related to audience design considerations: the perception of an agent with the potential for action. But Experiments 1 and 2 always involved scenes in which the agent had the opposite orientation, and so an opposite perspective, from the speaker. Is speakers' choice of perspective affected by the presence of an agent with the potential for action only when it has an opposite orientation to their own? Or is speakers' choice of perspective also affected by the perception of an agent with the potential for action even when it has the same orientation as their own?

In Experiment 3, we addressed this question by focusing on scenes involving an agentive entity. We manipulated whether the agent could or could not act on the objects, and whether the agent had the same orientation as the participant or the opposite orientation. Thus as in Experiments 1 and 2, we included conditions in which the agent had the opposite orientation to the participant and either could act on the objects (5a; Can act-Opposite orientation condition, corresponding to the Agent-Behind condition in Experiments 1 and 2)

or could not act on the objects (5b; Cannot act-Opposite orientation condition, corresponding to the Agent-Front condition in Experiments 1 and 2). But we additionally included conditions in which the agent had the same orientation as the participant and either could act on the objects (5c; Can act-Same orientation condition) or could not act on the objects (5d; Cannot act-Same orientation condition).

Experiments 1 and 2 showed that speakers would be more likely to produce non-self descriptions when the scene contained an agent that had the opposite orientation to their own and that could act on the objects (Can act-Opposite orientation condition; 5a) than when the scene contained an agent that had the opposite orientation to their own and that could not act on the objects (Cannot act-Opposite orientation condition; 5b). If the pattern in Experiments 1 and 2 occurred because speakers are more likely to take the perspective of an agent that can act on objects than an agent that cannot act on objects only if that agent has the opposite orientation to their own, then participants should produce more non-self descriptions in the Can act-Opposite orientation condition than in the Cannot act-Opposite orientation condition (i.e., as in Experiments 1 and 2) – but we would not expect to find differences in the use of non-self descriptions between the Can act-Same orientation and Cannot act-Same orientation conditions. This hypothesis predicts an interaction between the agent's orientation and their potential to act, and a simple effect of action potential within the Opposite but not the Same orientation condition.

In contrast, if this finding was because speakers are more likely to take the perspective of any agent that can act on objects (irrespective of its orientation) than an agent that cannot act on objects, then participants should again produce more non-self descriptions in the Can act-Opposite orientation condition (Figure 5a) than in the Cannot act-Opposite orientation condition (Figure 5b) (i.e., as in Experiments 1 and 2) - but moreover they should produce *fewer* non-self descriptions (i.e., more self-descriptions) in the Can act-Same orientation condition (Figure 5c) than in the Cannot act-Same orientation condition (Figure 5d), because

in the Can act-Same orientation condition the agent's perspective corresponds to, and hence reinforces, the speaker's own (i.e., self-) perspective. This hypothesis predicts an interaction between the agent's orientation and their potential to act, and a simple effect of action potential within both the Opposite orientation and the Same orientation conditions, (with a positive effect in the Opposite orientation condition, and a negative effect in the Same orientation condition).

4.1 Methods

The experiment was identical to Experiment 1 except as specified in sections 4.1.1-4.1.4-below.

4.1.1 Participants

288 participants took part in the online experiment.

4.1.2 Materials

We created 12 experimental items. As in Experiment 1, an item involved two experimental objects with no clear front/behind or left/right. For each item, we created two sets of five photographed scene: one set for each possible position of the two objects on the table (see Experiment 1), each set comprising five photographs corresponding to NO ENTITY, and the four unique combinations of ACTION POTENTIAL and ORIENTATION (see Figure 5).

We then created 12 lists, each comprising 24 experimental and 16 filler trials. Overall, each list contained 8 No Entity conditions and 16 Agent conditions, 4 for each combination of ACTION POTENTIAL and ORIENTATION. In each list, each item appeared twice but never in the same condition, and participants were asked about the position of each individual object only once.

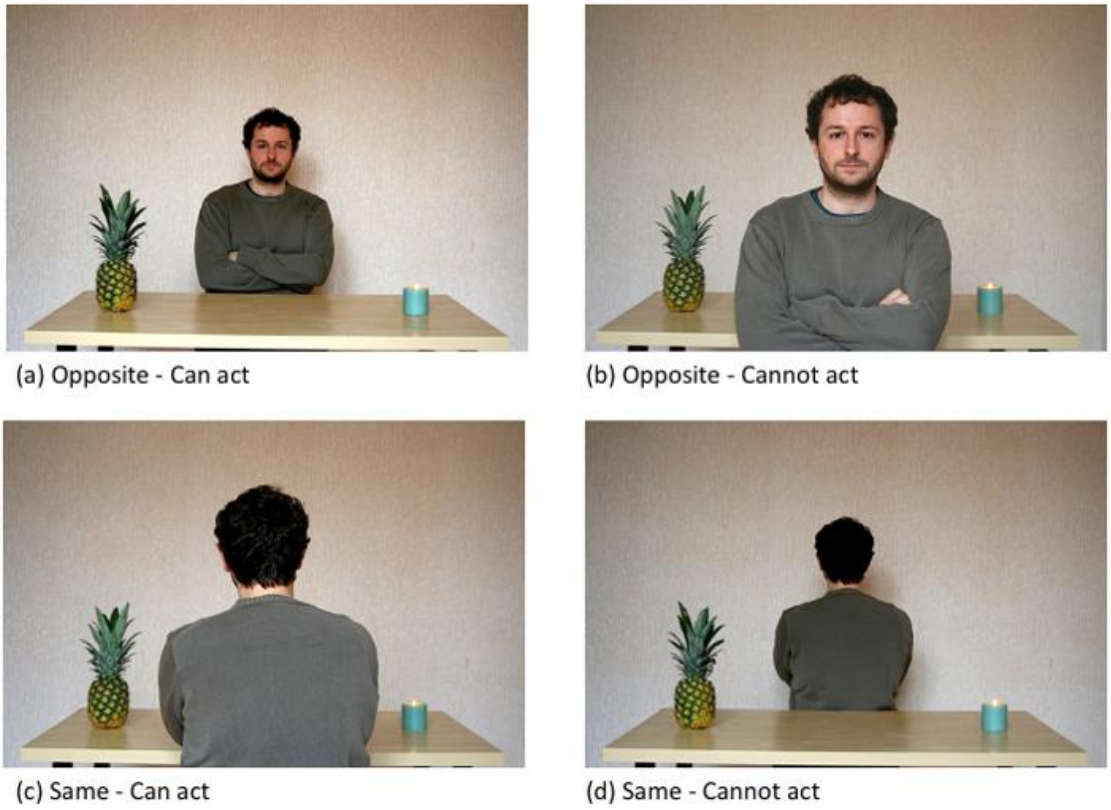

Figure 5 *Example of a set of photographed scenes (item: pineapple/candle) used in Experiment 3*

4.1.3 Design

As in the previous experiments, participants saw photos of scenes in which two objects were located on the opposite sides of a table. They described the spatial relation between the two objects by answering the question: *On which side of X is Y*? We manipulated: (i) whether an agentive entity (i.e., a person) was present in the scene (ENTITY: No Entity vs. Agent); (ii) whether the agent could act or not act on the objects (ACTION POTENTIAL: Can Act vs. Cannot Act; and (iii) whether the agent's orientation towards the scene was aligned with the speaker's or not (ORIENTATION: Same vs Opposite).

The study used an incomplete factorial design with NO ENTITY as control group and a 2×2 within-subject within-item design for ENTITY Person with ACTION POTENTIAL: Can Act vs. Cannot Act) and ORIENTATION (Same vs. Opposite) as factors.

4.1.4 Procedure

The procedure was as in Experiment 1.

4.2 Data analysis

4.2.1 Coding and exclusion criteria

We excluded 35 participants (12%) who taken part in one of our previous experiments, 26 participants (9%) for whom English was not their first language, and 4 participants (1.4%) who provided no responses. We further excluded 17 responses (0.25% of all observations) that contained explicit reference to both perspectives, 46 responses (0.67%) in which no perspective was taken (e.g., "the pear is at the opposite end of the table"), and 36 (0.52%) 'null' responses, where the participant did not produce a response. 5215 observations from 221 participants were included in the analysis (866 in Can act-Opposite, 873 in Cannot act-Opposite, 873 in Can act-Same, 871 in Cannot act-Same, and 1732 in No Entity condition).

4.3 Results

Participants used a non-self perspective to describe the spatial location of the objects on 10.2% of trials (see Table 7 and Figure 6). 105 participants (i.e., 48% of participants) never produced descriptions from a non-self perspective; in contrast, only 2 participants (i.e., 0.9% of participants) always produced descriptions from a non-self perspective.

Table 7

Frequency of *self* and *non-self* perspectives by LOCATION and ORIENTATION (N = 5215) in Experiment 3

		Perspective Responses	
Orientation	Action Potential	self	non-self
Opposite	Can act	741	125
	Cannot act	772	101
Same	Can act	805	68
	Cannot act	787	84
No Entity		1577	155

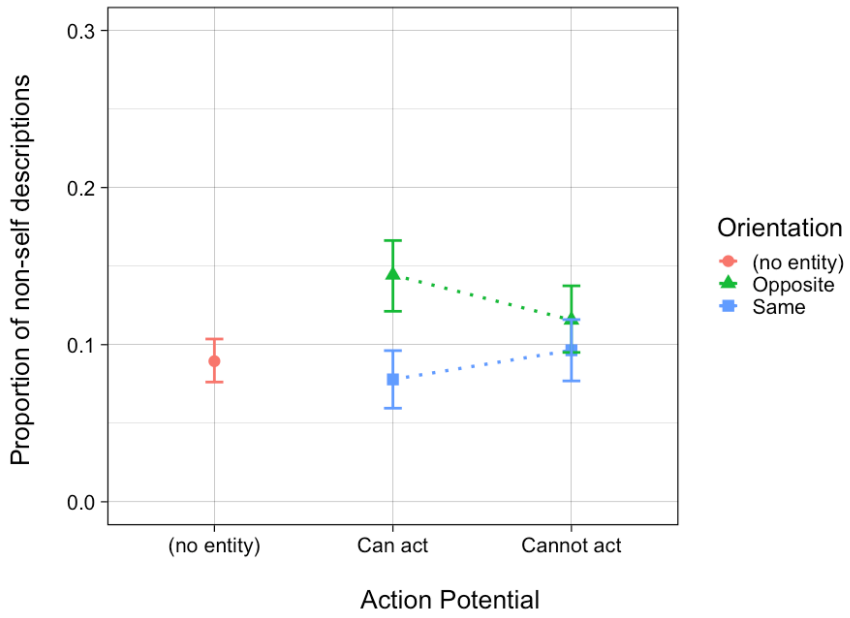

Figure 6 *Observed proportions of non-self descriptions in the five experimental conditions (Experiment 3). Error bars represent non-parametrically bootstrapped (BCa) 95% confidence intervals on the participant-wise conditional means.*

A logit mixed model with ACTION POTENTIAL, ORIENTATION and their interaction as fixed effects (see Table 8) revealed a significant main effect of ORIENTATION: Participants were more likely to produce non-self descriptions when the agent held an opposite orientation to their own compared to when it had the same orientation ($p < 0.01$). There was a significant interaction between ACTION POTENTIAL and ORIENTATION ($p < 0.01$), indicating that participants' likelihood of using a non-self perspective when the agent could act on the objects was modulated by whether the agent's orientation was the same as the participant's orientation or not. Simple effects analysis revealed a significant positive effect of ACTION POTENTIAL within the Opposite ORIENTATION conditions ($\beta = 0.42$, S.E. = .19, odds-ratio= 1.53, 95% CI [1.06; 2.21], $p = 0.02$), suggesting that when the agent's perspective was the opposite to the participant's perspective, participants were more likely to produce non-self descriptions when the agent could act on the objects than when it could not act on them. ACTION POTENTIAL had no significant effect within the Same ORIENTATION conditions ($\beta = -0.35$, S.E. = .21, odds-ratio= 0.71, 95% CI [0.47; 1.06], $p = 0.091$), suggesting that when the agent's perspective was the same as the participant's perspective, participants likelihood of producing non-self

descriptions was not affected by whether the agent could or could not act on the objects.

Table 8

Summary of the mixed logit model with Action Potential (Can Act = .5/ Cannot Act = -.5) and Orientation (Opposite = .5/ Same = -.5) as predictors for Experiment 3 (N=3483)[1]

Predictors (fixed effects)	Parameter estimates			Wald's test	
	Log-odds (β)	S.E.	Odds Ratio [95% CI]	Z	p(β = 0)
Intercept	-3.75	0.25	0.02 [0.01; 0.04]	-15.12	
Orientation	0.61	0.21	1.84 [1.23; 2.77]	2.95	0.003
Action Potential	0.02	0.14	1.02 [0.77; 1.35]	0.14	0.891
Orientation:Action Potential	0.77	0.28	2.15 [1.24; 3.72]	2.73	0.006

Random effects	Explained standard deviation estimates
Subjects: (intercept)	2.28
Subjects: *Orientation*	1.38
Items: *Orientation*	0.27
Items: *Action Potential*	0.09

[1] Because of singularity issues, random-effect terms estimated to -1/1 (if correlations) or to zero (if variance) were iteratively removed until this led to a non-singular fit of the model. All correlations between fixed effects r < |0.12|.

4.4 Discussion

Experiment 3 showed that speakers' use of a non-self perspective was influenced by the presence in the scene of an agent with action potential, but only when that agent had an opposite orientation to their own. They were more likely to use a non-self description when the agent had an opposite orientation to their own, and could see and potentially act upon objects (but evinced no current action or intention to act) than when the agent had an opposite orientation to their own, and could not see and potentially act upon objects. However, they were as likely to use a non-self description when the agent had the same orientation as their own, and could see and potentially act upon objects, as when the agent had the same orientation as their own, and could not see and potentially act upon objects. These findings suggest that when speakers select a spatial perspective, they are drawn to the

perspective of an agent who can act but only when that agent has a different perspective to the speaker.

5 General Discussion

When people describe the location of objects, they must choose a perspective from which to do so, and our experiments show that this choice is affected by more than just communicative factors. Three experiments showed that speakers' perspective was affected by the presence of another person in the scene, even though their utterances were not addressed to that person. Speakers were always more likely to adopt a self-perspective than a non-self perspective. However, they were more likely to adopt a non-self perspective when the scene included a person that had a different perspective to their own and that could see and act on the objects, compared to a scene that contained a person that could not see and act on the objects (Experiments 1 and 2). This tendency occurred even though the person was not actually acting on (or showing an intention to act on) the objects. They were also more likely to adopt a non-self perspective when the scene included a person that had a different perspective to their own and that could see and act on the objects, compared to a scene that contained no other entity, if they had not been previously exposed to other experimental conditions (i.e., Block 1 of Experiment 2). Additionally, , speakers were more likely to adopt a non-self perspective when the scene included a person with a different perspective who could see the objects, but not a personwith the same perspective who could see the objects (Experiment 3).

Participants were also more likely to adopt a non-self perspective when the scene contained a plant behind the table compared to when it contained a plant in front of the table, but they did so only when trials involving a plant were interspersed with trials involving a person (Experiment 1) and not otherwise (Experiment 2). We propose that speakers transferred intentionality and orientation from agents (humans) to non-agents (plants) in Experiment 1. Otherwise, speakers did not adopt a non-self perspective for a non-human entity. To summarise, our results suggest that speakers were drawn to the perspective of an

agent who could act, and even when they were not addressing their spatial descriptions to that agent, but only if that agent had a different perspective to the speaker

Our experiments therefore show the complex ways in which spatial perspective-taking can be affected by factors other than considerations of an addressee's immediate communicative needs. We therefore propose that speakers make an initial decision about perspective, during the phase when they are 'thinking for speaking' – developing a conceptualization of what they want to describe. They can always adopt a self-perspective, but are sometimes drawn to a non-self perspective by the presence of an entity (another human) that is sufficiently similar to themselves.

Our results demonstrate that this effect of another entity is crucially linked to the entity's potential for agency, i.e., the fact that it is able to act on the relevant objects in the scene. Previous research has suggested that people's conceptualization of spatial relations may be affected by the presence in a scene of entities with salient directional features, such as human agents (Santiesteban et al., 2014). But our studies show that the mere presence of such an entity does not in itself encourage speakers to adopt a non-self perspective when producing a spatial description: It is a necessary condition that the entity be able to act on relevant objects and offer an alternative perspective on the scene.

Other studies have shown that the presence of an agent that is in an opposite orientation and that is acting or showing an intention to act on objects can encourage speakers to adopt a non-self perspective (e.g., Furlanetto et al., 2013; Tversky & Hard, 2009). Our studies show the agent does not need to be acting on, or showing any intention to act on, the objects to affect speakers' choice of perspective: It simply needs to be able to act, as long as it has a different orientation to the speaker.

One way to interpret our findings is in terms of simulation – speakers sometimes put themselves into the shoes of another agent and speak as though they were the agent. On this account, our experiments identified some of the conditions under which speakers simulate,

specifically those in which the speaker's choice of perspective is not a consequence of audience design reasons. We suggest that when speakers viewed the scene in preparation for producing a spatial description, they incorporated the agent's location and orientation, and more specifically, the agent's visual perspective (i.e., what they could and could not see), into their representation of the objects' spatial arrangement. Thus, when the agent was oriented in such a way that the objects were in the agent's visual field, the speaker sometimes simulated the scene from the agent's perspective. In other words, apprehending the visual perspective of another person made the speaker aware of a spatial perspective other than their own. As a result, when the agent's perspective did not match participants' (self) perspective, participants sometimes took a non-self perspective when generating a description of how the objects were located. When the agent's perspective did match participants' perspective, however, such simulation did not reinforce their use of a self-perspective to generate a description. Our results suggest that the mechanisms underlying such simulation are linked to perceived agency but are independent of more explicit action-related considerations (e.g., action understanding or discernible intentionality).

Our findings are compatible with evidence from non-linguistic spatial cognition showing that another person's visual perspective can interfere with a participant's performance on spatial (Zwickel & Mueller, 2010) and non-spatial (Samson, Apperly, Braithwaite, Andrews, & Bodley Scott, 2010) target detection tasks, so long as that alternative perspective appears salient. It is also congruent with previous findings that people can easily and quickly process what another person can or cannot see, without resorting to effortful mental representations (Kessler & Rutherford, 2010; Michelon & Zacks, 2006; Surtees, Apperly, & Samson, 2013). The participants in our experiments seemed to have relied on this ability and quickly processed whether the agent could or could not see the objects (see e.g., Gardner et al., 2013, for evidence of how such processing might occur without simulation); this information then affected whether they simulated the agent's spatial perspective in the

scene as they produced their utterance. This is important, as determining what an agent can or cannot see before simulating their position in space can prevent us from unnecessary simulation (cf. Johnson & Demiris, 2005).

We cannot be sure how consistently the characteristics of the agent triggered participants to simulate its perspective when preparing their utterances. Note that people find it more effortful to make verbal left/right judgments that involve a non-self perspective than equivalent motor responses (Kessler & Thomson, 2010), most likely as a consequence of the ambiguity of *left* and *right* (Coventry & Garrod, 2004). Hence speakers may be less likely overall to simulate another agent's perspective in linguistic tasks than in non-linguistic tasks, and this may be exacerbated in contexts where there is no incentive to do so (as in our study). It is also possible that what an agent can see may not always be a sufficiently strong cue to the agent's relevance and perspective to trigger a non-self perspective in the speaker. Our results – in particular the fact that many participants only ever used a self-perspective – are also consistent with the existence of individual differences in spatial perspective-taking that can affect use of spatial language (Schober, 2009), though our experiments were not designed to investigate this question directly.

Nor can we be sure whether participants considered only one perspective when 'thinking for speaking', or whether they considered more than one perspective in parallel. On the serial account, the speaker may have initially adopted one perspective (we assume the self-perspective) when generating a description, and then sometimes been triggered to abandon it by the presence of the agent. On the parallel account, the two perspectives competed for selection, and the presence of the agent (with particular characteristics) supported the non-self perspective). Many studies have demonstrated parallel processing of alternatives at different levels of language production, including conceptualisation (see Melinger, Branigan, & Pickering, 2014), but our data are compatible with serial or parallel processing.

Our results show clearly, however, that the self perspective was most frequently selected as the basis for sentence formulation. The presence of an agent with action potential only sometimes induced speakers to produce non-self descriptions. Overall, the pattern of effects is consistent with other research on language production, which similarly finds that a range of factors such as visual salience and ease of processing affect speakers' choices in a probabilistic rather than deterministic way (Bock, 1986; Goudbeek & Krahmer, 2012).

Our experiments did not find evidence for agency-independent object-centred spatial coding, in which participants used the additional entity as a source of spatial coordinates (e.g., their intrinsic left/right axes) without simulating the alternative perspective (Heyes, 2014). But of course such mapping mechanisms using intrinsic axes might be relevant in other contexts, in particular where the relevant scenes include only the located object and the person (e.g., Mazzarella et al., 2012). This type of scene may induce participants to use the only other available entity (i.e., the person) as a source of spatial encoding for the located object, especially when the person's pose (e.g., head turned, arm stretched) created asymmetries that made their left/right axes more salient. Such an explanation could explain why previous studies using this paradigm found that, unlike a picture including a person in a neutral pose, a picture including a person engaged in action triggered more non-self descriptions than a baseline picture that did not include a person (e.g., Mazzarella et al., 2012). Hence it is possible that simulation mechanisms and intrinsic-axes mapping mechanisms may have additive effects, so that the adoption of a non-self perspective has both perceived intentional and non-intentional components. The non-intentional component might also encompass other potentially relevant considerations such as the intrinsic features or functional affordances of the located objects (e.g., Carlson-Radvansky & Radvansky, 1997; Johannsen & De Ruiter, 2013).

Overall, our findings are consistent with the view that spatial perspective taking during language production is linked to speakers processing others' representations of a

shared environment, and in particular what information those others are likely to be aware or unaware of. This conclusion is strengthened by previous findings that individuals tend to ignore stimuli that another person ignores (e.g., Frischen, Loach, & Tipper, 2009). It is also in line with evidence that knowing that another person is looking at the same stimuli affects how people encode spatial information to solve a target detection/inhibition task (Boekler, Knoblich, & Sebanz, 2011; Boekler & Zwickel, 2013). In our experiments, a more basic form of "shared attention" may be implicated. The person in our photographed scenes never directed their gaze at the objects, but appeared to experience the same scene as the participant (even if from a different perspective, and presumably at a different point in time). Speakers may have been sensitive to this "shared attention" when generating their spatial descriptions. An important question for future research is whether any such effects of shared attention depend on another person being represented as physically present in the observed environment or might also be induced by the mere belief that someone else is attending to the same environment at the same time.

In conclusion, our results demonstrate that speakers' choice of spatial perspective can be influenced simply by perceiving another agent with the potential to act, even when they are not addressing that agent. When we produce spatial descriptions, we sometimes spontaneously simulate another agent's perspective: We 'talk in another person's shoes'.

References

Avraamides, M.N., Mello, C., & Greenauer, N. (2013). Spatial representations for re-described and perceived locations. In: Tenbrink, T., Wiener, J., Claramunt, C. (eds.) *Representing Space in Cognition: Behaviour, Language, and Formal Models*, pp. 27-41. Oxford University Press.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68, 255–278. http://dx.doi.org/10.1016/j.jml.2012.11.001

Bates, D., Maechler, M., Bolker, B., & Walker, Steve. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1-48. R package version 1.1–21. Retrieved from http://CRAN.R-project.org/package=lme4

Bock, J. K. (1986). Syntactic persistence in language production. *Cognitive Psychology*, 18, 355–387.

Bock, K., & Levelt, W. J. M. (1994). Language production: Grammatical encoding. In M.A. Gernsbacher (Ed.), *Handbook of Psycholinguistics* (pp. 945-984). London: Academic Press.

Böckler, A., Knoblich, G., & Sebanz, N. (2011). Giving a helping hand: Effects of joint attention on mental rotation of body parts. *Experimental Brain Research*, 211, 531-545.

Böckler, A., & Zwickel, J. (2013). Influences of spontaneous perspective taking on spatial and identity processing of faces. *Social, Cognitive and Affective Neuroscience*, 8(7), 735-740.

Bryant, D. J., Tversky, B., & Franklin, N. (1992). Internal and external spatial frameworks for representing described scenes. *Journal of Memory and Language*, 31, 74-98.

Carlson, L. A. (1999). Selecting a reference frame. *Spatial Cognition and Computation*, 1(4), 365–379. http://doi.org/10.1023/A:1010071109785

Carlson-Radvansky, L. A., & Radvansky, G. A. (1996). The influence of functional relations on spatial term selection. *Psychological Science.* 7, 56–60.

Clements-Stephes, A.M., Vasiljevic, K., Murray, A.J., & Shelton, A.L. (2013). The role of potential agents in making spatial perspective taking social. *Frontiers in Human Neuroscience*, 7(497), 1-11.

Coventry, K.R., & Garrod, S. (2004) *Saying, Seeing and Acting: The Psychological Semantics of Spatial Prepositions.* Hove: Psychology Press.

Frischen, A., Loach, D., Tipper, S.P. (2009). Seeing the world through another person's eyes: Simulating selective attention via action observation. *Cognition*, 111, 212–218.

Furlanetto, T., Cavallo, A., Manera, V., Tversky, B., and Becchio, C. (2013). Through your eyes: incongruence of gaze and action increases spontaneous perspective taking. *Front. Hum. Neurosci.* 7:455. doi: 10.3389/fnhum.2013. 00455

Galati, A., & Avraamides, M. N. (2013). The conversational partner's perspective affects spatial memory and descriptions. *Journal of Memory and Language*, 68:2, 140-159.

Gardner, M. R., Brazier, M., Edmonds, C. J., & Gronholm, P. C. (2013). Strategy modulates spatial perspective-taking: evidence for dissociable disembodied and embodied routes. *Frontiers in Human Neuroscience*, *7*, 457. http://doi.org/10.3389/fnhum.2013.00457

Goudbeek, M., & Krahmer, E. (2012). Alignment in interactive reference production: Content planning, modifier ordering and referential overspecification. *Topics in Cognitive Science*, 4(2), 269-289.

Herskovits, A. (1986). *Language and spatial cognition: An interdisciplinary study of the Prepositions in English*, Studies in natural language processing, Cambridge: Cambridge University Press.

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

Johannsen, K., & de Ruiter, J. (2013). The role of Scene Type and Priming in the processing and selection of a spatial frame of reference. *Frontiers in Psychology*, *4*, 182.

Johnson, M., & Demiris, Y. (2005). Perceptual perspective taking and action recognition. *International Journal of Advanced Robotic Systems*, 1. http://doi.org/10.5772/5775

Kessler, K., & Thomson, L. A. (2010). The embodied nature of spatial perspective taking: embodied transformation versus sensorimotor interference. *Cognition*, 114, 72– 88. doi:10.1016/j.cognition.2009.08.015

Kessler, K., Rutherford, H. (2010). The two forms of visuo-spatial perspective taking are differently embodied and subserve different spatial prepositions. *Front. Psychol*., 1, 213. doi:10.3389/fpsyg.2010.00213

Levelt, W. J. M. (1984). Some perceptual limitations on talking about space, in A. J. van Doorn, W. A. van der Grindand, and J. J. Koenderink (eds.), *Limits in perception*, pp. 323-58. .Utrecht: VNU Science Press.

Levelt, W. J. M. (1996). Perspective Taking and Ellipses in Spatial Descriptions. *Language and Space*, 77-108.

Levinson, S. C. (1996). Language and Space. *Annual Review of Anthropology*, 25, 353-382.

Levinson, S. (2003). *Space in Language and Cognition: Explorations in Cognitive Diversity (Language Culture and Cognition)*. Cambridge: University Press.

Mainwaring, S.D., Tversky, B., Ohgishi, M., & Schiano, D.J. (2003). Descriptions of simple spatial scenes in English and Japanese. *Spatial Cognition and Computation,* 3, 3–42.

Masson, M. E. J. (2011). A tutorial on a practical Bayesian alternative to null-hypothesis significance testing. *Behavior Research Methods*, 43, 679–690. http://dx.doi.org/10.3758/s13428-010-0049-5

Mazzarella E., Hamilton A., Trojano L., Mastromauro B., & Conson M. (2012). Observation of another's action but not eye gaze triggers allocentric visual perspective. *Q. J. Exp. Psychol.,* 65, 2447–2460.

Melinger, A., Branigan, H. P., & Pickering, M. J. (2014). Parallel processing in language production. *Language, Cognition and Neuroscience*, *29*(6), 663–683. https://doi.org/10.1080/23273798.2014.906635

Michelon, P., & Zacks, J. M. (2006). Two kinds of visual perspective taking. *Perception & Psychophysics*, 68, 327–337.

Pick, H. L., Jr., & Lockman, J. J. (1981). From frames of reference to spatial representations. In L. S. Liben, A. H. Patterson, & N. Newcombe (Eds.), *Spatial representation and behaviouir across the life span: Theory and application* (pp. 39-60). New York: Academic Press.

Raftery, A. E. (1995). Bayesian model selection in social research (with Discussion*). Sociological Methodology*, 25, 111–196. http://dx.doi.org/10.2307/271063

R Core Team. (2015). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. http://www.R-project.org/

Roche, J. M., Dale, R. & Kreuz, R. J. (2010). The avoidance of ambiguity during conversation: More than mere priming or mimicry?. In S. Ohlsson & R. Catrambone (Eds.), *Proceedings of the 32nd Annual Meeting of the Cognitive Science Society* (pp. 206-211). Austin, TX: Cognitive Science Society.

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. https://doi.org/10.1016/J.JML.2014.04.003

Ryskin, R. A., Wang, R. F., & Brown-Schmidt, S. (2016). Listeners use speaker identity to access representations of spatial perspective during online language comprehension. *Cognition*, *147*, 75–84. https://doi.org/10.1016/J.COGNITION.2015.11.011

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

Santiesteban, I., Catmur, C., Hopkins, S.C., Bird, G., & Heyes, C. (2014). Avatars and arrows: Implicit mentalizing or domain-general processing? *J. Exp. Psychol.-Hum. Percept. Perform.,* 40, 929-937.

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

Schober, M.F. (2009). Spatial dialogue between partners with mismatched abilities. In K.R. Coventry, T. Tenbrink, & J.A. Bateman (Eds.), *Spatial language and dialogue* (pp. 23-39). Oxford: Oxford University Press.

Slobin, D. (1987). Thinking for speaking. In J. Aske, N. Beery, L. Michaelis, and H. Filip (Eds.), *Berkeley Linguistics Society: Proceedings of the Thirteenth Annual Meeting*, 435-444. Berkeley: Berkeley Linguistics Society.

Surtees, A., Apperly, I., & Samson, D. (2013b). The use of embodied self-rotation for visual and spatial perspective taking. *Frontiers in Human Neuroscience*, 7, 698. doi: Artn698 Doi 10.3389/Fnhum.2013.00698

Surtees, A.D.R., Noordzij, M.L., & Apperly, I.A. (2012). Sometimes losing your self in space: Children's and adults' spontaneous use of multiple spatial reference frames. *Developmental Psychology,* 48, 185-191.

Talmy, L. (1983). How language structures space, in H. Pick and L. Acredolo (eds.). *Spatial orientation: Theory, research and application*, pp. 225-82. NewYork: Plenum Press.

Tversky, B. (1996). Spatial perspective in descriptions. In P. Bloom, M. A. Peterson, L. Nadel, & M. F. Garrett (Eds.), *Language, speech, and communication. Language and space*, pp. 463-491. Cambridge, MA, US: The MIT Press.

Tversky, B., & Hard, B.M. (2009). Embodied and disembodied cognition: spatial perspective-taking. *Cognition,* 110, 124–129.

Wagenmakers, E. J. (2007). A practical solution to the pervasive problems of p values. *Psychonomic Bulletin & Review*, 14, 779–804. http://dx.doi.org/10.3758/BF03194105

Yoon S. O., Koh S., & Brown-Schmidt S. (2012). Influence of perspective and goals on reference production in conversation. *Psychon. Bull. Rev.,* 19, 699–707

Zwickel, J. (2009). Agency attribution and visuospatial perspective taking. *Psychon. Bull. Rev.* 16, 1089–1093.

Zwickel, J., & Mueller, H.J. (2010). Observing fearful faces leads to visuo-spatial perspective taking. *Cognition*, 117, 101-105.

Notes

[i] We use ‘speaker’ throughout to refer to someone producing language, whether in the spoken or written modality.

[ii]The distinction between a self- and a non-self perspective (i.e., a self vs. non-self origin of the coordinate system) is independent of the distinction between different types of reference frame, such as Levinson’s (2003) contrast between a relative reference frame, based on the viewpoint of a perceiver, and an intrinsic reference frame, based on the directional features of another entity in the environment. Crucially for our study, both relative and intrinsic reference frames are compatible with a self-perspective and with a non-self perspective. I can describe the same object as The ball to the left of the vase with respect to my own viewpoint (self relative) or another person’s viewpoint (non-self relative), or as The ball to the left with respect to my own axes (self intrinsic), or another person’s axes (non-self intrinsic).