On the automaticity of egocentricity: A review of the Egocentric Anchoring and Adjustment model of perspective taking

PAULA RUBIO-FERNÁNDEZ

Abstract

This review article makes a critical evaluation of the Egocentric Anchoring and Adjustment model of perspective taking put forward by Keysar, Barr and collaborators (Horton & Keysar, 1996; Keysar et al., 2003; Barr & Keysar, 2007). According to this model, in taking other people's perspectives (e.g. in verbal communication), our own, egocentric perspective works as an automatic default that needs to be deliberately corrected by a sequential mechanism of perspective adjustment. This two-stage model of perspective taking is distinguished from more general accounts of an egocentric bias in human cognition. The author argues that a design feature biasing participants towards the egocentric response makes it impossible to evaluate to what extent egocentric anchoring is indeed an automatic process in perspective taking, or simply the result of a systematic confound in Keysar et al.'s experiments. In this sense, their findings do not necessarily contradict other studies that have shown rapid and effective use of theory of mind abilities in communication using unbiased paradigms and multiple measures of egocentricity.

1 Introduction

It is a common assumption of pragmatic theories that in everyday interaction – particularly in communication, adults systematically rely on their theory of mind (i.e. our capacity to take somebody else's perspective and represent their beliefs and intentions, especially when they are different from our own). Contrary to this common assumption, a growing body of experimental literature has offered strong evidence that adults design and interpret utterances from an egocentric perspective, adjusting to the other's perspective only when they make an error (Keysar, Barr & Horton, 1998b;

Keysar, Barr, Balin & Brauner, 2000; Keysar, Lin & Barr, 2003; see Barr & Kesyar (2007) for a detailed review). The aim of this paper is to examine the strength of this twofold claim in view of the experimental paradigm used to investigate it. This critical evaluation seems necessary on the face of conflicting empirical evidence that shows, with the use of similar techniques, that perspective taking is fast and effective in communication (e.g. Nadig & Sedivy, 2002; Hanna, Tanenhaus & Trueswell, 2003; Hanna & Tanenhaus, 2004; Tanenhaus & Brown-Schmidt, 2008).

Before reviewing the relevant literature, a distinction needs to be drawn between the general view that the egocentric perspective is the dominant perspective in human cognition (a view which in its broadest formulation would follow simply from the fact that we perceive the world from our own sensory organs) and the twofold claim that the egocentric perspective is an automatic default and theory of mind works only as a sequential, deliberate correction mechanism of this automatic bias (Keysar et al., 2003; Epley, Keysar, Van Boven & Gilovich, 2004a; Epley, Morewedge & Keysar, 2004b). Despite the relation between a general egocentric bias and Keysar et al.'s claims, the latter does not necessarily follow from the former, with other views of the egocentric bias remaining impartial as to whether another's perspective can be generated in parallel with our own or only sequentially (e.g. Birch & Bloom, 2004, 2007)[1].

In his critical review of early experimental work on common ground, Keysar (1997) argues that these studies might have confounded common ground with privileged ground (i.e. the information that is mutually known or salient to both speaker and addressee with the information that is only known or salient to the addressee). Imagine that a man walks into a flower shop and asks the florist "How much is that flower?" According to Keysar, if the florist understands that the costumer is referring to a flower that is mutually salient to both of them, this does not necessarily show that she selected that particular flower because it was part of her common ground with the costumer: she might have selected it simply because it was salient to her (1997:258). Following this argument, Keysar's research on perspective taking has been characterized by distinguishing common and privileged ground in their experimental designs. Despite the advantage that such an important distinction should in principle

---

[1] Keysar and his colleagues describe egocentric anchoring and adjustment as sequential phases in processing: "people adopt others' perspectives by initially anchoring on their own perspective and only subsequently, serially and effortfully accounting for differences between themselves and others" (Epley et al., 2004a:328; see also Keysar, Barr, Balin & Paek, 1998a; Keysar et al., 2003; Keysar, 2007). It must be noted, though, that Barr and Keysar (2002) mention the possibility that these two processes might operate in parallel or in cascade (p.392). Nonetheless, it is unclear how their model could predict that perspective taking operates in parallel with egocentric anchoring, when they argue that the former process works only as a deliberate *correction* mechanism of the latter, normally playing no role in language processing unless an error is detected (e.g. Keysar et al., 1998a, 1998b; Barr & Keysar, 2002, 2007; Epley et al., 2004b).

offer to the investigation of perspective taking, my argument in this paper is the reverse of Keysar's: if a certain flower was made highly salient to the florist alone, the fact that she may initially disregard other flowers that are mutually salient to her and her costumer does not necessarily show that she normally makes limited use of common ground in interpreting language.

2 Egocentric Anchoring and Adjustment – *or just egocentric favouring?*

For more than a decade, Boaz Keysar, Dale Barr and their various collaborators have steadily provided empirical evidence that perspective taking – understood as part of the normal adult theory of mind, is an effortful, controlled mechanism that corrects the automatic bias of our own, egocentric perspective (e.g. Horton & Keysar, 1996; Keysar et al., 1998b, 2003; Epley et al., 2004a; Keysar, 2007). In linguistic communication, this two-stage model of perspective taking results in utterances being planned and interpreted egocentrically, with common ground being taken into account only in a later, optional monitoring phase (Barr & Keysar, 2005, 2007). Their investigation has covered both language production and interpretation, and their experimental methods have usually included innovative on-line measures of language processing, often combining some version of the referential communication task (Glucksberg, Krauss & Higgins, 1975; Krauss & Glucksberg, 1977) with eye-tracking techniques. Off-line experiments or studies that did not aim at drawing a distinction between automatic and controlled processes in perspective taking will not be reviewed here, although their results offer support to a more general view of the egocentric bias (e.g. Keysar, 1994b, 2000; Keysar, Ginzel & Bazerman, 1995; Keysar & Henly, 2002; Epley et al., 2004a; Wu & Keysar, 2007a; Shintel & Keysar, 2007; cf. Gerrig, Ohaeri & Brennan, 2000).

   The tenets of the Egocentric Anchoring and Adjustment model of perspective taking can be summarised as follows:

i)   People usually operate egocentrically in their interaction with others, correcting their own perspective only when they detect an error (e.g. when deriving the wrong interpretation of an utterance, not having considered the speaker's intentions).
ii)   When people take somebody else's perspective, their own, egocentric perspective works as an *automatic default* that needs to be corrected by a *sequential* and *deliberate* mechanism of perspective adjustment.
iii)   Even though perspective taking works as a reflective correction mechanism, the adjustment might not always be sufficient, with people showing a bias towards their own perspective even when trying to take somebody else's.

It follows from this model that an egocentric bias affects adult perspective-taking at two different levels: first, the egocentric perspective is always the first one to be adopted and is only corrected if an error is detected; and second, the correction of the egocentric perspective may often be insufficient[2].

In recent years, empirical evidence of an egocentric bias in human cognition has been observed in different areas of psychology (see Birch and Bernstein (2007) for a review). For example, individuals are more likely to predict that somebody else will appreciate the ironic meaning of an expression if they themselves know that it was intended ironically than if they do not know the intention behind the message (Keysar, 1994b). A more common experience of this egocentric bias may arise in a classroom setting, for example, where the teacher might think that he has given a clear explanation of a simple problem while most of his students did not understand the explanation and actually found the problem quite difficult. It is important to note, however, that this general egocentric bias can be explained by claim (iii) above, without necessarily supporting the two main claims of the Egocentric Anchoring and Adjustment model. In this sense, it is possible to conceive an alternative model of perspective taking where people systematically rely on their theory of mind in their interaction with others and can easily access their perspectives without needing to deliberately correct their own perspective first. This alternative model of perspective taking would be in principle compatible with a general egocentric bias, which would result from the natural dominance of our own perspective – something like the pull of our own subjectivity.

This paper will focus on the first two claims of the above summary, which are specific to Keysar et al.'s model and are not necessarily confirmed by empirical evidence of a general egocentric bias in human cognition. The paper will be structured as follows: first, the experimental paradigm used to test the Egocentric Anchoring and Adjustment model will be examined in a critical literature review. Secondly, various baseline issues around the experimental results supporting the two-stage model of

---

[2] A pragmatic model of language interpretation that would fit the main claims of the Egocentric Anchoring and Adjustment model of perspective taking is the Gricean view of figurative language, known in the psycholinguistics literature as 'the standard pragmatic view' (Grice, 1975, 1989; see Glucksberg, 1989, 2001): in processing a figurative expression, interpreters derive first the literal meaning and when they realise the infelicity of this initial interpretation, they then derive the figurative interpretation that was intended by the speaker. Ironically (pun intended), Keysar's early work in collaboration with Sam Glucksberg offered conclusive empirical evidence against this sequential model of metaphor interpretation and in favour of the view that the literal, unintended meaning of a metaphoric expression is derived in parallel with the figurative, communicated meaning (e.g. Glucksberg & Keysar, 1990, 1993; Keysar, 1989, 1994a). Although beyond the scope of this paper, it is debateable whether Keysar's early work on metaphor interpretation is compatible with his current two-stage model of language processing (Barr & Keysar, 2005, 2007; Keysar, 2007).

perspective taking will be discussed. To conclude, seemingly contradictory empirical evidence showing that perspective taking can operate fast and effectively in communication will be reconciled with Keysar et al.'s work. Finally, I will explore an alternative conclusion to Keysar et al.'s claim that important elements of theory of mind are not fully incorporated into the language processing system.

In one of their earliest studies, Horton and Keysar (1996) asked participants to describe simple figures to a confederate listener in the context of a similar figure that could be either present to both the speaker and the listener (Common Ground condition) or only to the speaker (Privileged condition). Descriptions under no time constraints appeared to incorporate common ground with the listener (i.e. the target object was described with reference to the contrast object only in the Common Ground condition). However, common ground was not used when the speakers were under time pressure (i.e. descriptions were made with reference to the contrast object irrespective of the experimental condition). Horton and Keysar argue that these results support their Monitoring and Adjustment model of language production: speakers plan their utterances using information that is available to them regardless of considerations of common ground, which only come into play as a correction mechanism in the monitoring phase of production[3]. When participants were given enough time to revise the planning of their utterances, they produced descriptions that were sensitive to differences in common ground. However, under time pressure, they produced descriptions that had not been properly monitored and therefore did not observe considerations of common ground.

Crucially for their investigation, participants were aware of whether the contrast object was present only to them, or to their listener as well. However, this simple design can be argued to favour the egocentric response in both time conditions: the targets that participants had to describe to the confederate listener were simple clip-art images that differed from the contrast object along one dimension (e.g. a light green and a dark green dinosaur). Because of the simplicity of these figures, the dimension of contrast with the paired object would have been clearly salient and therefore favoured in terms of the possible dimensions of attribution of the target object (e.g. hue would have been more salient than size in the above example). Thus, describing a dinosaur as "a light green dinosaur" in the context of a dark green dinosaur or a circle as "a small circle" in the context of a larger circle seems easier than thinking of an

---

[3] As it will become clear in this literature review, Keysar and his collaborators have referred to their model of perspective taking by different names during the years (e.g. 'Monitoring and Adjustment' or simply 'Perspective Adjustment'). In this paper, I adopt the most recent name, 'Egocentric Anchoring and Adjustment' (Epley et at., 2004a; Barr & Keysar, 2005), which is also the one that best describes the tenets of the model: perspective taking in general, and language interpretation in particular, operate in two stages, an early, automatic egocentric phase and a controlled, common-ground sensitive monitoring phase (Epley et al., 2004b; Barr & Keysar, 2007).

alternative description in absolute rather than relative terms – especially under time constraints. Critically, this difference in difficulty can be observed not only in the Speeded condition, but also in the Unspeeded condition, where it took participants on average 543ms longer to start producing the critical descriptions in the Privileged condition than in the Common Ground condition – a significant difference both by Subjects and by Items (see reference for details). Even though Horton and Keysar (1996) take these results as supportive of their Monitoring and Adjustment model, this bias towards the egocentric response stacks the deck against the alternative, Initial Design model of common ground proposed by Clark and Marshall (1981) (see Polichak & Gerrig (1998) for a different, theoretical critique and Keysar & Horton (1998) for a rebuttal).

Keysar, Barr, Balin and Paek (1998a) also investigated two models of common ground in the comprehension of referential expressions. In their first experiment, a confederate played a communication game with the participant. The confederate was presented with a series of scenarios (e.g. a short story about Joe and Rachel, who work for a delivery company) and asked the participant a comprehension question for each scenario (e.g. "What did she deliver?"). In order to be able to answer these questions, participants were given a critical sentence (e.g. "Rachel delivered the sofa"). In addition to this main task, participants had to memorise another sentence that would potentially interfere with the first (e.g. "Marla delivered a cake"). This second sentence was supposed to be just a memory load for the participants, who were specifically instructed to use only scenario-relevant information in answering the comprehension questions. Thus, even though participants received both scenario-relevant and memory-load sentences, only the former were considered as part of their common ground with the confederate.

The results show that entities not in common ground interfered with the processing of the comprehension questions, both in terms of error rates and response latencies (see reference for details). Keysar et al. (1998a) take these results as supportive of the Unrestricted Search hypothesis, according to which the search for referents is not restricted to entities in common ground. Analogously to the language production version of their model, Keysar et al. argue that "the role of mutual knowledge is only to correct interpretation errors" (1998a:16). However, because participants were instructed not only to read the sentences but also to memorise them and bear them in mind (see the Appendix in Keysar et al. (1998a) for their instructions) and because the memory-load sentence was always presented after the scenario-relevant sentence, the interference effect was clearly maximised. The authors argue that the memory-load sentence was always presented in second position to make sure that participants were not confused about which sentence was the scenario-relevant one and which was the memory load (Keysar et al., 1998a:7). Although this could have been a potential problem if the order of the sentences had been manipulated within participants, it

could have been avoided simply by presenting half of the participants with the memory-load sentence first, and the other half with the memory-load sentence second. Only by controlling the order of the sentences would it be possible to evaluate the extent to which the results of Keysar et al. (1998a) indeed reflect an initial blindness to considerations of common ground, rather than simple recency and salience effects in reference resolution.

The second experiment in Keysar et al. (1998a) suffered from a similar bias towards the egocentric response, this time using a version of the referential communication task coupled with an eye-tracking technique (see below for a description of this method). In assessing how generalizable the results of these two experiments are, Keysar et al. argue that unintended antecedents should interfere with the process of reference assignment in every situation "when the perspectives [of the speaker and the addressee] diverge and when the addressee has privileged knowledge of potential referents" (1998a:16). However, given the bias towards the egocentric response in their two experiments, it is possible that their results might only extend to cases where the unintended antecedent is not only part of the privileged knowledge of the addressee but also the *most salient* antecedent at the time of processing the referential expression.
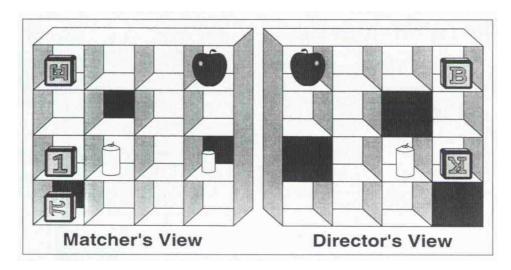
The unbalanced salience of referential candidates is a common artefact in Keysar's studies, which is not only limited to those experiments that tried to distinguish between the processes of egocentric anchoring and perspective adjustment. Barr and Kesyar (2002), for example, report a study that investigated to what extent language comprehension is guided by linguistic precedents and, most importantly for the present discussion, to what extent these linguistic precedents need to be part of the common ground between the speaker and the addressee. In the two experiments that manipulated common ground, the interpreter-privileged precedent was highly salient at the time of processing the corresponding referential expression. In one of these experiments, participants had to choose between two pictures following the instructions of a speaker. In the pre-test phase of the experiment, participants were presented with a picture of a flower and a picture of a car, for example, and had to select the latter when hearing the word 'car'. In the entrainment phase that immediately followed, the context of the targets changed and the referential expressions became more precise. For example, when the same flower was presented in the context of another flower, it was referred to as a 'carnation' (in contrast with the other flower, which was a daisy). In the last phase of the experiment, participants were presented again with the original pictures of the flower and the car. In this context, if the instructions referred to the flower target as a carnation they would be overspecific since the basic-level term would be enough to distinguish it from the car. However, participants' expectations for the subordinate-level precedent interfered with the identification of the car target, showing an initial preference for the flower (note that

'car' overlaps phonologically with the onset of 'carnation', a sophisticated feature that characterised all critical items). Crucially, this interference was observed even when the speaker in the last phase of the experiment was different from the speaker in the previous phases (i.e. the instructions had been recorded by a different speaker in the last phase).

In view of these results, Barr and Keysar (2002) argue that listeners systematically anchor comprehension in linguistic precedents, to the extent that they expect speakers to overspecify referents in certain contexts. Moreover, listeners' expectations that speakers will use linguistic precedents are so strong that they are extended to other speakers who have not previously used the same referential expressions, thus disobeying partner-specific pacts (Barr & Keysar, 2002:413). However, a simpler explanation of their results emerges from the details of their experimental design: whereas participants heard the terms 'car' and 'flower' once during the initial pre-test phase, the subordinate-level terms 'sportscar' and 'carnation' were used five and six times, respectively, during the intermediate entrainment phase. Moreover, even though different fillers were used in all three phases, the critical subordinate-level precedents were always used in the last block of trials before the final phase where the targets were referred to again as 'car' and 'flower'. It is clear that, at that point in the task, the words 'sportscar' and 'carnation' would have been more highly activated than the basic-level counterparts, both in terms of frequency and recency of processing. The fact that these lower-level processing effects are more powerful than higher-level considerations of context and common ground would in itself be an important finding for psycholinguistics research. However, the bias towards the egocentric response weakens Keysar et al.'s results as empirical evidence in favour of their theoretical claims and against the alternative, partner-specific model of lexical entrainment put forward by Brennan and Clark (1996) (see Metzing & Brennan (2003) for a different methodological critique and further results, and Shintel & Kesyar (2007) for a follow-up).

Keysar, Barr and Horton (1998b), Keysar, Barr, Balin and Brauner (2000), Keysar, Lin and Barr (2003) and Epley, Morewedge and Keysar (2004b) (see also Keysar et al. (1998a), Experiment 2) tested the Egocentric Anchoring and Adjustment model of perspective taking using an eye-tracking technique with a version of the referential communication game in which a participant is asked to move objects around a vertical grid of squares following the instructions of a confederate sitting on the other side of the grid. Strategically, some of the slots in the grid are visible to both participant and confederate, whereas others have been blinded so that the contents are only visible to the participant. In a critical trial, the confederate asks the participant to "put the bottom block below the apple", for example; when the lowest of the three blocks on display is in the participant's privileged view. A diagram with the display of objects from the participant's and the confederate's perspectives is shown in *Figure 1*. Participants

showed a tendency to look first at the block at the bottom of the grid before reaching for the one in the middle (i.e. the bottom one from the confederate's perspective).

Figure 1

The display of objects from the participant's perspective (i.e. the matcher's) and the confederate's perspective (i.e. the director's). The critical instructions were "Put the bottom block below the apple" (from Keysar et al., 1998b:49).

In these studies, both the objects and the occluded cells were normally moved around the grid between trials. However, the interference of the egocentric target (i.e. the occluded block at the bottom of the grid) was observed even when the occluded cells were kept constant during the experiment (Keysar et al., 2000, Experiment 2)[4].
Participants also experienced the interference of the occluded block at the bottom of the grid when this object was put inside a paper bag so that the participants were aware of the contents of the bag but could not see them directly (Keysar et al., 2003). This is an important finding since it extends the effect of the egocentric bias from the perceptual to the epistemic level. In the same study, participants showed an egocentric bias even when they thought that the confederate was not only ignorant about the contents of the paper bag but actually mistaken, thus showing a potential limitation in their theory of mind use (Keysar et al., 2003; Barr & Keysar, 2007). Epley et al.

---

[4] A possible test of the Egocentric Anchoring and Adjustment model in this condition would be to see whether participants experienced less interference across trials. If participants got better at ignoring the occluded cells during the course of the experiment, such practice effect would challenge the view that egocentric anchoring is an automatic process, given that such processes are by definition impervious to one's control or learning. Unfortunately, Keysar et al. (2000) do not report that analysis of the data of their second experiment where the occluded cells were always in the same position.

(2004b) observed similar results when testing children and adults[5]: both groups processed language egocentrically, the only difference observed being the speed of correction of the automatic egocentric anchoring. Epley et al. argue that both children and adults have to overcome their initial egocentrism *each time* they attempt to adopt another's perspective (2004b:765). Overall, Keysar and collaborators argue that "although adults can reflectively and deliberately use this sophisticated aspect of their theory of mind [i.e. their ability to distinguish between their own beliefs and those of others], this ability is not yet incorporated enough into the routine operation of the interpretation system to allow spontaneous, non-reflective use" (Keysar et al., 2003:28).

While these results converge in appearing to support the Egocentric Anchoring and Adjustment model of perspective taking, the findings are difficult to interpret because all the experiments share a bias towards the egocentric response. In these studies, the object that *best* fit the referential description in the instructions (e.g. the block at the bottom of the grid for "the bottom block", or the smallest of three candles for "the small candle") was always visible to the participant but occluded from the confederate giving the instructions. It can therefore be argued that the results in Keysar et al. (1998a, 1998b, 2000, 2003) and Epley et al. (2004b) do not necessarily show that the participants' egocentric view is an automatic bias, but simply that the experimental design systematically favoured the egocentric response, requiring correction to perform optimally[6].

3 Baseline issues around the Egocentric Anchoring and Adjustment model
3.1 Tasks and conclusions

Wu and Keysar (2007b) investigated whether people from collectivistic and individualistic cultures were different in applying their perspective-taking abilities. For

---

[5] The average age of the children in Epley et al. (2004b) was 6.2 years, with the actual ages ranging between 4 and 12 years. The authors report that removing the older children in the group from the statistical analyses did not make a significant difference to the results.

[6] In addition to the space ambiguity ("move the bottom block") and the size ambiguity ("move the small candle") used in the other studies, Epley et al. (2004b) used lexical ambiguity in one of their critical trials. Participants were presented with an occluded stuffed rabbit and a chocolate figure in common view, both of which could be potential referents for the expression "the bunny". It is likely that, since Easter bunnies are seasonal whereas children interact with stuffed animals all year around, the favoured candidate in this trial would have also been the occluded one. For what is worth, if one looks up the term 'bunny' in Google, eight images of stuffed bunnies appear before the first image of an Easter bunny comes up. In any case, the experimenters did not counterbalance the type of presentation of the two referential candidates.

this purpose, they compared the performance of Chinese and American participants in the referential communication game. However, in this study, the experimental design did not favour the egocentric response: participants from both cultures were asked to "put the block below the apple", for example; when there were only two blocks in the grid, one in common view and the other in the privileged view of the participant. It is important to note that, contrary to what happened with the occluded block at the bottom of the grid in "move the bottom block" (see *Figure 1*), in Wu and Keysar's design the occluded block did not fit the description in the instructions *better* than the block in common view. That is, other than for the fact that one was in the participant's privileged view while the other was in common view, both blocks were equally good candidates for the referential expression "the block", and so the design did not favour the egocentric response.

Wu and Keysar (2007b) observed that Chinese participants were indeed significantly better than American participants at focusing on the intended target. As in their previous eye-tracking experiments, effective use of perspective taking was measured with reference to an unrelated object in an occluded cell, which served as a baseline. Unlike the Americans, when Chinese participants processed an instruction such as "Put the block below the apple", they did not experience more interference from an occluded block than from an occluded toy monkey, for example. Wu and Keysar conclude from these results that "the interdependence that pervades Chinese culture has its effect on members of the culture over time, taking advantage of the human ability to distinguish between the mind of the self and that of the other, and developing this ability to allow Chinese to unreflectively interpret the actions of another person from his or her own perspective" (2007b:605).

The results in Wu and Keysar (2007b) are not only interesting but also important for our understanding of adult theory of mind across different cultures. The advantage of Chinese over American participants in a game that requires perspective taking might be independent of the details of the task. However, in order to extend their previous findings (Keysar et al., 2000, 2003) and conclude that, because of the collectivistic nature of their cultural background, Chinese people have better integrated their perspective-taking abilities into their language processing system and so are able to take the perspective of others *spontaneously* while Americans can do so only *reflectively* (Wu & Keysar, 2007b:605), Chinese participants need to be tested on the original task where other participants experienced difficulty. It is still possible that, even if better than the Americans in their earlier experiments, Chinese participants might first fixate on the egocentric target if asked to "put the bottom block below the apple" when a third block at the bottom of the grid is occluded from the confederate – thus showing a less spontaneous and efficient use of their perspective-taking abilities. Such a pattern of results could be taken as evidence that the findings of Wu and Keysar (2007b) and Keysar et al. (2000, 2003) are more the result of the different

manipulations of their task than of a differential integration of theory of mind in the language processing system of Chinese and American individuals.

3.2 Measures of egocentricity

Another important test for the Egocentric Anchoring and Adjustment model of perspective taking would be the relative speed with which American participants fixated on the occluded and the common-ground blocks in the unbiased test used in Wu and Keysar (2007b). According to the predictions of the Egocentric Anchoring and Adjustment model, if people experience an 'automatic moment of egocentrism' in taking another's perspective (Epley et al., 2004b:760) – or at least people from individualistic cultures, American participants should have been equally likely to fixate on the occluded block than on the common-ground block during the initial stages of processing the instruction "Put the block below the apple" (see Barr & Keysar (2007:317) for a similar prediction). Unfortunately, Wu and Keysar (2007b) do not report that analysis of their eye-tracking data. However, a similar study by Nadig and Sedivy (2002) showed a different pattern of results with 5- to 6-year-old children: these young participants used common-ground information from the initial stages of language processing. For example, by the offset of the noun 'glass' in the instruction "Pick up the glass", children began to fixate on the intended glass more frequently than on the occluded competitor. It seems safe to assume that, if the American children in Nadig and Sedivy (2002) could make such an efficient use of their perspective-taking abilities, both the Chinese and the American adults in Wu and Kesyar (2007b) would have performed at least as well – even though the Chinese might have outperformed the Americans on this measure as well[7].

Barr and Keysar (2007) argue against the relative measure of egocentricity used by Nadig and Sedivy (2002) by pointing out that, *in general*, participants might fixate more frequently on the co-present objects in the grid than on the private objects. Therefore, a preference for the target object in common view might not be due to a linguistic use of common ground but rather to "task-based strategic effects of common ground mixed with automatic *egocentric* effects" (Barr & Keysar, 2007:918). Even if participants might fixate more frequently on the co-present objects in the grid (see Keysar et al., 2000), Barr and Keysar's criticism does not explain why they have not

---

[7] Note that Keysar (2007) refers to an unpublished study presented at the *2005 Meeting of the Psychonimic Society* where adult participants were also presented with two similar objects – among others, one of them in common ground and the other in privileged view. Surprisingly, when asked to move the truck, for example, participants in this study "tended to ask *Which truck?*" (Keysar, 2007:74). Unfortunately, Keysar does not report details of the statistical significance of this tendency, so it is difficult to evaluate this result and compare it with the conflicting evidence observed by Nadig and Sedivy (2002) when testing children.

included this second measure of egocentricity in their studies, given that a baseline difference between co-present and privileged objects could always be corrected before doing further statistical analyses. In this sense, Nadig and Sedivy (2002) provide a critical analysis of their eye-tracking data which is missing in the studies by Keysar and collaborators: namely, a plot of the participants' fixations on the different objects in the grid during the processing of the instructions[8].

Regarding a potential baseline difference between co-present and privileged objects, Nadig and Sedivy (2002) used a baseline condition similar to the one in Keysar et al.'s studies, which should balance out any possible preference for common-view objects in the critical condition. When participants were asked to "pick up the glass", they were presented with two glasses in the critical condition – one in common view and another in the participant's privileged view, among other objects. In the baseline condition, they were presented with one glass in common view and an unrelated object in the occluded spot. If participants were generally more likely to fixate on the glass in common view than on the occluded competitor, this would have also been the case in the baseline condition. Therefore, a differential advantage of the glass in common view over the occluded object between these two conditions should be a reliable measure of egocentricity. Moreover, Nadig and Sedivy (2002) used a second control condition where participants were presented with two glasses in common view, thus rendering the instruction "Pick up the glass" ambiguous. According to their results, in the earliest time window (i.e. 200-760ms after the onset of the noun), when an automatic moment of egocentrism should have been observed according to Keysar et al., the advantage of the intended glass was not significantly different in the critical and the baseline conditions, although a significant difference was observed with the ambiguous condition.

I should emphasize, however, that when Nadig and Sedivy (2002) used the same measure of egocentricity as Keysar et al. and looked only at participants' fixations on the target glass in the critical and control conditions, they did not find evidence of common-ground use during the earliest stages of processing (see reference for discussion). A comparison between gaze fixations on the intended targets and the hidden competitors is therefore crucial for the investigation of perspective taking. Going back to Keysar et al.'s studies, consider again *Figure 1* and the instructions "Put the (bottom) block below the apple" (Keysar et al., 1998b, 2000, 2003; Wu & Keysar, 2007b): even if participants might have been *slower* at fixating the intended block in

---

[8] Unlike in their other studies, Epley et al. (2004b) used a baseline measure only to control for overall differences in fixation latencies between children and adults. Their analysis of the critical trials therefore compared the relative speed with which participants fixated on the intended and hidden objects. However, because this study used the egocentric-favouring paradigm (i.e. "Put the *bottom* block above the apple"), it is not possible to make a fair comparison between their results and those of Nadig and Sedivy (2002).

the middle of the grid when there was an occluded block at the bottom of the grid than when there was an occluded toy monkey, as long as they fixated *more frequently* on the intended block in common ground than on the occluded block during the earliest stages of processing the instructions, it could be argued that participants were using common-ground information rather than being egocentric[9].

Therefore, in order to conclude that participants were momentarily egocentric, Keysar and collaborators should have made a second comparison between participants' fixations on the intended block in the middle of the grid and the occluded block at the bottom – correcting for any possible general preference for objects in common ground. In this sense, it is still possible that their participants might have shown a preference for the intended block over the competitor during the early stages of interpretation. Such mixed results could be taken to support the view that the egocentric and the allocentric perspectives compete in parallel – rather than serially, so participants would be "egocentric enough" to show a preference for the occluded block over the occluded toy monkey, but "allocentric enough" to prefer the common-view block over the occluded block. In conclusion, as informative as it might be in its own right, evidence of 'the egocentric error' with respect to an unrelated occluded object (Keysar, 2007:74) needs to be matched with a relative measure of gaze-fixation between the intended target and the occluded competitor if a complete evaluation of the participants' egocentricity is to be made. In this sense, Keysar et al.'s analysis of their eye-tracking data might not be comprehensive enough to warrant the general conclusions that they draw from their results.

3.3 Spontaneous vs. deliberate perspective-taking

In defence of the egocentric-favouring paradigm used in their eye-tracking experiments and against the unbiased paradigm used by Nadig and Sedivy (2002), Keysar and collaborators argue that, if the hidden and the intended objects were equally good referents for the target noun (as it was the case in Nadig & Sedivy (2002)), the instructions would become ambiguous and their hypothesis about the limits of perspective taking could not be tested. "To resolve such ambiguity [between two similar objects, one in common view and one occluded from the speaker] participants would have been forced to employ their theory of mind, thus obscuring the phenomenon we are attempting to uncover" (Keysar et al., 2003:31; see also Epley

---

[9] It might initially seem obvious that an occluded block would cause more interference than an occluded toy monkey in processing the referential expression "the bottom block". However, the rationale of this measure of egocentricity is that, if participants are taking into account the speaker's perspective, they should not consider *any* of the occluded objects when processing the instructions, so no difference should be observed between the occluded block and the occluded toy monkey.

et al. (2004b) for a similar argument). If this had been the case and the children in Nadig and Sedivy (2002) had initially perceived the instruction "Pick up the glass" as ambiguous, one might have expected them to show more hesitation between the intended glass and the occluded competitor during the earliest stages of processing. However, their participants were as fast in their use of common ground as to show a preference for the intended glass right from the offset of the noun in the instructions, "indicating that fixations programmed prior to the end of that word were [already] influenced by common-ground information" (Nadig & Sedivy, 2002:334)[10]. This evidence contradicts the claim that monitoring for common ground in language processing is a *slow* and *effortful* process (Epley et al., 2004b; Barr & Keysar, 2005, 2007). However, Keysar et al.'s further claim that common-ground monitoring is *optional* remains an empirical question.

Despite the strong evidence of effective perspective-taking offered by Nadig and Sedivy (2002), Keysar's counterargument is legitimate in terms of a valid test for their model: given the "egocentric ambiguity" in their instructions, the experimental design used by Nadig and Sedivy does not allow determining whether their participants employed their theory of mind spontaneously or only deliberately. It is interesting to note, however, that the criticism that Keysar et al. (2003) raise against Nadig and Sedivy (2002) also applies to Wu and Keysar (2007b), given that Wu and Keysar used the same unbiased, potentially ambiguous paradigm as Nadig and Sedivy. Crucially, in the case of Wu and Keysar (2007b), Keysar et al.'s counterargument invalidates their main claim that the Chinese participants in their study were able to use their theory of mind spontaneously while American participants did so reflectively. At most, Chinese participants would have been faster at fixating on the intended target, but whether they used their perspective-taking abilities spontaneously or only to disambiguate the instructions is an open question.

3.4 Possible confounding variables

One last critical issue in evaluating the results and conclusions of Wu and Keysar (2007b) is the possibility that the two populations tested in their study might have differed not only in their cultural background but also in some other relevant ability or experience that could potentially challenge Wu and Keysar's attribution of significant cognitive effects to cultural differences. Given the nature of the referential communication task used by Keysar and his collaborators, the good performance of the Chinese participants in Wu and Keysar (2007b) might have had to do with factors other than their theory of mind abilities: for example, selective attention, inhibition of competing information, or more generally, executive control (see Neill, Valdes &

---

[10] Note that the approximate time to launch a saccade is 200ms.

Terry, 1995; Perner & Lang, 1999). If this were the case, a question would remain as to why the Chinese participants would have outperformed the Americans.

Research on bilingualism and executive control might suggest a possible answer to this question: over more than a decade of experimental research, Bialystok and her collaborators have found strong empirical evidence that bilingual individuals – both children and adults, have more effective controlled processing than monolinguals (Bialystok 2001, 2006; Bialystok, Craik, Klein & Viswanathan, 2004; Bialystok, Craik & Ryan, 2006). These studies used executive control tasks that involve focusing on one type of stimuli while inhibiting a competing response. The referential communication game used by Keysar et al. could be understood as an executive control task where participants have to select objects from the common-view slots in the grid while ignoring potential targets in the privileged-view slots.

Bialystok and her colleagues attribute the bilinguals' advantage in executive control to their need to simultaneously monitor two language systems in order to prevent interference from the unwanted language. The description that Wu and Keysar (2007b) make of their participants does not allow properly evaluating their linguistic background. However, they specify that the American participants did not have an Asian background and were native speakers of American English studying in an American university. In contrast, the Chinese participants were native speakers of Mandarin who had been born and raised in mainland China and had been living in the United States from 2 to 9 months. Their level of English was not specified in the technical report of the experiment but they must have been fluent enough to be studying at the same university as the American participants.

Obviously, the above description does not show that the Chinese participants were Mandarin-English bilinguals, just as it does not preclude the American participants from being bilingual in some other language. Nonetheless, the linguistic immersion in an English-speaking environment that the Chinese participants would have experienced when moving to the United States and joining an American university might have helped them improve their controlled processing compared to other American students (see Bialystok, Craik & Ruocco, 2006). This advantage might have in turn improved their performance on the referential communication task, helping them focus their attention on the common-ground objects while ignoring the hidden slots in the grid.

Research on collectivistic and individualist cultures also suggests relevant differences in attentional processes between these two groups: using a variety of tasks, Nisbett and collaborators have observed that East Asians are significantly more likely to allocate their attention to contextual information than North Americans, who focus more rapidly and longer on central objects (Masuda & Nisbett, 2001, 2006; Chua, Boland & Nisbett, 2005). In the referential communication task used by Wu and Keysar (2007b), the greater sensitivity to background information characteristic of

East Asians might have helped Chinese participants to distinguish more rapidly between the open cells and the occluded cells in the grid and thus focus their attention more effectively on the shared-view objects. In contrast, American participants might have generally focused their attention on the objects in the grid, showing less of a preference for objects in shared view. As Masuda and Nisbett conclude, "Westerners are relatively more likely to see objects, whereas Easterners are relatively more likely to see contexts" (2006:394).

In conclusion, even if research on bilingualism and attention might not necessarily explain the good performance of Chinese participants in Wu and Keysar (2007b), the results observed by Bialystock et al. and Nisbett et al. suggest the possibility that confounding variables might have led Wu and Keysar to overinterpret their results. An evaluation of the cognitive demands of the referential communication task as well as of the linguistic and attentional abilities of the participants in the study might therefore be necessary before concluding that Chinese people are able to use their perspective-taking abilities spontaneously in communication, while American people can do so only reflectively.

4 Conflicting empirical evidence

The Egocentric Anchoring and Adjustment model is challenged by various eye-tracking studies that have observed fast and effective use of common-ground information in language interpretation with experimental designs that did not favour the egocentric response (e.g. Hanna & Tanenhaus, 2004; Tanenhaus & Brown-Schmidt, 2008). Hanna, Tanenhaus and Trueswell (2003), for example, observed that participants were fast at identifying an intended target (e.g. "Pick up the empty martini glass") even when their perspective was different from that of the confederate giving the instructions (i.e. participants knew that the confederate, who was behind a blind, had been misinformed about the contents of their display of objects). Thus, even when their display included *two* empty martini glasses, participants were able to quickly identify "the empty martini glass" if they knew that the speaker thought that one of their glasses was actually full.

In evaluating these results, Keysar and his collaborators could argue that participants in this condition were also 'forced' to take the confederate's perspective in order to unambiguously interpret the instructions and choose one of the two empty martini glasses in their display. As in the case of Nadig and Sedivy (2002), if the participants in Hanna et al. (2003) had initially perceived the instructions as ambiguous and monitored for the speaker's perspective in a later phase of processing, one would expect that these participants would have been relatively slow at selecting the intended martini glass. However, the participants in this study were actually *faster* to interpret

the instruction "Pick up the empty martini glass" when they could use their knowledge of the confederate's mistaken perspective to choose between two empty martini glasses than when they shared perspective with the confederate but the description in the instructions was applicable to two objects (e.g. when there was an empty martini glass and an empty jar in the display). This pattern of results therefore contradicts Keysar et al.'s claim that monitoring for the speaker's perspective is a slow, effortful process that takes place in a late phase of language processing (Keysar et al., 1998a; Epley et al., 2004b; Barr & Keysar, 2005, 2007). More importantly, these results challenge Keysar et al.'s claim that the theory of mind ability to distinguish between one's beliefs and those of another is not fully integrated into the language processing system (Keysar et al., 2003; Barr & Keysar, 2007): if this were the case, the participants in Hanna et al. (2003) should not have been able to use their knowledge of the confederate's false belief as an on-line cue to unambiguously interpret her instructions[11].

Hanna and collaborators conclude from their results that "while the referential domain is not completely restricted to objects in common ground, common ground does have an immediate effect on reference resolution" (2003:50). According to Keysar et al., restricting common ground use to a late, optional phase of interpretation would be advantageous in terms of reducing processing load, given that interpreters would not need to constantly monitor for the speaker's perspective and intentions (see Keysar et al. (1998b, 2003), Epley et al. (2004b) and Keysar (2007) for discussion). Hannah et al. (2003), on the other hand, endorse a constraint-based account of language processing where common ground is a type of contextual constraint that has *immediate* and *probabilistic* effects on interpretation, with these effects varying depending on how salient and relevant the speaker's perspective is to the addressee[12]. Thus, in a co-operative game where participants had to help a cook follow a recipe and

---

[11] The results of Hanna et al.'s experiment are also important because they do not involve a comparison between common-view and privileged-view objects in a grid. In this sense, their second experiment escapes Barr and Keysar's criticism that the results of Nadig and Sedivy (2002) and Hanna et al. (2003) might reflect non-linguistic, task-based strategic effects of common ground resulting from an overall preference for objects in common view over hidden objects (2007:918). Further empirical evidence against Keysar et al.'s model, which cannot be reduced to 'different baseline probabilities of fixation', comes from Hanna and Tanenhaus (2004).

[12] I want to point out that Hanna et al. (2003) tried to reconcile their results with those of Keysar et al. (2000) by arguing for a *moderate* version of the Perspective Adjustment model in which egocentricity was understood as a tendency rather than a complete insensitivity to common ground (see Keysar et al., 2000; Footnote 1). However, rather than adopting a more moderate position in view of the conflicting empirical evidence presented by Nadig and Sedivy (2002), Hanna et al. (2003) and Hanna and Tanenhaus (2004), Keysar, Barr and collaborators have developed their model of perspective taking in a more radical direction in recent years (see Keysar et al., 2003; Barr, 2004; Epley et al., 2004a, 2004b; Barr & Keysar, 2005; Wu & Keysar, 2007b).

so his perspective was both salient and relevant to them, Hanna and Tanenhaus (2004) observed that participants continuously modified the referential domain of interpretation according to the cook's abilities and intentions.

Following on the above quote, it must be noted that Hanna et al. (2003) did find evidence of egocentric language processing in their study: in their first experiment, for example, when their participants were instructed to "put the blue triangle on the red one", they considered a blue triangle in their privileged view more often than other irrelevant figures in common ground. In view of this pattern of results, Hanna and her collaborators argue that people cannot ignore perceptually salient objects outside their common ground with a speaker when these objects fit a referential description in the speaker's instructions (2003:59). This is a crucial finding for perspective-taking research, both for its theoretical and methodological implications: common ground does not completely delimit the referential domain of a referring expression, with highly-accessible potential candidates in the addressee's privileged ground being considered for reference assignment.

Note that this pattern of results confirms, with an unbiased test, the Unrestricted Search Hypothesis of Keysar et al. (1998a) – at least in an interactive setting involving physical objects as referents. As well as supporting the earliest, more moderate claims of Keysar and his collaborators regarding reference resolution, this empirical evidence could be taken to support a *general* formulation of the egocentric bias in language processing. However, the critical test for Keysar et al.'s two-stage model of perspective taking as egocentric anchoring and adjustment is not whether participants experienced *any* interference from the competitor target in their privileged view, but whether this interference resulted in an initial, automatic moment of egocentrism that was subsequently corrected by taking into account the speaker's perspective (Keysar et al., 1998b, 2003; Keysar, 2007).

Contrary to this prediction, Hanna et al. (2003) observed that even from the earliest stages of processing, participants showed a preference for the intended target in common ground over the privileged competitor. That is, in processing the instructions "Put the blue triangle on the red one", participants started to fixate their gaze more frequently on the blue triangle in common ground than on the blue triangle in their privileged view within 400ms after the onset of the colour word. In view of these results and the conflicting pattern observed by Keysar and his collaborators with the egocentric-favouring paradigm, it is possible to conclude that, in order for participants to disregard common ground and show an *initial preference* for an object in their privileged view as a referential candidate, this object should be not simply a *possible* referent for the description (e.g. a blue triangle for "the blue triangle"), but actually the *best* possible candidate available (e.g. the darkest blue triangle for "the dark blue triangle"). It seems clear that the latter case does not extend to all instances of

reference assignment in language processing, with the processes of egocentric anchoring and correction being therefore limited to very specific contexts.

My point in contrasting these studies is that it should not be necessary to bias an automatic process in order to observe an interference with a controlled process: creating a balanced tension between the two should suffice. If egocentric anchoring was indeed an automatic process and perspective taking worked only as a controlled correction mechanism, the unbiased designs used by Hanna, Tanenhaus, and their collaborators should have shown evidence of that particular time-course of perspective taking. However, even when their participants showed interference from potential referents in privileged ground, Hanna and collaborators "found no evidence for an initial stage of processing where addressees ignored information from common ground" (2003:59).

Let me illustrate this point with an analogy with the Stroop effect – probably the best documented automatic processes in psycholinguistics to date (Stroop, 1935). In this task, participants have to name the colour of the ink of a series of words. In critical trials, the actual word is a colour word, which is different from the colour of the ink in which it is printed (e.g. 'blue' printed in red). In such trials, participants experience interference between reading the colour word – which is an automatic process, and naming the colour of the ink – which is a controlled process. Analogously to the egocentric-favouring paradigm used by Keysar, Barr and collaborators, the automatic response in the Stroop task would probably be favoured if a large font was used for the colour words, while they were printed in a faint, undistinguishable ink colour (see Alter, Oppenheimer, Epley & Eyre, 2007). However, the interference of an automatic process in a controlled process has been observed with the Stroop task without having to resort to that type of manipulation. I therefore want to conclude that the empirical evidence provided by Keysar, Barr and collaborators in support of their two-stage model of perspective taking leaves open the question of whether the processes of egocentric anchoring and adjustment characterise all instances of perspective taking in language processing – as Keysar and colleagues claim (e.g. Keysar et al., 1998b, 2003; Epley at al., 2004b; Barr & Keysar, 2002, 2005), or only those situations where the egocentric response is reinforced and therefore needs to be corrected[13].

---

[13] The Egocentric Anchoring and Adjustment model is based on a clear distinction between automatic and controlled processes: "As with other dual-process accounts of human judgement, the automatic default occurs quickly and rapidly whereas the corrective process must be activated by motivation and sustained by attention" (Epley et al., 2004b:761). It is therefore puzzling that Epley and his colleagues suggest that these processes may not apply in interaction with family or close friends whose perspectives are well known to us and need not be inferred (2004a:328). One would assume that, by virtue of being automatic, the process of egocentric anchoring would not be sensitive to how well we might know our addressees. Given that the focus of Epley et al. (2004a) was on people's judgements of others' perceptions, one possibility is that they might have used the term

5 An open answer

Contrary to the predictions of the Egocentric Anchoring and Adjustment model, Tanenhaus and collaborators have shown that common-ground information can be used early on in the language interpretation process rather than being limited to a late monitoring phase. However, these studies have also shown that common ground does not completely restrict the referential domain of a referring expression, with participants considering potential referents outside their common ground with the speaker. This latter pattern of result leaves open the question of whether our perspective-taking abilities are fully integrated into our language interpretation system, or whether as Keysar and collaborators propose, our ability to distinguish our own perspective from that of another can only be applied reflectively in communication (Keysar et al., 2003).

   In the setting of a collaborative task involving physical interaction with objects, optimal monitoring for the speaker's perspective might involve constraining the search for potential referents to those objects that are visible to both the speaker and the addressee. Although this simplistic notion of common ground might be valid for the rather restrictive conditions of the referential communication games used in perspective-taking experiments, it is clearly too narrow for everyday communication. Imagine, for example, that John and Mary are cooking dinner together. If Mary points to a drawer and asks John to give her the scissors, she might not be able to *see* the contents of the drawer, but this should not prevent John from understanding that she *thinks* that the scissors are in the drawer. Given that most referring expressions are normally used in the absence of physical referents, the importance of bearing in mind the speaker's beliefs and intentions in interpreting a referring expression is comparable to the importance of monitoring for the speaker's visual perspective. This importance is illustrated by the fact that, even in the simple setting of a referential communication game, Keysar et al. went to great lengths to make sure that their participants did not suspect that the confederate giving the instructions was actually aware of the contents of the occluded cells in the grid: if their participants knew that the confederate knew, their interpretation of her instructions would change completely (e.g. "the bottom block" could be taken to refer to the lowest block in the grid, even if it was hidden from the confederate). Intuitively, it seems obvious that monitoring for a speaker's visual perspective is a much simpler task than monitoring for what she might know or have in mind. In this sense, optimal use of our theory of mind abilities might still be insufficient to read somebody else's mind.

'perspective' more generally to include views and opinions. In any case, Epley et al. (2004a) motivate their experiments and discuss their results in line with Keysar's model of perspective taking.

The real test for theory of mind use in language processing probably arises in those situations where we know that the speaker is mistaken (i.e. the real-life equivalent of a false-belief task; see Baron-Cohen, Leslie & Frith, 1985). Imagine, for example, that John knows that the scissors that Mary wants are on the kitchen table and not in the drawer where she thinks they are. The question here would be whether John would look first inside the drawer that Mary is pointing at or go directly to the table. We can probably tell from our everyday experience that the answer to this question might simply depend on how alert John is at the time of processing Mary's request: if he is somewhat absentminded, he might need to look inside the drawer and realize that the scissors are not there before he remembers that he had actually left the scissors on the kitchen table. On the other hand, if this piece of information was highly accessible at the time of processing Mary's utterance, John might go directly to the kitchen table and take the scissors for Mary without even opening the drawer.

It is intriguing, however, that most participants in Keysar et al. (2003) made an egocentric error and considered an object inside a paper bag as a possible referent for a definite description when they had been told that the confederate giving the instructions was not only *ignorant* about the contents of the bag but actually *mistaken*. One would expect – or at least hope, that participants in an experiment would not be absentmindedly thinking about other things while being tested, the way John might do while cooking dinner with Mary. The question therefore remains as to why the participants in Keysar et al. (2003) would have made the egocentric error when they were fully aware of the confederate's false belief (cf. Hanna et al., 2003; Experiment 2). One possibility is that by putting one of the objects in the grid inside a paper bag and making the participant believe that the confederate giving the instructions was misinformed about the contents of the bag, the actual contents might become salient to the participant – especially if they corresponded with the object that best fit the description in the instructions (e.g. the lowest block in the grid). In the scenario where Mary asks John to give her the scissors in the drawer, a similar situation might arise if John had just put a pair of scissors inside the drawer: even though John may be aware that those are not the scissors that Mary wants (e.g. if she was cleaning fish and needed the appropriate scissors), having just put a pair of scissors in the drawer might make John more prone to look inside the drawer when interpreting Mary's request.

Nonetheless, even if the contents of the paper bag might have been salient for the participants in Keysar et al. (2003), one might still expect that if these participants were making optimal use of their theory of mind abilities, they would simply disregard the object inside the paper bag as a possible referent (the same way that John should disregard the small scissors inside the drawer) on the grounds that the mistaken confederate could not be referring to this object in his instructions. However, even in a simple referential communication game, it seems reasonable to consider to what extent it would be worth limiting the referential domain of a referring expression to those

objects that we *think* the speaker can see or has knowledge of. After all, just as a speaker might know more than what she can see, she might also know more than what the addressee thinks – or for that matter, have an intention that he could not foresee[14]. Evidence comes again from the actual perspective-taking experiments where the confederate giving the instructions was aware of the contents of the occluded cells, contrary to what the participants thought. In this sense, our knowledge of other people's beliefs and intentions is inevitably *tentative*. Therefore, monitoring for the speaker's beliefs to the extent of *blocking* all accessible potential referents that could not be intended by the speaker according to our tentative representation of her beliefs might require a great expense of processing resources for no certain gain. After all, Mary could be referring to the small scissors inside the drawer, contrary to what John first thought.

It seems undeniable that the inferential nature of linguistic communication leaves ample room for miscommunication. Even to the best of our knowledge and our mind-reading abilities, monitoring for a speaker's perspective involves an element of speculation. This uncertainty might not always warrant that we narrow down interpretation to a single reading from the earliest stages of processing, but rather that we construct the intended interpretation as part of a complex, ongoing process where different processing factors and levels of meaning converge. The results of Nadig and Sedivy (2002) and Hanna et al. (2003), among others, show that interpreters are sensitive to the speaker's beliefs and intentions from the earliest stages of processing, even though more "egocentric" interpretations might be processed in parallel with the intended one. Contrary to the conclusions of Keysar et al. (2003), who explain the underperformance of their participants in terms of 'limits of theory of mind use', I want to suggest an alternative interpretation of the partial egocentric bias observed in the common-ground literature: the fact that an addressee might not immediately discard possible interpretations that are less likely to be intended by the speaker might be the reflection of the inferential nature of communication in general, and the tentative nature of interpretation in particular, rather than empirical evidence that important elements of theory of mind are not fully incorporated into the language processing system[15].

---

[14] One such possibility would arise if the confederate in Keysar et al. (2003) used the referring expressions in her instructions *attributively* rather than *referentially* (Donnellan, 1966): given this distinction, a definite description such as "the bottom block" could be used to refer to the absolute lowest block in the grid (regardless of whether it was visible or hidden from the confederate), rather than to a particular block (which would be the lowest one from the confederate's perspective).

[15] Another way of looking at this issue is to see the limitation, not in the integration of theory of mind abilities in the comprehension system, but in the theory of mind abilities *per se*: given that as addressees we normally have to infer the speaker's beliefs and intentions on the basis of rather limited evidence, in order to be *certain* about a speaker's mental state and systematically use this certainty as

Other bottom-up processes that have been documented in the psycholinguistics literature and which might be comparable to the accessibility of unintended salient candidates in reference resolution is the spread activation of the various meanings of a homonym – even in contexts where the intended meaning is clear (Swinney, 1979; Simpson, 1981), or the accessibility of the literal interpretation of a metaphoric expression (Keysar, 1994a; Rubio-Fernández, 2007). More intuitive evidence of the openendedness of the language interpretation process comes from our ability to play with words and appreciate irony and double entendres. It therefore seems possible to conclude that accessing an alternative, less likely interpretation of a linguistic expression may not necessarily reveal an egocentric default, or even a dissociation between theory of mind and interpretation. On the contrary, accessing parallel interpretations – mostly at the local level of constituents but sometimes even at the global level of the utterance, might simply be part of a cost-effective comprehension procedure aiming at the interpretation that was intended by the speaker; which is rarely unequivocal given how linguistic meaning falls short of determining what is communicated (see Carston, 2002).

6 Summary and Conclusions

I started this review paper by distinguishing general views and empirical evidence of an egocentric bias in human cognition (e.g. Birch & Bloom, 2007; Birch & Bernstein, 2007) from the particular predictions of the Egocentric Anchoring and Adjustment model of Keysar, Barr and collaborators regarding the *time course* of perspective taking. According to Keysar et al., perspective taking operates as a two-stage process where our egocentric perspective serves as an automatic default that needs to be corrected by a sequential, reflective mechanism of perspective adjustment (Epley et al., 2004a, 2004b). In the case of verbal communication – which has been the focus of this paper, language would usually be interpreted egocentrically, with common ground playing a role in language processing only in a later, monitoring phase of interpretation, which is characterised as being *optional* and *slow* (Barr & Keysar, 2005, 2007).

Keysar, Barr and their colleagues have provided robust empirical evidence of egocentric anchoring and perspective adjustment over more than a decade of researching common ground. However, in a critical review of their experimental work on perspective taking, I have pointed out a systematic design feature which biases

---

a reliable filter in the language interpretation process, we would probably need to be able to mind-read the speaker in the paranormal sense of the word. On a more positive note, it is remarkable how far our mind-reading capacity can get us in communication to be "just normal".

participants towards the egocentric response. Moreover, their use of a single measure of egocentricity makes their analysis of their eye-tracking data insufficient, at least for the broad conclusions that they draw from their results. Given these potential confounds, I argue that Keysar et al. may have constructed a theoretical model of perspective taking in language processing on the basis of an *overgeneralisation* of their empirical findings, which goes well beyond the specific experimental paradigm used in their studies. Far from limiting their conclusions to situations where the egocentric perspective might be highly salient and therefore require conscious correction, Keysar, Barr and collaborators have extended their theoretical claims to all instances of perspective taking in human interaction and communication, going as far as to argue that important elements of our theory of mind are not fully integrated into our language processing system (Keysar et al., 2003; Barr & Keysar, 2005; Wu & Keysar, 2007b)[16].

Contrary to this radical claim, other studies using unbiased tasks have provided clear empirical evidence that common ground distinctions are applied fast and effectively in communication (e.g. Nadig & Sedivy, 2002; Hanna, Tanenhaus & Trueswell, 2003; Hanna & Tanenhaus, 2004). The egocentric-ambiguous paradigm used in these studies only leaves open the question of whether people are able to apply their perspective-taking abilities optimally and spontaneously in all situations, or only in those contexts where they need to do so in order to disambiguate a linguistic expression. At any rate, I argue that evidence of a partial egocentric bias in language processing does not need to result from an insufficient integration of theory of mind into the language interpretation system, but rather from the cost-effective operation of this system, which responds to the natural indeterminacy of linguistic expressions.

To conclude, in their steady attempt to 'unconfound common ground' (Keysar, 1997), Keysar, Barr and their colleagues may have confounded the most salient response in their experiments with an automatic egocentric process, and the lesser favoured response with a controlled correction mechanism.

References

Alter, A., Oppenheimer, D., Epley, N. & Eyre, R. (2007). Overcoming intuition: Metacognitive difficulty activates analytical thought. Journal of Experimental Psychology: General, 136, 569-576.

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

---

[16] In recent years, the scope of Keysar et al.'s research has been broadened even further with computer simulations supporting their claim that conventional communication systems may be established by joint action, without the members of the linguistic community resorting to their mutual beliefs (see Barr, 2004; Keysar & Barr, 2005).

Barr, D. J. (2004). Establishing conventional communication systems: Is common knowledge necessary? Cognitive Science: A Multidisciplinary Journal, 28, 937-962.

Barr, D. J. & Keysar, B. (2002). Anchoring comprehension in linguistic precedents. Journal of Memory and Language, 46, 391-418.

Barr, D. J. & Keysar, B. (2005). Mindreading in an exotic case: The normal adult human. In B. F. Malle & S. D. Hodges. (Eds.), Other minds: How humans bridge the divide between self and other. New York: Guilford Press, pp. 271-283.

Barr, D. J. & Keysar, B. (2007). Perspective taking and the coordination of meaning in language use. In M. J. Traxler & M. A. Gernsbacher (Eds.), Handbook of Psycholinguistics, 2nd Edition. San Diego: Academic Press, pp. 901-938.

Bialystok, E. (2001). Bilingualism in development: Language, literacy, and cognition. New York: Cambridge University Press.

Bialystok, E. (2006). Effect of bilingualism and computer video game experience on the Simon task. Canadian Journal of Experimental Psychology, 60, 68–79.

Bialystok, E., Craik, F. I., Klein, R. & Viswanathan, M. (2004). Bilingualism, aging, and cognitive control: Evidence from the Simon task. Psychology and Aging, 19, 290-303.

Bialystok, E., Craik, F. I. & Ruocco, A. C. (2006). Dual-modality monitoring in a classification task: The effects of bilingualism and ageing. The Quarterly Journal of Experimental Psychology, 59, 1968–1983.

Bialystok, E., Craik, F.I. & Ryan, J. (2006). Executive control in a modified antisaccade task: Effects of aging and bilingualism, Journal of Experimental Psychology: Learning, Memory and Cognition, 32, 1341–1354.

Birch, S. & Bernstein, D. (2007). What kids can tell us about hindsight bias: A fundamental constraint on perspective-taking? Social Cognition, 25, 98-113.

Birch, S. & Bloom, P. (2004). Understanding children's and adult's limitations in mental state reasoning. Trends in Cognitive Science, 8, 255-260.

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

Brennan, S. E. & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. Journal of Experimental Psychology: Learning, Memory, and Cognition, 22, 1482-1493.

Carston, R. (2002). Thoughts and utterances. The pragmatics of explicit communication. Oxford: Blackwell.

Chua, H. F., Boland, J. & Nisbett, R. E. (2005). Attention to object vs. background: Eye-tracking evidence comparing Chinese and Americans. Poster presented at the Annual Meeting of the Society of Personality and Social Psychology, New Orleans, LA.

Clark, H. H. & Marshall, C. R. (1981). Definite reference and mutual knowledge. In A. K. Joshi, B. L. Webber & I. A. Sag (Eds.), Elements of discourse understanding. Cambridge, England: Cambridge University Press, pp. 10-63.

Donnellan, K. S. (1966). Reference and definite descriptions. Philosophical Review, 75, 281-304.

Epley, N., Keysar, B., Van Boven, L. & Gilovich, T. (2004a). Perspective taking as egocentric anchoring and adjustment. Journal of Personality and Social Psychology, 87, 327-339.

Epley, N., Morewedge, C. K. & Keysar, B. (2004b). Perspective taking in children and adults: Equivalent egocentrism but differential correction. Journal of Experimental Social Psychology, 40, 760-768.

Gerrig, R. J., Ohaeri, J. O. & Brennan, S. E. (2000). Illusory transparency revisited. Discourse Processes, 29, 137-159.

Glucksberg, S. (1989). Metaphors in conversation: How are they understood? Why are they used? Metaphor and Symbolic Activity, 4, 125-143.

Glucksberg, S. (2001). On the fate(s) of literal meanings. Paper presented at the Workshop towards an Experimental Pragmatics, Lyon, France.

Glucksberg, S. & Keysar, B. (1990). Understanding metaphorical comparisons: Beyond similarity. Psychological Review, 97, 3-18.

Glucksberg, S. & Keysar, B. (1993). How metaphors work. In A. Ortony (Ed.), Metaphor and Thought. Cambridge, England: Cambridge University Press.

Glucksberg, S., Kraus, R. & Higgins, E. T. (1975). The development of referential communication skills. In F. D. Horowitz. (Ed.), Review of Child Development Research. Chicago: Chicago University Press, pp. 305-345.

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. Morgan (Eds.), Syntax and Semantics III: Speech Acts. London: Academic Press, pp. 41-58.

Grice, H. P. (1989). Studies in the way of words. Cambridge, Mass: Harvard University Press.

Hanna, J. & Tanenhaus, M. (2004). Pragmatic effects on reference resolution in a collaborative task: Evidence from eye movements. Cognitive Science 28, 105-115.

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

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

Keysar, B. (1989). On the functional equivalence of literal and metaphorical interpretations in discourse. Journal of Memory and Language, 28, 375-385.

Keysar, B. (1994a). Discourse context effects: Metaphorical and literal interpretations. Discourse Processes, 18, 247-269.

Keysar, B. (1994b). The illusory transparency of intention: Linguistic perspective taking in text. Cognitive Psychology, 26, 165-208.

Keysar, B. (1997). Unconfounding common ground. Discourse Processes, 24, 253-270.

Keysar, B. (2000). The illusory transparency of intention: Does June understand what Mark means because he means it? Discourse Processes, 29, 161-172.

Keysar, B. (2007). Communication and miscommunication: The role of egocentric processes. Intercultural Pragmatics, 4, 71-84.

Keysar, B. & Barr, D. J. (2005). Coordination of action and belief in communication. In J. C. Trueswell & M. K. Tanenhaus (Eds.), Approaches to Studying World Situated Language Use: Bridging the Language-as-Product and Language-as-Action. Cambridge, Mass: MIT Press, pp. 73-94.

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

Keysar, B., Barr, D. J., Balin, J. A. & Paek, T. S. (1998a). Definite reference and mutual knowledge: Process models of common ground in comprehension. Journal of Memory and Language, 39, 1-20.

Keysar, B., Barr, D. J. & Horton, W. S. (1998b). The egocentric basis of language use: Insights from a processing approach. Current Directions in Psychological Sciences, 7, 46-50.

Keysar, B., Ginzel, L. E. & Bazerman, M. H. (1995). States of affairs and states of mind: The effect of knowledge about beliefs. Organizational Behavior and Human Decision Processes, 64, 283-293.

Keysar, B. & Henly, A. S. (2002). Speakers' overestimation of their effectiveness. Psychological Science, 13, 207-212.

Keysar, B. & Horton, S. W. (1998). Speaking with common ground: From principles to processes in pragmatics. A reply to Polichak and Gerrig. Cognition, 66, 191-198.

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

Krauss, R. M. & Glucksberg, S. (1977). Social and nonsocial speech. Scientific American, 236, 100-105.

Masuda, T. & Nisbett, R. E. (2001). Attending holistically versus analytically: Comparing the context sensitivity of Japanese and Americans. Journal of Personality and Social Psychology, 81, 922–934.

Masuda, T. & Nisbett, R. E. (2006). Culture and change blindness. Cognitive Science, 30, 381-399.

Metzing, C. & Brennan, S. E. (2003). When conceptual pacts are broken: Partner specific effects in the comprehension of referring expressions. Journal of Memory & Language, 49, 201-213.

Nadig, A. & Sedivy, J. (2002). Evidence of perspective-taking constraints in children's on-line reference resolution. Psychological Science, 13, 329-336.

Neill, W. T., Valdes, L. A. & Terry, K. M. (1995). Selective attention and the inhibitory control of attention. In F. N. Dempster & C. J. Brainerd (Eds.), Interference and inhibition in cognition. San Diego, CA: Academic Press, pp. 207-261.

Perner, J. & Lang, B. (1999). Development of Theory of Mind and executive control. Trends in Cognitive Sciences, 3, 337-344.

Polichak, J. W. & Gerrig, R. J. (1998). Common ground and everyday language use: Comments on Horton and Keysar (1996). Cognition, 66, 183-189.

Rubio-Fernández, P. (2007). Suppression in metaphor interpretation: Differences between meaning selection and meaning construction. Journal of Semantics, 24, 345-371.

Shintel, H. & Keysar, B. (2007). You said it before and you'll say it again: Expectations of consistency in communication. Journal of Experimental Psychology: Learning, Memory & Cognition, 33, 357-369.

Simpson, G. (1981). Meaning dominance and semantic context in the processing of lexical ambiguity. Journal of Verbal Learning and Verbal Behavior, 20, 120-136.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. Journal of Experimental Psychology, 18, 643-662.

Swinney, D. A. (1979). Lexical access during sentence comprehension: (Re)consideration of context effects. Journal of Verbal Learning and Verbal Behavior, 18, 645-659.

Tanenhaus, M. & Brown-Schmidt, S. (2008). Language processing in the natural world. Philosophical Transactions of the Royal Society B, 363, 1105–1122.

Wu, S. & Keysar, B. (2007a). The effect of communication overlap on communication effectiveness. Cognitive Science, 31, 169-181.

Wu, S. & Keysar, B. (2007b). Cultural effects on perspective taking. Psychological Science, 18, 600-606. Also featured in Science (June 15, 2007) in the Editor's Choice section.