RESPONSE

Gaze following: how (not) to derive predictions from a computational model

Jochen Triesch[1,2], Christof Teuscher[3] and Gedeon O. Deák[1]

1. *Department of Cognitive Science, University of California, San Diego, USA*
2. *Frankfurt Institute for Advanced Studies, Johann Wolfgang Goethe University, Germany*
3. *Los Alamos National Laboratory, Los Alamos, USA*

This is a response to the commentaries by Csibra (2006), Moore (2006) and Richardson and Thomas (2006).

We thank the four commentators for carefully evaluating our model (Triesch, Teuscher, Deák, & Carlson, 2006) and sharing their opinions. We will respond to the commentaries one by one.

Chris Moore focuses on two important limitations of our model: our choice not to incorporate attentional cueing mechanisms, and our choice to ignore spatial aspects of gaze following. We agree with his comments. It is important to emphasize, however, that there are good reasons to develop computational models in an incremental fashion. Richardson and Thomas, in their commentary, put it very nicely: 'Overly complex models are time consuming to build and run the risk of revealing little about the potential causes of a particular behavior, since credit and blame assignment can become opaque.' Our choice was thus to start with a very simple model, which we view as a useful stepping stone for the development of more complex and powerful models. In fact, our current work has been extending the present model in the suggested directions (see Jasso, Triesch & Teuscher, 2005, for recent results).

Gergely Csibra's commentary nicely illustrates some pitfalls of deriving and interpreting the predictions of a computational model. In a first manipulation of our model, Csibra fixes the only rewarding visual stimulus (apart from the caregiver) to one location. In this situation, there are only two interesting things to see in the environment and they are always in the same place. Clearly, the model infant 'growing up' in this environment has nothing to gain by learning to follow gaze because the caregiver provides no additional information regarding where the other interesting thing is. In fact, whenever the caregiver looks away from the fixed target, *it is guaranteed that there will be nothing interesting to see there*. Not surprisingly, the model infant does not learn to follow gaze, but learns to largely ignore the caregiver's looking behavior and look directly to the fixed target. Thus, the model behavior is easily explained and understood given the assumptions made. But what conclusions should be drawn from this finding? Csibra thinks it means that our model predicts that 'Infants raised in stable environments will be slower, or less likely, to develop a gaze following response than infants raised in unpredictably changing environments.' What he ignores is that his manipulation also made the caregiver's gaze predict the *absence* of a rewarding sight in all but one location, violating a fundamental assumption of our model. Thus, Csibra's interpretation is misleading.

His second prediction is somewhat better justified. Based on our experiments with multiple targets he suggests that: 'The richer the environment in which an infant grows up, the less likely (or slower) that she will learn to follow gaze.' This statement is essentially correct, but it ignores the fact that in an impoverished environment where there is nothing interesting to see at certain times, gaze following will also be slow to emerge. Thus, too much or too little 'richness' of the environment are both detrimental. We find neither prediction surprising. That too rich an environment may not be optimal for the emergence of gaze following is arguably supported by the observation that most experiments on gaze following find it necessary to use rooms with plain, undistracting walls.

Finally, Csibra extends our model (a good idea in principle) and derives two further predictions from his extension to our model. It is very difficult to comment on these predictions, simply because Csibra's description

Address for correspondence: Jochen Triesch, Department of Cognitive Science, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0515, USA; e-mail: triesch@cogsci.ucsd.edu

© 2006 The Authors. Journal compilation © 2006 Blackwell Publishing Ltd. Published by Blackwell Publishing Ltd., 9600 Garsington Road, Oxford OX4 2DQ, UK and 350 Main Street, Malden, MA 02148, USA.

leaves many questions open, making it impossible to reproduce his experiments. Regardless, in the extreme case of an omniscient infant who knows exactly where the target object is no matter where she is looking, it makes perfect sense to predict that gaze following will not emerge. This is because, similar to the situation discussed above, the caregiver's looking behavior does not provide any additional information about the target location. Gaze following would have no utility.

More fundamentally, however, Csibra has extended our model, and now blames the original model for what are cast as implausible predictions. By contrast, our own recent extension of the present model, which also introduces peripheral vision (Jasso *et al.*, 2005), appears to make very sensible predictions. For example, it correctly captures the stage-wise development of gaze following skills to targets within and outside of the infant's visual field (Butterworth & Jarrett, 1991). This suggests a problem with Csibra's extension, not the original model.

Richardson and Thomas (R&T) argue that architectural assumptions, such as the one about separate when- and where-agents in our model, can form a key aspect of a model. We agree with this general view but would maintain that in our specific case, the choice was not crucial to the functioning of the model. R&T also argue that our model may be viewed as a 'realistic innately modular' system. We are sympathetic to this view but we would like to emphasize that our model, as Gergely Csibra's commentary has underscored, learns different kinds of behaviors if 'raised' in different environments. Thus, the *function* of our 'modules' is not pre-specified but emerges through a learning process that depends crucially on the interaction with the environment.

We are thrilled to see that Richardson and Thomas, following in the footsteps of Williams and Dayan (2005), have used our model to explore the potential effects of Attention Deficit/Hyperactivity Disorder (ADHD) on the emergence of gaze following. This leads R&T to a discussion of multiple causality, which is very useful. Several different manipulations of a model might all produce a similar deficit in a particular behavior metric. In that case, it is important to simultaneously consider other metrics of behavior in order to distinguish different potential causes of a deficit. In our paper, for example, we found that two manipulations to the reward structure of the model mimicking autism and Williams syndrome could both lead to similar deficits in our gaze following metric, the GFI. At the same time, however, other aspects of the model's behavior were vastly different in these two cases, as can be seen by, for example, looking at the model's behavior at a microscopic level (compare Figure 3 of our paper). We view it as a great strength of computational models that they can make specific predictions about complex patterns of differences in multiple behavior metrics.

References

Butterworth, G.E., & Jarrett, N. (1991). What minds have in common in space: spatial mechanisms serving joint visual attention in infancy. *British Journal of Developmental Psychology*, **9**, 55–72.

Csibra, G. (2006). Blind infants in random environments: further predictions. *Developmental Science*, **9** (2), 148–149.

Jasso, H., Triesch, J., & Teuscher, C. (2005). A reinforcement learning model explains the stage-wise development of gaze following. Poster presented at the 12th Joint Symposium on Neural Computation, JSNC 2005, Los Angeles, CA.

Moore, C. (2006). Modeling the development of gaze following needs attention to space. *Developmental Science*, **9** (2), 149–150.

Richardson, F., & Thomas, M.S.C. (2006). The benefits of computational modelling for the study of developmental disorders: extending the Triesch *et al.* model to ADHD. *Developmental Science*, **9** (2), 151–155.

Triesch, J., Teuscher, C., Deák, G.O., & Carlson, E. (2006). Gaze following: why (not) learn it? *Developmental Science*, **9** (2), 125–147.

Williams, J.O.H., & Dayan, P. (2005). Dopamine, learning, and impulsivity: a biological account of ADHD. *Journal of Child and Adolescent Psychopharmacology*, **15** (2), 160–179.

Received: 13 December 2004
Accepted: 2 June 2005

© 2006 The Authors. Journal compilation © 2006 Blackwell Publishing Ltd.