



# How culture influences perspective taking: differences in correction, not integration

Shali Wu<sup>1,2\*</sup>, Dale J. Barr<sup>3\*</sup>, Timothy M. Gann<sup>4</sup> and Boaz Keysar<sup>5</sup>

<sup>1</sup> School of Economics and Management, Tsinghua University, Beijing, China

<sup>2</sup> School of Management, Kyung Hee University, Seoul, South Korea

<sup>3</sup> Institute of Neuroscience and Psychology, University of Glasgow, Glasgow, UK

<sup>4</sup> University of California at Merced, Merced, CA, USA

<sup>5</sup> Department of Psychology, The University of Chicago, Chicago, IL, USA

## Edited by:

Antonia Hamilton, University of Nottingham, UK

## Reviewed by:

Elizabeth Sheppard, University of Nottingham Malaysia Campus, Malaysia

Jessica Wang, University of Birmingham, UK

## \*Correspondence:

Shali Wu, School of Management, Kyung Hee University, 1 Hoegi-Dong, Seoul, 130-701, South Korea  
e-mail: shalichina@gmail.com;  
Dale J. Barr, Institute of Neuroscience and Psychology, University of Glasgow, 58 Hillhead Street, Glasgow G12 8QB, UK  
e-mail: dale.barr@glasgow.ac.uk

Individuals from East Asian (Chinese) backgrounds have been shown to exhibit greater sensitivity to a speaker's perspective than Western (U.S.) participants when resolving referentially ambiguous expressions. We show that this cultural difference does not reflect better integration of social information during language processing, but rather is the result of *differential correction*: in the earliest moments of referential processing, Chinese participants showed equivalent egocentric interference to Westerners, but managed to suppress the interference earlier and more effectively. A time-series analysis of visual-world eye-tracking data found that the two cultural groups diverged extremely late in processing, between 600 and 1400 ms after the onset of egocentric interference. We suggest that the early moments of referential processing reflect the operation of a universal stratum of processing that provides rapid ambiguity resolution at the cost of accuracy and flexibility. Late components, in contrast, reflect the mapping of outputs from referential processes to decision-making and action planning systems, allowing for a flexibility in responding that is molded by culturally specific demands.

**Keywords:** perspective taking, comprehension, cultural differences, ambiguity, reference

## INTRODUCTION

The human language comprehension system is shaped by informational demands related to communication that are relatively universal, as well as by demands of a more social nature that can vary widely across cultures. On the universal side, spoken language is inherently ambiguous at multiple levels, from lexical processing all the way up to the identification of speech acts and resolution of referential ambiguity. In addition, the speech signal itself is evanescent, requiring language comprehenders to rapidly commit to specific parsing decisions and interpretations. On the culturally specific side, cultures vary in the underlying norms and values that regulate social behavior, including norms for participation in conversational interaction. Do the cultural norms governing language and social interaction impact language processing as immediately and as powerfully as the universal demands for rapid ambiguity resolution? Or do they mainly determine how outputs from relatively universal processes are mapped onto later decisions and actions?

One way of addressing these general questions is to compare language users from different cultures in terms of how they integrate social and linguistic information during the online processing of referring expressions. In this study, we investigated cultural differences in how Chinese vs. Western (U.S.) language users take into account a speaker's diverging perspective when they resolve ambiguous references such as *the candle*. Referring expressions are of theoretical interest not only because they are ubiquitous in conversation, but also because they require listeners to go beyond the input – an expression such as the *candle*

denotes a particular class of object, not any particular individual object, and so listeners must access further information to determine which candle is being spoken about. When speakers and listeners have different visual perspectives, reference resolution will only be consistently successful if listeners take these differences into account. It is also methodologically convenient to study visual perspective taking during reference resolution, because a listener's eye gaze during the search for a referent provides an external index of the moment-by-moment process of language interpretation (Cooper, 1974; Tanenhaus et al., 1995). The waxing and waning of referential alternatives during processing will be reflected in moment-by-moment changes in the probability distribution of eye gaze over these alternatives.

Studies of perspective taking during reference resolution have experimentally created differences in perspective between speakers and listeners, and monitored listeners' interpretations as they interpret speakers' instructions to manipulate objects (Keysar et al., 2000; Nadig and Sedivy, 2002; Hanna et al., 2003). These studies suggest that listeners momentarily experience egocentric interference, with listeners considering "privileged" information that they know is unavailable to the speaker. For example, when searching for a referent for the expression *the candle*, listeners will temporarily consider a candle that is hidden from the speaker's view, in spite of their knowledge that the speaker does not know about it and therefore could only be referring to another, mutually visible candle (Keysar et al., 2000, 2003). Ultimately, listeners tend to eventually choose the mutually visible candle, although sometimes they may exhibit signs of confusion. For example, listeners

frequently ask the speaker to clarify the reference, even though if they took the speaker's point of view, they would realize the reference was perfectly clear.

Although the basic phenomenon of egocentric interference has been replicated in numerous studies, recent evidence suggests that it might be specific to the Western (European and North American) populations that have been the traditional object of study (Wu and Keysar, 2007). Cultures differ in the extent to which they emphasize the thoughts and beliefs of the individual versus those of the larger group, with cultures of East Asia exhibiting a more "collectivist" character relative to Western cultures, which tend to be more "individualist" in nature (Triandis et al., 1988; Markus and Kitayama, 1991; Ross et al., 2002). Lifelong membership in a particular culture may shape one's tendency or ability to take another's perspective into account while comprehending language. If so, then people from East Asian backgrounds should show more reliable and effective perspective taking than Westerners in resolving references.

To test this prediction, Wu and Keysar (2007) conducted an eye-tracking study using the basic visual perspective-taking task of Keysar et al. (2000), comparing the performance of Mandarin-speaking Chinese to English-speaking North Americans (from the U.S.). Each group performed the task in its participants' native language. Participants played the role of "listener," sitting across a table from a confederate "director," with a set of shelves placed between them. The contents of some of the shelves were visible from both sides, while others were hidden from the speaker's view. The director had a picture of how the objects in the shelves should be arranged, and told the listener which objects to move and where to move them. Embedded within the interaction were certain pre-scripted test instructions designed to be ambiguous from the listener's perspective, in that they could refer either to a mutually visible "target" object, or a privileged "competitor" object that was visible only to the listener. For example, in one such instruction the director told the listener to "move the candle to the top row," in a context where the listener saw two identical candles, only one of which was visible to the speaker. Listeners' eyes were tracked as they interpreted these test instructions. To provide a baseline, in a control condition, the competitor object was replaced with a non-competitor (an object that did not match the description of the target, such as a toy truck for the "candle" instruction). Egocentric interference would lead to an elevated probability of looking at the hidden competitor (candle) relative to the hidden non-competitor (toy truck), as well as in a delayed latency to fixate on the competitor.

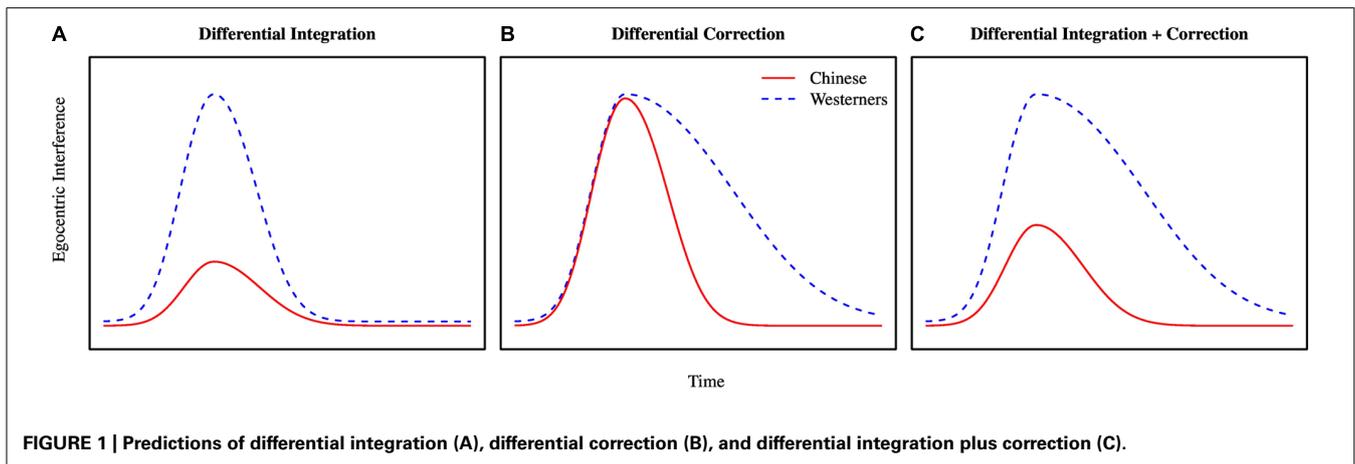
Wu and Keysar (2007) found that while Western participants showed the typical pattern of strong egocentric interference, Chinese participants showed virtually no interference. Unlike their American counterparts, Chinese participants were far less likely to fixate on privileged objects or to ask the speaker to clarify a reference that was ambiguous from their own perspective. In short, the Chinese participants were much more effective overall at taking the speaker's perspective into account.

How might these cultural differences be explained in terms of underlying cognitive processing? Wu and Keysar (2007) measured egocentric interference in terms of first fixation latency and

fixation duration, measures that can detect overall differences between groups, but that do not provide information about when such differences might emerge. To gain further insight into the underlying processes, we reanalyzed the data from Wu and Keysar (2007) using a more time-sensitive analysis in order to investigate the time-course of these cultural differences. Our analysis sought to test whether cultural differences emerged early or late relative to the onset of referential processing. On the one hand, cultural differences in egocentric interference may be present from the earliest moments of referential processing, suggesting that Chinese are able to more effectively use information about perspective to constrain the online processing of referring expression. On the other hand, it is possible that cultural differences emerge late, with both groups showing similar levels of egocentric interference early on, and only diverging later. This latter pattern would imply that the earliest moments of processing are unaffected by social information, and are driven largely by egocentric heuristics that enable rapid ambiguity resolution. Under this view, cultural differences would emerge late because participants from a Chinese background would be faster and more effective than Westerners at suppressing the pragmatically inappropriate information. In other words, cultural differences would not reflect differences in the ability to integrate social information into language processing, but instead would reflect differences in how listeners connect the outcome of basic referential processes to further thought and action.

Having laid out these possibilities in general terms, let us now consider in more detail the nature of the analysis, the possible outcomes, and their implications for theories of language processing and social cognition. Our analysis focused on the temporal profile of egocentric interference across the two cultural groups. We define egocentric interference as the difference in the likelihood of gazing at a hidden competitor (e.g., candle) versus gazing at a hidden non-competitor (e.g., toy truck). Note that we expect interference to show a curvilinear effect over time as shown by the curves in **Figure 1**, climbing from zero up to a peak from which it will eventually drop (as the listener will ultimately ignore the competitor and select the target).

Based on previous literature, we identify three different effect profiles that would be consistent with three different theoretical accounts. The first account, which we term the *differential integration* account, assumes that the cultural difference reflects the enhanced ability of Chinese to integrate information about the speaker's perspective with incoming linguistic information. This account would be consistent with constraint-based models of perspective use in language comprehension (Nadig and Sedivy, 2002; Hanna et al., 2003), as these models assume that information about a speaker's perspective is one of many cues that are simultaneously and interactively integrated during processing. Critically, the account does not differentiate between different types of cues, assuming that any available cue can influence any level of processing from its earliest moments, regardless of its source (e.g., whether it is derived from the unfolding syntax or semantics of the utterance or from situational pragmatics); the influence of a given cue depends only on its salience and reliability. Under this view, the shared perspective between the speaker and listener is a more salient and reliable cue for Chinese than for Westerners. Thus,



Chinese should show less egocentric interference than Westerners from the earliest moments of referential processing – in other words, the onset of the cultural difference should be simultaneous with the onset of the overall effect of egocentric interference (Figure 1A).

An alternative possibility is suggested by the autonomous activation hypothesis of Barr (2008) which, in contrast to constraint-based accounts, assumes that information about a speaker's perspective is a kind of situational cue that influences comprehension through anticipatory or post-lexical decision processing, but is not integrated into online lexical processing. Anticipatory processing refers to those steps taken by the listener in preparation for a referring expression, such as increasing attention to shared (mutually visible) objects. Barr (2008) found that comprehenders strongly *anticipated* that speakers would refer to referential candidates that were shared with the speaker, as evidenced by a higher probability of fixating shared than privileged objects. However, supporting autonomous activation, while interpreting the referring expression, listeners did not show any less interference from privileged than from shared competitors: the probability of gazing at a privileged competitor increased from its (lower) baseline at the same rate as the increase in probability for a shared competitor. Strikingly, in one experiment Barr (2008) found that unlike information about the speaker's perspective, listeners could very efficiently integrate contextual constraints derived from verb semantics. Based on these findings, Barr (2008) argued that lexical processing is encapsulated from high-level information about a speaker's perspective, and perhaps from other kinds of situational information, but is not strictly modular in the sense of being completely cognitively impenetrable.

The autonomous activation account would predict that Chinese participants might be more sensitive overall to a speaker's perspective, but without showing any greater ability to integrate this information with the linguistic input. Under this view, they should experience comparable levels of egocentric interference to Westerners, at least during the earliest moments of comprehension. In the current paradigm, this difference would be expressed as *differential correction*: Chinese participants would not initially experience less egocentric interference, but would be faster and

more effective at suppressing this interference than Westerners (Figure 1B)<sup>1</sup>. Of course, the integration and correction accounts are not mutually exclusive. A third possibility would be that the groups differ in both integration and correction, such that not only do Chinese participants experience lower interference from the earliest moments of comprehension, but they also are more efficient at suppressing this interference (Figure 1C). This pattern would be consistent with constraint-based models.

## MATERIALS AND METHODS

Additional details regarding experimental and data collection procedures are available in the original report (Wu and Keysar, 2007).

Our analyses considered looks to the competitor/non-competitor object from 250 ms after the onset of the critical word (e.g., the word “candle” in the phrase “move the candle. . .”) until 3000 ms. Observations for a given trial were terminated when listeners touched the target. These points varied from trial to trial, with a median of 3306 ms (2808 vs. 3844 for Chinese vs. U.S. participants, respectively), and a standard deviation of 4729 ms. For those trials that were terminated before 3000 ms, we replaced the missing frames with 0 s (representing the absence of a look to the competitor/non-competitor object).

Our goal was to test whether there was a time-lag between the onset of egocentric interference and the onset of cultural differences. To give an overview of our analysis method, we applied the *cluster randomization* method that has become popular in neuroimaging for determining the spatial and temporal extent of experimentally induced effects (Bullmore et al., 1999; Maris and Oostenveld, 2007; for prior adaptation of the solution to the analysis of visual-world data, see Barr et al., 2013). This approach is attractive for localizing effects in time in a visual-world study because it takes advantage of temporal correlations among adjacent data points to overcome the problem of multiple comparisons. The approach proceeds as follows. First, a significance test is performed at each time slice for a given effect (e.g., main effect

<sup>1</sup>Unlike Barr (2008), the current study does not offer the possibility of determining whether listeners were attentionally biased toward shared referential alternatives before hearing the referring expression. This would require additional conditions in which the competitor or non-competitor would be shared with the speaker.

or interaction). Then, “clusters” are defined by identifying adjacent time slices where the effect reaches significance, and where all effects are in the same direction. For example, consider tests performed at six subsequent time slices,  $t_1$ ,  $t_2$ ,  $t_3$ ,  $t_4$ ,  $t_5$ , and  $t_6$ , with tests significant at the 0.05 level only at  $t_2$ ,  $t_3$ ,  $t_5$ , and  $t_6$ . If  $t_2$  and  $t_3$  have effects in the same direction, then they form a cluster; likewise, if  $t_5$  and  $t_6$  are in the same direction, they also form a cluster. There are two separate clusters rather than a single one because of the intervening non-significant test at  $t_4$ . Once the clusters have been identified, a “cluster mass statistic” is calculated for each one, typically the sum of all of the individual test statistics (e.g.,  $t$  values) for that cluster. One obtains a null-hypothesis distribution for this cluster mass statistic through randomization (permutation tests); i.e., by randomly shuffling the condition labels across trials to create a large number of new datasets, repeating the above procedure on these datasets, and then storing the maximum obtained cluster mass statistic for each one. Finding a significant cluster between  $t_i$  and  $t_j$  with, say, 1000 additional randomized datasets and  $p = 0.002$  means that the cluster mass statistic for the original data was matched or exceeded in only 2 of the 1000 randomly created datasets.

We did this procedure twice, once to test for the main effect of Competition (competitor vs. non-competitor, e.g., egocentric interference), and once to test for the Culture-by-Competition interaction. The cluster randomization procedure provides only  $p$ -values; however, we were also interested in defining confidence limits for our effects. To obtain these confidence limits we used bootstrapping (details below). The remainder of this section provide further technical details regarding how these analyses were implemented.

Rather than comparing the observed probabilities at each time point, we fit a time-series model to the data and compared predictions from the model, following Barr et al. (2013). The time-series model smooths the data over time, thus minimizing noise and facilitating the detection of clusters (see **Figure 2B**). In the model, time was represented as a 7th order polynomial. We determined the order of the polynomial using a model search procedure, in which we calculated the Akaike Information Criterion (AIC) value for all models ranging from a 3rd order to a 16th order polynomial, and then selected the model with the lowest AIC, which was a 7th order polynomial. This was done on the grand-averaged data (i.e., without any predictors for Competition or Culture) so as not to bias the cluster randomization procedure.

Logistic regression models were fit to the data using the `multinom()` procedure of the `nnet` package (Venables and Ripley, 2002) of R statistical software (R Core Team, 2013), treating the outcome for each sample as binary. The cluster randomization procedure was performed twice, once treating subjects as random and items as fixed ( $p_1$ ), and once treating items as random and subjects as fixed ( $p_2$ ). For simplicity, we describe the procedure treating subjects as random factors. In addition to the parameter estimates from a fit to the original data, we created 999 additional data sets by randomly permuting the condition labels (competitor vs. control) independently for each unit (subject or item). To obtain orthogonality between the main effects and the interaction, the relabeling followed a “synchronized” permutation logic

(Pesarin, 2001). For a given culture group, a permutation was created by randomly choosing, with equal probability, whether or not to block-exchange all competitor and non-competitor labels for each subject. The same number of exchanges was then performed for the other culture group (with the units undergoing the exchange chosen at random). The “synchronization” of exchanges across groups (i.e., ensuring that the same number of exchanges of Competition occurs at each of the two levels of Group) ensures that the tests for the main effects and interaction are orthogonal (Pesarin, 2001). The parameter estimates for the model fit to each of these data sets were stored as a row in the matrix.

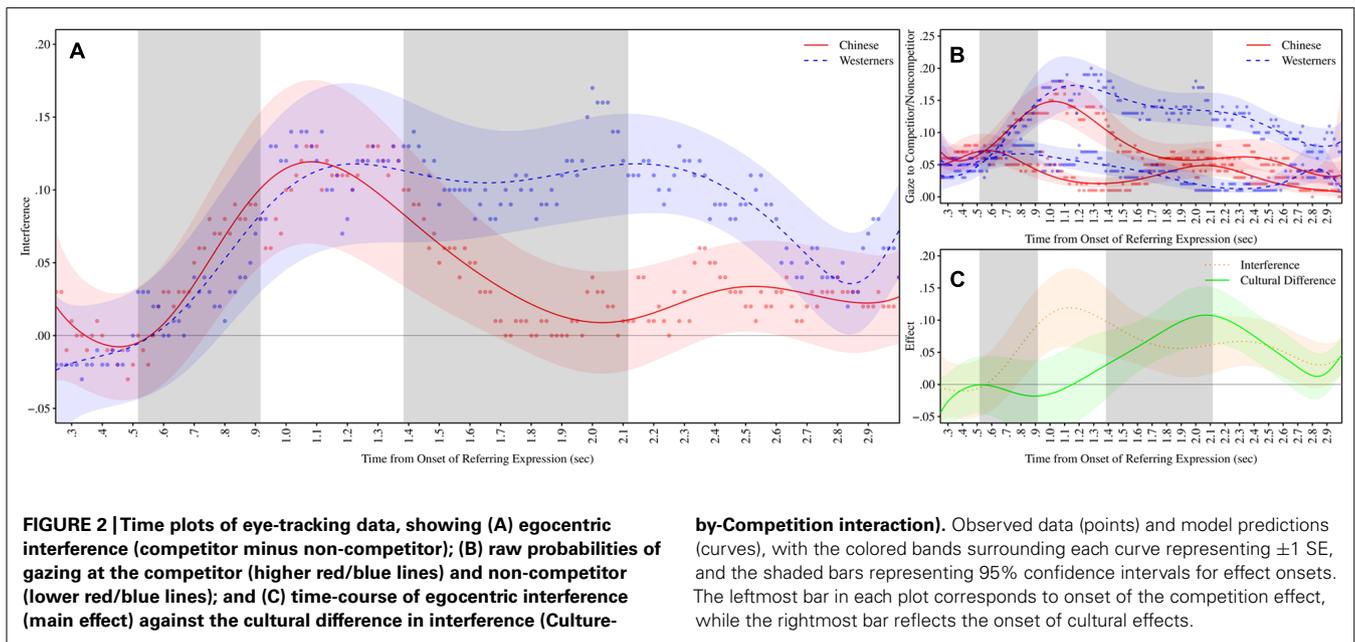
After all datasets were created, we then calculated the predicted log odds of a gaze to the competitor/non-competitor object at each 1/60 of a second (i.e., for each frame of data sampled at 60 Hz), deriving main effects and interactions at each time point. The predicted effects for each of the 1000 datasets (including the original) were stored as separate rows in a matrix. The  $p$ -value for each effect (main effect of competition or interaction) at each time point was given as the number of rows in the effect matrix exceeding the original value divided by the number of rows in the matrix. Then, we identified clusters by grouping together all temporally adjacent time-frames where the effect reached significance. A cluster mass statistic (Bullmore et al., 1999) was calculated for each cluster by summing together the negative (natural) logarithm of each  $p$ -value belonging to the cluster, such that smaller  $p$ -values would contribute to a larger cluster mass statistic; for example, for 0.05 the negative log is 3, and for 0.0001 it is 9.21. This cluster mass statistic was calculated for each cluster in the original data. Then, a null-hypothesis distribution for the statistic was derived by treating each permuted data set as if it was the “original” data, calculating  $p$ -values and cluster mass statistics in the manner described above, and storing the maximum observed statistic for each permutation. This allowed us to identify the onset of the first significant cluster for both the main effect as well as for the interaction.

To obtain confidence limits, we repeated the complete analysis described above for 999 bootstrapped versions of the data set, wherein we sampled subjects with replacement from each group at random. If for a given bootstrapped dataset, no significant cluster for the main effect or interaction was detected at  $\alpha = 0.05$ , the  $\alpha$  level was progressively lowered until a cluster was detected, stopping at  $\alpha = 0.2$ . Although the confidence limits derived from bootstrapping provide useful information, the main inferential focus is on the results of the cluster randomization on the original data.

## RESULTS

The time-course data appear in **Figure 2**. Note that 0 ms does not correspond to the onset of the utterance (e.g., “move” in “move the candle”), but to the onset of the referring expression within the utterance (e.g., “candle”). Thus any differences in timing between groups cannot be attributed to possible linguistic differences between Chinese and English in the duration of the material preceding the referring expression.

The cluster randomization procedure detected significant overall egocentric interference from 750 to 2800 ms, with the 95% confidence interval for the onset of interference ranging from 517



to 917 ms ( $p_1 < 0.001$ ,  $p_2 = 0.003$ )<sup>2</sup>. As **Figure 2** clearly shows, there was a large time-lag between the onset of egocentric interference and the onset of a cultural difference in this interference (as given by a Culture-by-Competition interaction). There was little evidence that Chinese participants experienced any less interference than U.S. participants until 1767 ms, approximately 1000 ms after the onset of interference. The Culture-by-Competition interaction was significant from 1767 to 2483 ms ( $p_1 = 0.009$ ,  $p_2 = 0.049$ ), with the 95% confidence interval for the onset ranging from 1383 to 2117 ms. Note that there was no overlap between the confidence interval for the onset of the cultural difference (1383–2117) with that for the onset of egocentric interference (517–917). Furthermore, we directly computed the delay between the onsets for each bootstrapped sample, which yielded a 95% confidence interval for the lag between 600 and 1400 ms.

## DISCUSSION

Overall, our findings support the hypothesis that language users from different cultures share a common stratum of referential processing, with cultural variation in how the products of these early referential processes are used in the higher-level processes governing thought and action. Specifically, whereas neither Chinese nor Western participants were able to integrate the situational cue of the speaker's perspective into lexical processing, Chinese participants were better able to suppress the interference.

Could our findings of common interference and differential correction be alternatively explained in terms of linguistic differences between Mandarin Chinese and English? One potentially

relevant difference is that Mandarin lacks definite marking, such that the Mandarin version of the English expression “move the candle” might be glossed in English as “move candle.” It might be argued that the Chinese participants were interpreting the descriptions as if the speaker had said, “move any candle.” This would indeed predict that the Chinese participants would experience less interference than the U.S. participants because they would not need to decide between the two possible referents, but could pick either one. However, if this were the case, then Chinese participants should have shown a stronger tendency than U.S. participants to move the hidden candle, since any candle would suffice. However, the data showed the exact opposite. While the U.S. participants sometimes moved the occluded candle, the Chinese participants never did.

One possible concern might be that the later correction for Chinese participants reflects shorter referring expressions in Chinese, or more rapid speech when the confederate spoke Chinese. Although we lack the data to directly address this question, the overall patterns shown in **Figure 2** make this explanation seem unlikely. First, if the earlier correction occurred because the Chinese expressions were briefer or spoken more rapidly, then not only would the correction process take place earlier, but so would the egocentric interference; specifically, the initial rising slope of the curve should have been much steeper for the Chinese group than for the Western group, and should have reached its peak much earlier. However, egocentric interference seems to rise at similar rates for both groups, and both seem to initially reach their maximum values at roughly the same time (1000–1200 ms). Second, whereas the correction process seems to begin at around 1000 ms for the Chinese group, it seems delayed until about 2200 ms for the American group. This is far too great of a disparity to be explained by differences in the spoken expressions, given that expressions in these types of experiments typically last no more than 1 s. Finally, the groups differ not only in the timing of the correction, but also

<sup>2</sup>Although an onset of 750 ms is quite late relative to typical visual-world studies (250–350 ms), this is not surprising given that our paradigm presented participants with a more demanding search task than in a typical study. Whereas a typical grid in Wu and Keysar (2007) contained nine alternatives appearing in any of 16 possible locations, a typical visual-world task presents no more than four referential alternatives in fixed locations (Huettig et al., 2011).

in the *efficacy* of the correction, with a sudden sharp decline for the Chinese group, and more of a lingering pattern for the Western group. Thus, these patterns seem less likely to be driven by differences in the stimuli, and more likely to reflect true cultural differences in linguistic interpretation.

Constraint-based views would have difficulty accounting for the extreme delay in the emergence of cultural differences relative to the onset of egocentric interference. If, as constraint-based views predict, language users can integrate perspective information from the earliest moments of processing, and Chinese participants attend more strongly to the shared perspective than Westerners, then Chinese participants should have shown less egocentric interference from the very earliest moments of processing. Our view, then, is that despite attending more strongly to shared information, Chinese participants are no better at integrating it into referential processing. However, an alternative view must be considered, which is that perhaps the late emergence does not reflect a standalone correction process, but simply reflects delayed activation of shared information relative to other kinds of information. Under this view, had the shared knowledge become activated earlier, perhaps we would have seen its effects earlier in processing. However, it is unclear what would account for the delayed activation of shared knowledge within the current paradigm. For one, in the current experimental situation, listeners knew well before hearing the referring expression which items their partner could see and which they could not see. In other words, information about what was shared was available to participants even before any referential information became available. It is therefore not clear why listeners would wait for a referring expression to activate the shared knowledge, rather than using it to predict potential referents in advance. It is not possible to tell whether listeners in fact made such predictions, because this requires comparing shared to privileged objects, and our analysis only considered privileged objects. However, experiments using a similar setup have found that in the interval preceding the onset of the referring expression, listeners are more likely to look at shared objects (Keysar et al., 2000). Furthermore, recent experiments including conditions where competitors/non-competitors are shared show that listeners spontaneously access shared knowledge prior to the onset of referring expressions, but are unable to integrate this information into early referential processes (Barr, 2008). Specifically, listeners attend less overall to privileged objects than to shared objects, but nonetheless experience similar levels of interference from competitors regardless of whether they are shared or not. It would be of interest to repeat these experiments with East Asian participants. Our account predicts greater access to shared knowledge among East Asians, but without any reduction in the size of the interference produced by competitors.

Our view that information about perspective is involved in correction is consistent with an anchoring and adjustment view of perspective taking (Keysar et al., 2000), in which listeners anchor interpretation in their own perspectives, and use information about the speaker's perspective to incrementally adjust away from the anchor. However, distinct from Keysar et al.'s (2000) original formulation, our findings, together with those of Barr (2008), suggest that listeners do not strategically "anchor" in their own

egocentric perspective as a kind of reasoning heuristic; rather, their anchoring is forced upon them by the autonomous activation of referents by low-level interpretation processes that are blind to information about the speaker's perspective (Barr, 2008). Under this view, the noted egocentrism of listeners might be best characterized as a form of "mental contamination" – i.e., the result of rapid, automatic processes that are beyond control and possibly even awareness (Wilson and Brekke, 1994).

Consistent with the use of common ground in correction, other research shows that perspective taking involves cognitive effort (Rosnagel, 2000; Brown-Schmidt, 2009; Nilsen and Graham, 2009; Lin et al., 2010), and recent neuroimaging evidence suggests a role for the medial pre-frontal cortex in the adjustment process (Tamir and Mitchell, 2010). Furthermore, the correction account is also consistent with dual process views of perspective taking, which assume that social judgments reflect the combination of both efficient but inflexible processing that uses limited information and more flexible but effortful processing that can draw upon a broader set of information (Apperly and Butterfill, 2009). However, the current data offer no insight into why the adjustment process might differ across the groups. One possibility, consistent with the collectivist vs. individualist distinction, is that information about a speaker's perspective is simply more available to people from a collectivist background, since their cultures require greater attunement to one another's knowledge. Another is that perhaps Chinese participants are more motivated to perform the task "correctly" due to heightened concerns about self-presentation. A further possibility is that membership in a Chinese culture, where self-control is valued, results in better executive control abilities. This explanation is supported by research that finds enhanced executive control abilities among Chinese as opposed to North American children (Sabbagh et al., 2006), who nonetheless showed comparable performance on a belief reasoning task. As we have argued here and elsewhere (Keysar et al., 2003; Barr, 2008) listeners' difficulty in identifying the intended referent in conversational perspective-taking tasks is unlikely to be the result of a failure to have the appropriate beliefs about what is shared with the speaker. Instead, it seems to reflect difficulty using this information to constrain the processing of the linguistic input. To the extent that early referential processes are not guided by beliefs about the speaker, these processes will boost activation of referents that are pragmatically implausible, even in spite of correct and accessible representations of shared knowledge. Because suppressing this knowledge will involve executive control, it is here where we would expect to see strong individual (and cultural) differences. Although in this respect our view is consistent with Sabbagh et al.'s (2006) developmental findings, it is important to note that it is not yet known whether the differences in executive function that Sabbagh et al. (2006) noted extend into adulthood.

Whatever the explanation for the cultural differences, a recent study suggests that it might be possible to induce cultural effects through priming. Luk et al. (2012) replicated Wu and Keysar's (2007) study but with Chinese-Westerner bi-cultural individuals. Participants primed by images from Western culture committed more egocentric errors on the perspective-taking task relative to participants who were primed by images from Chinese culture.

The fact that cultural differences can be situationally induced in bicultural individuals suggests that they arise from flexible modes of processing. This flexibility is consistent with our explanation of such differences in terms of differential correction – it would seem easier to override a deliberative and effortful correction process than an integration process that is largely routinized and automatic.

In sum, our data suggest that people from different cultures share a common core of ambiguity resolution processes, but differ in how the output from these processes is linked to higher-level systems governing thought and action. The two cultures we have studied show systematic differences in how they prioritize the individual vs. the social (Triandis et al., 1988; Markus and Kitayama, 1991; Ross et al., 2002). Finding equivalent interference from privileged information in spite of such differences suggests that such egocentrism might be a universal consequence of rapid ambiguity resolution during spoken language comprehension.

## ACKNOWLEDGMENTS

This research was supported by National Science Foundation of China Grants 71002014 and 71110107027 to Shali Wu. Boaz Keysar also received partial support from a grant from the University of Chicago's Wisdom Research Project and the John Templeton Foundation.

## REFERENCES

- Apperly, I. A., and Butterfill, S. A. (2009). Do humans have two systems to track beliefs and belief-like states? *Psychol. Rev.* 116, 953–970. doi: 10.1037/a0016923
- Barr, D. J. (2008). Pragmatic expectations and linguistic evidence: listeners anticipate but do not integrate common ground. *Cognition* 109, 18–40. doi: 10.1016/j.cognition.2008.07.005
- Barr, D. J., Jackson, L., and Phillips, I. (2013). Using a voice to put a name to a face: the psycholinguistics of proper name comprehension. *J. Exp. Psychol. Gen.* doi: 10.1037/a0031813 [Epub ahead of print].
- Brown-Schmidt, S. (2009). The role of executive function in perspective taking during online language comprehension. *Psychon. Bull. Rev.* 16, 893–900. doi: 10.3758/PBR.16.5.893
- Bullmore, E. T., Suckling, J., Overmeyer, S., Rabe-Hesketh, S., Taylor, E., and Brammer, M. J. (1999). Global, voxel, and cluster tests, by theory and permutation, for a difference between two groups of structural MR images of the brain. *IEEE Trans. Med. Imaging* 18, 32–42. doi: 10.1109/42.750253
- Cooper, R. M. (1974). The control of eye fixation by the meaning of spoken language: a new methodology for the real-time investigation of speech perception, memory, and language processing. *Cogn. Psychol.* 6, 84–107. doi: 10.1016/0010-0285(74)90005-X
- Hanna, J. E., Tanenhaus, M. K., and Trueswell, J. C. (2003). The effects of common ground and perspective on domains of referential interpretation. *J. Mem. Lang.* 49, 43–61. doi: 10.1016/S0749-596X(03)00022-6
- Huetting, F., Rommers, J., and Meyer, A. S. (2011). Using the visual world paradigm to study language processing: a review and critical evaluation. *Acta Psychol.* 137, 151–171. doi: 10.1016/j.actpsy.2010.11.003
- Keysar, B., Barr, D. J., Balin, J. A., and Brauner, J. S. (2000). Taking perspective in conversation: the role of mutual knowledge in comprehension. *Psychol. Sci.* 11, 32–38. doi: 10.1111/1467-9280.00211
- Keysar, B., Lin, S., and Barr, D. J. (2003). Limits on theory of mind use in adults. *Cognition* 89, 25–41. doi: 10.1016/S0010-0277(03)00064-7
- Lin, S., Keysar, B., and Epley, N. (2010). Reflexively mindblind: using theory of mind to interpret behavior requires effortful attention. *J. Exp. Soc. Psychol.* 46, 551–556. doi: 10.1016/j.jesp.2009.12.019
- Luk, K. K. S., Xiao, W. S., and Cheung, H. (2012). Cultural effect on perspective taking in Chinese–English bilinguals. *Cognition* 124, 350–355. doi: 10.1016/j.cognition.2012.05.016
- Maris, E., and Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *J. Neurosci. Methods* 164, 177–190. doi: 10.1016/j.jneumeth.2007.03.024
- Markus, H., and Kitayama, S. (1991). Culture and the self: implications for cognition, emotion, and motivation. *Psychol. Rev.* 98, 224–253. doi: 10.1037/0033-295X.98.2.224
- Nadig, A. S., and Sedivy, J. C. (2002). Evidence of perspective-taking constraints on children's on-line reference resolution. *Psychol. Sci.* 13, 329–336. doi: 10.1111/j.0956-7976.2002.00460.x
- Nilsen, E., and Graham, S. (2009). The relations between children's communicative perspective-taking and executive functioning. *Cogn. Psychol.* 58, 220–249. doi: 10.1016/j.cogpsych.2008.07.002
- Pesarin, F. (2001). *Multivariate Permutation Tests: With Applications in Biostatistics*. Chichester: Wiley.
- R Core Team. (2013). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.
- Ross, M., Xun, W., and Wilson, A. (2002). Language and the bicultural self. *Pers. Soc. Psychol. Bull.* 28, 1040–1050. doi: 10.1177/01461672022811003
- Rosnagel, C. (2000). Cognitive load and perspective taking: applying the automatic-controlled distinction to verbal communication. *Eur. J. Soc. Psychol.* 30, 429–445. doi: 10.1002/(SICI)1099-0992(200005/06)30:3<429::AID-EJSP3>3.0.CO;2-V
- Sabbagh, M. A., Xu, F., Carlson, S. M., Moses, L. J., and Lee, K. (2006). The development of executive functioning and theory of mind: a comparison of Chinese and U.S. preschoolers. *Psychol. Sci.* 17, 74–81. doi: 10.1111/j.1467-9280.2005.01667.x
- Tamir, D. I., and Mitchell, J. P. (2010). Neural correlates of anchoring-and-adjustment during mentalizing. *Proc. Natl. Acad. Sci. U.S.A.* 107, 10827. doi: 10.1073/pnas.1003242107
- Tanenhaus, M. K., Spivey, M. J., Eberhard, K. M., and Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science* 268, 1632–1634. doi: 10.1126/science.7777863
- Triandis, H. C., Bontempo, R., Villareal, M. J., Asai, M., and Lucca, N. (1988). Individualism and collectivism: cross-cultural perspectives on self-in-group relationships. *J. Pers. Soc. Psychol.* 54, 323–338. doi: 10.1037/0022-3514.54.2.323
- Venables, W. N., and Ripley, B. D. (2002). *Modern Applied Statistics with S*, 4th Edn. New York: Springer. doi: 10.1007/978-0-387-21706-2
- Wilson, T. D., and Brekke, N. (1994). Mental contamination and mental correction: unwanted influences on judgments and evaluations. *Psychol. Bull.* 116, 117–142. doi: 10.1037/0033-2909.116.1.117
- Wu, S., and Keysar, B. (2007). The effect of culture on perspective taking. *Psychol. Sci.* 18, 600–606. doi: 10.1111/j.1467-9280.2007.01946.x

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 30 June 2013; accepted: 13 November 2013; published online: 02 December 2013.

Citation: Wu S, Barr DJ, Gann TM and Keysar B (2013) How culture influences perspective taking: differences in correction, not integration. *Front. Hum. Neurosci.* 7:822. doi: 10.3389/fnhum.2013.00822

This article was submitted to the journal *Frontiers in Human Neuroscience*.

Copyright © 2013 Wu, Barr, Gann and Keysar. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.