Evidence for a two-stage account of prediction

Ruth Corps[1,2], Charlotte Brooke[2], & Martin Pickering[2]
[1] Psychology of Language Department, Max Planck Institute for Psycholinguistics, Ruth.Corps@mpi.nl
[2] Department of Psychology, University of Edinburgh

Comprehenders often predict what they are going to hear. For example, they will preferentially look at edible objects immediately after hearing *The boy will eat…*, and thus predict that the speaker is about to mention such an object [1]. But what exactly do comprehenders predict? And more importantly, what information do they use to make these predictions? Do they immediately make the best (most appropriate) predictions they can, or do such predictions take time and resources?

Comprehenders may immediately predict *appropriately* (from the speaker's perspective), because their predictions will tend to correspond to what the speaker actually says – a one-stage account. But considering perspective is effortful [2], and so comprehenders may initially predict on the basis of automatic *associations* [3], or on the basis of what they would (*egocentrically*) say if they were speaking [4] – two different two-stage accounts.

We tested among these alternatives in three experiments using the visual-world paradigm, in which participants listened to sentences (N=28; e.g., *I would like to wear…*), while viewing four objects on-screen. We manipulated the gender of the speaker (as indexed by their voice and face; [5]), the participants, and the characters in the sentences. In particular, participants heard a male or a female speaker producing sentences about gender-stereotyped objects (as assessed in a pre-test; N=80). One target (a dress) and one distractor (a hairdryer) were stereotypically female; the other target (a tie) and one distractor (a hairdryer) were stereotypically female; the other target (a tie) and distractor (a drill) were stereotypically male. To make different perspectives salient, sentences began with *I* in Experiment 1 (speaker's perspective), *You* in Experiment 2 (participant's perspective), and the name *James* or *Kate* in Experiment 3 (character's perspective). We fitted Bayesian generalized linear mixed effects models to binomial fixations in 50 ms time bins from 1000 ms before to 1500 ms after critical verb onset (*wear*).

In Experiment 1, participants (N=24, 12 males) fixated targets more than distractors from 450 ms after verb onset (ps < .05), before the target was mentioned, suggesting they predicted associatively. Participants also fixated appropriate targets (which matched the speaker's gender) more than inappropriate targets (which matched their own gender) from 600 ms (ps < .05; see Figure 2) and there was no point at which they predicted egocentrically. This appropriateness effect occurred later than the associative effect. For example, a male participant listening to a female speaker initially fixated the dress and the tie (over the hairdryer and the drill), and then homed in on the dress (over the tie).

We found similar effects in Experiment 2, in which sentences used the pronoun *You* rather than *I*, so that appropriate prediction was not tied to the speaker's perspective. Participants (N=32, 16 males) predicted associatively from 300 ms after verb onset (ps < .05) and appropriately from their own perspective from 1000 ms (ps < .05). Note that this appropriate effect was later in Experiment 2 than Experiment 1, perhaps because there is some ambiguity as to who *You* refers to [6]. But importantly, participants again predicted appropriately later than they predicted associatively, providing further evidence for a two-stage account.

In Experiment 3, participants (N=32, 16 males) listened to sentences referring to a male (James) or a female (Kate) character. Participants predicted associatively from 300 ms after verb onset (ps < .05). They also predicted appropriately (looking at the target that matched the character's gender) from 450 ms, again later than the associative effect.

We conclude that comprehenders predict in two different ways – associatively, by drawing on information associated with the verb, and appropriately, by drawing on relevant contextual information. We show how these findings are compatible with initial resource-free prediction-by-association, followed by slower resource-intensive prediction-by-production [7].

References

[1] Altmann, G. T. M. & Kamide, Y. (1999). Incremental interpretation at verbs: restricting the domain of subsequent reference. *Cognition, 73,* 247-264.
[2] Lin, S., Keysar, B., & Epley, N. (2010). Reflexively mindblind: Using theory of mind to interpret behavior requires effortful attention. *Journal of Experimental Social Psychology, 46,* 551-556.
[3] Neely, J. H. (1977). Semantic priming and retrieval from lexical memory. Roles of inhibitionless spreading activation and limited-capacity attention. *Journal of Experimental Psychology: General, 106,* 226-254.
[4] Keysar, B., Barr, D. J., Balin, J. A., & Brauner, J. S. (2000). Taking perspective in conversation: The role of mutual knowledge in comprehension. *Psychological Science, 11,* 32-38.
[5] Van Berkum, J. J., Van den Brink, D., Tesink, C. M., Kos, M., & Hagoort, P. (2008). The neural integration of speaker and message. *Journal of cognitive neuroscience, 20,* 580-591.
[6] Brunyé, T. T., Ditman, T., Mahoney, C. R., Augustyn, J. S., & Taylor, H. (2009). When you and I share perspectives: Pronouns modulate perspective taking during narrative comprehension. *Psychological Science, 20*, 27-32.
[7] Pickering, M. J. & Gambi, C. (2018). Predicting while comprehending language: A theory and review. *Psychological Bulletin, 144,* 1002-1044.

Figure 1. Examples of stimuli used in the three experiments

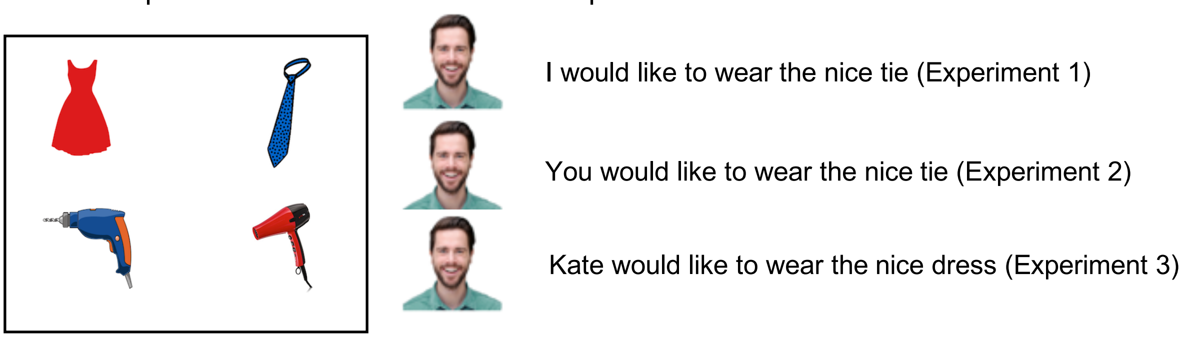

Figure 2. Eye-tracking results for Experiment 1. Shapes at the top of the graph show significant differences ($p < .05$) for the time bin on the x-axis between the critical pairs of pictures.