Reply

Reply to Brown-Schmidt and Heller

In reaction to Bezuidenhout (2013), Brown-Schmidt and Heller argue that the Visual World Paradigm (VWP) can tell us something about the macro-processes of verbal communication, despite what I called its individual-centered, narrow temporal focus methodology. They do an excellent job of arguing for their position, and I completely accept what they say. In my paper I never doubted that this methodology could be helpful, and I certainly never said: ''it is not a suitable methodology for studying perspective-taking in conversation.'' What I actually said was:

> I am not claiming that experimental methods in psychology are useless and/or irrelevant to understanding perspective taking abilities. When pragmatic factors are carefully controlled for, as they are in the sorts of experiments carried out by Tanenhaus and his collaborators, individualistic methods can give us a glimpse of the essentially social processes of conversational interaction. However, the findings from these psycholinguistic experiments should be considered alongside the findings of sociolinguists who use other methodologies, such as conversational analysis (Bezuidenhout, 2013:15).

The other methodologies I had in mind are ones that are multi-agent centered and that operate with a broad temporal focus (typically looking at multiple turns in an evolving conversation, as opposed to just reactions to a single utterance or to repeated utterances of the same sort).

But this is no fun, and may seem like a cop out, to simply accept everything Brown-Schmidt and Heller say! So I will try to respond to some of their points, taking a slightly more skeptical stance to the virtues of the VWP than they do.

When I referred to methodologies that have a ''narrow temporal focus'', what I had in mind are methods where the data collected consists of two or three millisecond snap shots of some temporally unfolding, observable process – typically a person's eye gaze fixations as he/she scans a visual display while simultaneously listening to a verbal instruction to manipulate an item in the display in some way. This eye-fixation data is assumed to be a window into the person's unfolding process of language comprehension (and sometimes language production, depending on the experimental setup). There will typically be many trials in a single experiment and many participants. That is, a single individual will hear multiple instructions over the course of the experiment and other individuals, possibly in different treatment groups, will also see similar visual displays and be given similar instructions. So the eye fixation data will eventually be aggregated over trials and over participants. Moreover, there will be other sorts of aggregating too – for example, the visual display will typically be segmented into just a few regions and the participant's eye gazes anywhere within in a single region within a certain time bin will be summed. Nevertheless, none of this aggregating makes the data any less individual-centered and does not transform the data into the sort of multi-agent, broadly temporally focused data I had in mind – e.g., data from multiple turns in a conversation between two or more people.

Why does this matter? As the quotation above suggests, I think that data from all these sources should be taken into account in any full picture of perspective taking in conversation. So I am not trying to dismiss this eye-monitoring data. The ability to monitor other people's perspectives is a social skill, although one that is of course possessed by individuals. So it is not inappropriate to study individuals as they perform tasks that can reasonably be taken to trigger perspective-taking abilities. However, I do think that we should be careful not to over-interpret the data we can get from the VWP.

In Bezuidenhout (2013), I suggested that the sorts of experimental setups used in the VWP are more likely to trigger a game strategy mind-set in participants rather than a conversational perspective taking mind-set. However, here I will not

push this skeptical line. I will assume that the referential communication games used in these experiments do trigger perspective-taking abilities (even if not quite the same ones used in ordinary conversation). We can then try to catch people in the act of doing such perspective-monitoring and vary the social conditions under which they do such monitoring to determine whether and which social variables have an impact on these processes.

For example, in the ''entrainment'' experiments mentioned by Brown-Schmidt & Heller on pp. 280–81, a matcher and a director must work together to get the matcher's display to correspond to the director's display. At some point in the game, matchers hear a new object description for an object already established as part of the common ground from either the same or a new director. Matchers take longer to identify the object when the original director uses this new label, as compared with the time taken when the instruction comes from a new director. If one is dealing with a new interlocutor, one should not act as though the common ground is the same as the one built up over time with one's previous interlocutor. So these results show that people are sensitive to what is old or new information for their interlocutors. (However, when the old label is used, matchers are just as fast to identify the object, whether the instruction comes from the original or a new director. The use of an old label by a new director should cause a bit of a boggle, but apparently it does not. Perhaps the labels used in these experiments were generic enough that it was not too much of a surprise that a new director would use the same label as the original director – the labels used were descriptions such as 'the shiny cylinder' and 'the silver pipe').

Furthermore, in Bezuidenhout (2013), I discussed the work of Brown-Schmidt et al. (2008), which used WH-questions to ask about objects in a visual display where some items are in the addressee's privileged ground and some are in the common ground. Such questions presuppose that the speaker is ignorant of something that the addressee knows. What is at issue is whether people's perspective taking responses are sensitive these sorts of manipulations – that is, whether they will distinguish between objects in privileged and common ground. And it turns out that people are sensitive to these distinctions about who knows what when.

So I concur wholeheartedly that one can draw conclusions about an essentially social process by looking at what individuals do on a moment-to-moment basis over the course of many individual trials in an experiment, with the aim of comparing trials across two or more conditions where some social variable is being manipulated – e.g., where one varies who the participant's interlocutor is, or what WH-question has just being asked, etc.

Does this show that language comprehension is a wholly mind-internal process that can be studied in a laboratory without any concerns that we may be looking at the phenomenon through distorting lenses or, worse yet, not even looking at language comprehension but rather some ersatz process invoked in participants in the lab?

In Bezuidenhout (2013), I argued for an interactionist picture of communication. Hence I very much resist the idea that all that matters is what goes on inside an individual's head during the few seconds it takes them to hear an instruction, search for an object in a visual display, and then click on the object or interact with it in some way. Of course what goes on inside people's heads is a crucial part of the whole story. If nothing went on inside their heads, there would be no possibility of an interaction. However, one must understand the conversational context in a more holistic fashion.

The sort of interactionist picture I have in mind is the one defended by Gallagher (2012). He argues that monitoring people's beliefs and intentions is not a matter of building elaborate mental (meta-)representations, as the so-called Theory-theory of mind supposes, nor of undergoing an internal simulation of mental states, as the Simulation theory of mind supposes. Rather, tracking other people's minds is a matter of being engaged in a joint activity that includes tracking the other person's eye gaze and jointly attending to objects. (Brown-Schmidt & Heller agree that monitoring eye gaze is an important component of conversational interactions. See p. 282. However, one of the features of the referential communication task used in VWP experiments is that the setup *prevents* interlocutors from using eye gaze as a cue for interpretation, as there is typically a barrier set up between the matcher and director).

In contrast to this interactionist picture, Brown-Schmidt & Heller assume a very internalist conception of language comprehension. In the final paragraph of their comments, they mention:

A. Mechanisms that use perspective representations
B. The representation of perspective information that is developed over the course of a conversation
C. Mechanisms of perspective taking that are influenced by narrow and broad processes
   (e.g., word formation processes and entrainment processes respectively).

I have to confess to being a bit unsure of what sort of cognitive model Brown-Schmidt & Heller are operating with. The mechanisms/processes that I have numbered (A)–(C) are mentioned within a single concluding paragraph, as though they all referred to basically the same thing. But they are not quite the same. (A) assumes that there are already perspective representations there to be used. (B) assumes that these perspective representations must be built up over time. (C) is ambiguous as to which mechanisms are influenced by narrow and broad processes. Is it the mechanisms that *use* perspective representations, or is it the mechanisms that *build* them? Also, what is the difference between a mechanism and a process? For current purposes, I take a mechanism to be a crystalized ability (mental structure) that can