Findings from the developmental investigation of false-belief understanding in preverbal human infants, based on looking time (and other kinds of looking behavior) are relevant to hypotheses about the ontogenetic and the phylogenetic origins of human mindreading capacities. According to Cecilia Heyes (2012), "recent empirical work in comparative psychology, developmental psychology and cognitive neuroscience provides surprisingly little evidence of genetic adaptation, and ample evidence of cultural adaptation."

In particular, she has argued that human infants do not genetically inherit the capacity to *mentalize* (or to read minds). Instead, they learn to read minds (or to mentalize) pretty much in the way they learn to read words by cultural transmission, i.e. via explicit verbal instruction from knowledgeable adults (cf. Heyes and Frith, 2014 and Strickland and Jacob, 2015 for a critique).

On Cecilia Heyes's (2014a, 2014b, 2015) view, explicit mentalizing, implicit mentalizing and even behavior-reading all rest on *domain-specific* social cognitive resources that are definitely not part of the *domain-general* "start-up kit" genetically inherited by human infants. This "start-up kit" is domain-general in the sense that its computations are taken to apply to social and non-social stimuli alike. What she calls *submentalizing* is the human capacity to predict human behavior on the basis of low-level domain-general psychological processes that *simulate the effects of mentalizing in social contexts*. In a nutshell, when faced with social stimuli, humans are genetically predisposed to submentalize.

According to Heyes's interpretation, the "dot perspective" study by Samson and colleagues (2010) shows that submentalizing is at work, not just in human infants, but in human adults as well (Heyes, 2014b, Frith and Heyes, 2014). In the dot perspective study, adult participants see an avatar in the middle of a room with two red dots on two of the three walls. The avatar can either see the same number of dots as participants or one less. When participants are asked how many dots they can see, they are slower to answer and make more mistakes if what they see is different from, rather than the same as, what the avatar can see. Samson et al. (2010) argue that this *altercentric* effect is evidence that human adults automatically compute another's visual perspective. However, Santiesteban et al. (2014) also found evidence for the altercentric effect when the avatar had been replaced by an arrow. They argue that this is evidence that processes of (non-social) attentional orientation (i.e. particular instances of domain-general processes of submentalizing) can simulate the effects of mentalizing (or visual perspective-taking) in social contexts.[1]

One particularly important test of Heyes's hypothesis that humans are genetically prepared to submentalize in the face of social stimuli is her *low-level perceptual novelty* hypothesis put forward to explain many of the findings based on preverbal human infants' looking behavior, which, according to other psychologists, are evidence of early false-belief understanding.[2] On Heyes's (2014a) approach, far from reflecting infants' expectations about an agent's mental states or behavior, the infants' looking behavior reflects the degree to which the low-level properties (the colors, shapes and movements) of test events depart from the low-level properties of the earlier events encoded and remembered by the infants.

In effect, her account of the infant data rests on the interplay between two kinds of domain-general cognitive processes. On the one hand, Heyes's low-level perceptual novelty account commits her to the assumption that infants can only encode a highly *restricted* subset of the full set of properties (the *low-level* properties) that are instantiated by both the test events and the earlier events perceived by the infants. On the other hand, Heyes appeals to the phenomenon of so-called *retroactive interference*, whereby the infants' perception of some of the specific *later* events *impairs* their *memory* of some of the *immediately preceding* events in the sequence.[3]

As Heyes (2014a, p. 647) rightly says about the infant data, “the devil is in the details.” The question is: can the details of her own account perspicuously explain the relevant data? In what follows, I will proceed in two steps: I will first review her account of findings based on Onishi and Baillargeon’s (2005) study and ask whether her account is consistent. Secondly, I will argue that findings based on a study by Kovács, Téglás and Endress (2010) fail to support one crucial component of her account (retroactive interference).

1. The water-melon, the green box and the yellow box

Onishi and Baillargeon’s study involves three temporal stages: a set of familiarization trials, a set of belief-induction trials, and a pair of test trials. In the first of three successive familiarization trials, 15-month-olds first saw a melon toy surrounded by a pair of opaque boxes (a green and a yellow box) against the background of a closed window. After the window was opened, the infants saw a human (female) agent manipulate the toy before placing it into the green box with her right hand. In the two following familiarization trials, they first saw the two opaque boxes against the background of the closed window. Then the window was opened and they saw the human (female) agent reach into the green box with her right hand (cf. Figure 1).

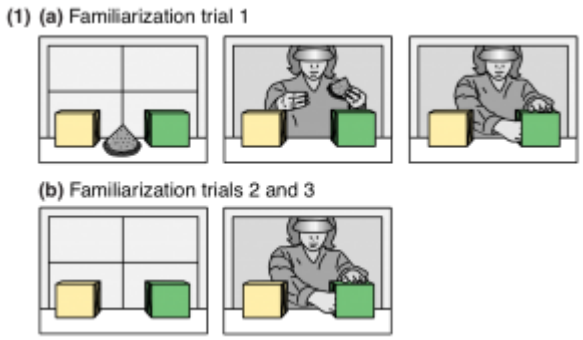

Figure 1 – From Onishi & Baillargeon (2005)

In a series of four distinct belief-induction trials generating a pair of true-belief (TB) conditions and a pair of false-belief (FB) conditions, the infants saw the toy either move (by its own self-propelled motion) from the green to the yellow box or not, either in the presence of the agent or not (cf. Figure 2). In the TB-green condition, the yellow box moved towards the green box and back to its initial position (as indicated by the arrow in Figure 2a), but the toy did not move and the agent was present. In the TB-yellow condition, the toy moved from the green to the yellow box in the agent’s presence (as indicated by the arrow in Figure 2b). In the FB-green condition, the toy moved to the yellow box in the agent’s absence (cf. Figure 2c). In the FB-yellow condition, the toy first moved from the green to the yellow box in the agent’s presence and moved back into the green box in the agent’s absence (cf. Figure 2d).

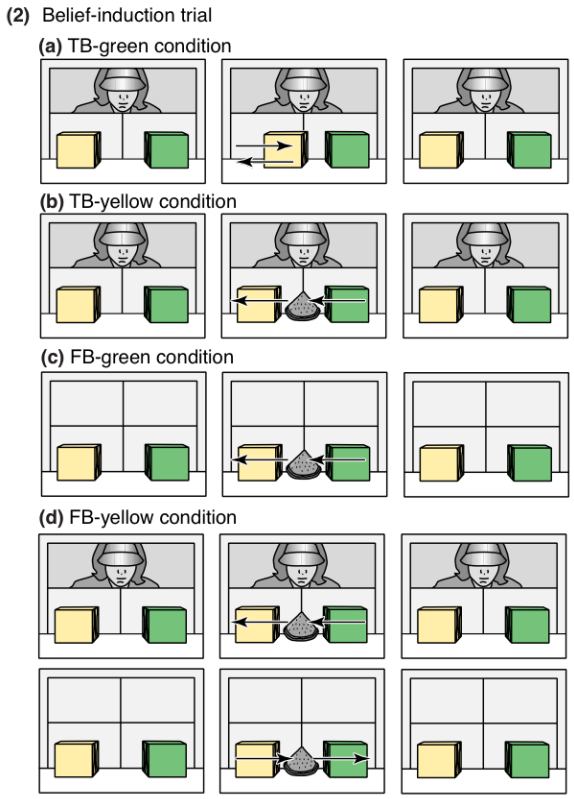

Figure 2 – From Onishi & Baillargeon (2005)

Finally, in the test trials, the infants first saw the two opaque boxes at their previous location against the background of the closed window. The window opened and they saw the agent reach either for the green box with her right hand (green test event) or for the yellow box with her left hand (yellow test event) (cf. Figure 3).

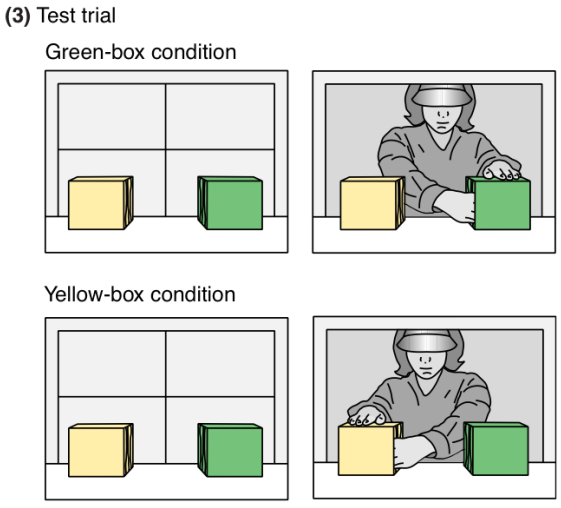

Figure 3 – From Onishi & Baillargeon (2005)

Onishi and Baillargeon found that infants in the TB-green condition looked reliably longer at the yellow rather than at the green test event. Infants in the TB-yellow condition looked reliably longer at the green rather than at the yellow test event. Infants in the FB-green condition looked reliably longer at the yellow rather than at the green test event. Infants in the FB-yellow condition looked reliably longer at the green rather than at the yellow test event. According to Onishi and

Baillargeon's mentalistic interpretation, infants looked reliably longer either when the agent reached to the empty location with a true rather than a false belief or when she reached to the toy's actual location with a false rather than a true belief.

Heyes first asks: why did infants in both the TB-green and the FB-green conditions look reliably longer at the yellow than at the green test event? On the mentalistic account, the answer to this question is: because infants expected the agent to reach for the box in which she believed (truly or falsely) the toy to be. Of course, this cannot be Heyes's non-mentalistic answer.

Heyes's own answer is two-tiered. On the one hand, she argues that to infants in the TB-green condition (in which the toy stayed in the green box), the yellow test event must have looked perceptually more novel than the green test event, relative to their *familiarization* experience (in which on three repeated occasions they saw the agent reach for the green box). On the other hand, she must explain away the fact that infants in the FB-green condition saw the toy move to the yellow box (in the agent's absence). The way Heyes cancels the effect of infants' seeing the toy move to the yellow box in the FB-green condition is by appealing to the phenomenon called *retroactive interference*. In the FB-green condition, the infants see the toy move to the yellow box, *in the agent's absence*. Heyes argues that the agent's unexpected reappearance in test event retroactively interferes with infants' memory of the immediately preceding event in the FB-green condition: "their memory for this event was impaired because it was immediately followed by a salient distractor event – the unexpected reappearance of the agent at the beginning of the test phase." In effect, she argues that retroactive interference cancels the difference between the TB-green and the FB-green condition. So in both cases, the yellow test event must have looked to infants perceptually more novel than the green test event, relative to their familiarization experience.

Secondly, Heyes asks: why did infants in the TB-yellow and the FB-yellow conditions look reliably longer at the green than at the yellow test event? On the mentalistic account, the answer to this question is: because infants expected the agent to reach for the box in which she either truly or falsely believed the toy to be.

Here again, Heyes's answer based on her low-level perceptual novelty account is two-tiered. On the one hand, she argues that in the TB-yellow condition, "after familiarization and before the test, these infants saw an event (movement of the toy-shape towards yellow), that was visually similar to the yellow test event (movement of the agent-shape towards yellow), and therefore reduced the novelty of the yellow test event." As a result, these infants looked longer at the green than at the yellow test event. On the other hand, Heyes must also explain away the fact that in the FB-yellow belief-induction trial, after seeing the toy move to the yellow box in the agent's presence, the infants saw the toy move back to the green box in the agent's absence. Here again, Heyes appeals to retroactive interference: she hypothesizes that the unexpected reappearance of the agent at the beginning of the test phase impairs infants' memory of the immediately preceding event whereby the toy moved back to the green box in the agent's absence. As a result, the difference between the TB-yellow and the FB-yellow conditions vanishes: in both cases, what matters is that after familiarization and before test, the infants saw "an event (movement of the toy-shape towards yellow), that was visually similar to the yellow test event (movement of the agent-shape towards yellow), and therefore reduced the novelty of the yellow test event."

I now want to call into question the internal consistency of Heyes's perceptual novelty account of Onishi and Baillargeon's findings. On the one hand, her low-level perceptual novelty account can only be satisfied if the novelty generating the surprise is present at a low representational level "where the events witnessed by the infants are represented as colors, shapes and movements, rather than as actions on objects by agents." For example, it is a critical assumption of the low-level perceptual novelty account that to infants the movement of the *toy* to the *yellow* box (in the TB-

yellow belief-induction trial) is visually *similar* to the yellow test event whereby the *agent* reaches for the *yellow* box. In the TB-yellow belief induction trial, infants see the toy move to the yellow box; they see the agent watch the movement of the toy; but they only see the agent' head, not her full upper body parts. In the yellow test event, however, they see the agent reach for the visible yellow box with her fully visible left arm, but they do not see the toy. So it is crucial to the low-level perceptual novelty account that infants be blind to the many differences between the event of the toy moving to the yellow box in the TB-yellow belief-induction trial and the yellow test event of the agent's reaching for the yellow box.

But on the other hand, Heyes appeals to retroactive interference in both the FB-green and the FB-yellow conditions. So infants must be vulnerable to the relevant instance of retroactive interference whereby the *agent*'s unexpected reappearance in the test event impairs their memory of the immediately preceding event in which the agent was absent. Only if the infants can encode the property of *being an agent* (capable of executing e.g. reaching arm movements) could they be vulnerable to relevant instances of retroactive interference. But according to the low-level perceptual novelty account, infants should not be able encode the property of being an agent. (Remember: by Heyes's own criteria, even *behavior-reading* rests on domain-specific cognitive mechanisms that are unavailable to human infants who are limited to domain-general cognitive resources.) So there is a conflict between the two components of her non-mentalistic explanation of the findings reported by Onishi and Baillargeon (2005).

2. The Blue Smurf and the ball

Heyes (2014a, 2014b) has also tried to explain away the famous Blue Smurf studies by Kovács, Téglás and Endress (2010). In these studies with adults and infants, participants watch one of four versions of a video. In all four conditions, an agent (a blue Smurf) places a ball on a table in front of an occluder and then the ball rolls behind the occluder. From then on, the four different conditions diverge: participants see the ball either stay behind the occluder or leave and they see the agent leave the scene either before or after the ball reaches its final location. Finally, in all four conditions, the agent comes back, the occluder is lowered in his presence (cf. Figure 4).

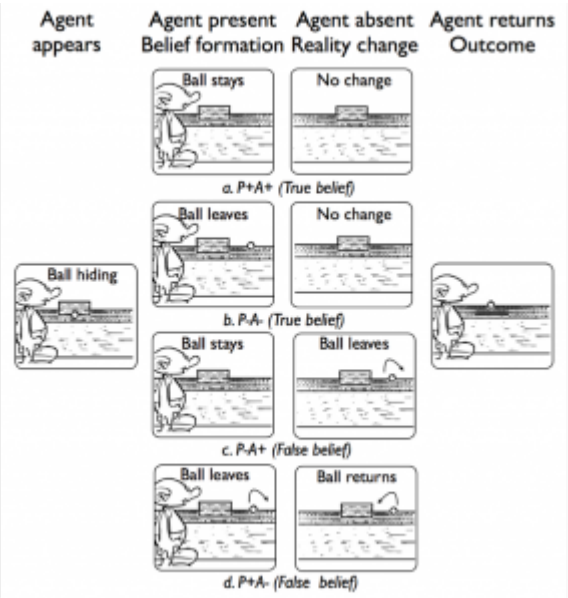

Figure 4 – From Kovács, Téglás and Endress, (2010)

In the adult study, participants were instructed to press a button as fast as possible in the final stage

when the agent is back, if they saw the ball when the occluder was lowered down (which was the case 50% of the time in all four conditions). Kovács and colleagues measured participants' reaction times in performing this task. Not surprisingly, they found that adults were significantly faster to respond in the P+A+ condition (when both participants P and the Smurf agent A expected the ball to be behind the occluder) than in the P-A- condition (when neither participants P nor the agent A expected the ball to be behind the occluder). The surprising finding was that they were also significantly faster in the P-A+ condition (when A falsely expected the ball to be there, but participants did not) than in the P-A- condition. Kovács and colleagues argue that this surprising finding is evidence that participants automatically computed the content of the agent's false belief.[4]

Heyes (2014b, p. 137) has offered a non-mentalistic alternative explanation of the adults' findings based on retroactive interference. In both the P-A- and the P+A+ conditions, the Smurf is present during the last event that is relevant to the participants' own expectations about the location of the ball. However, in the P-A+ condition, the Smurf is absent during the last event that is relevant to the participants' own expectations about the location of the ball, i.e. when the ball finally leaves the scene. Heyes argues that the "perceptually salient" reappearance of the Blue Smurf in the final stage of condition P-A+ (when the occluder is being lowered) is likely to have impaired participants' memory of the immediately preceding event (i.e. the last motion of the ball) by a process of retroactive interference. Heyes's non-mentalistic account makes a straightforward prediction: in the P+A- condition, when participants see the ball finally move back behind the occluder, the Blue Smurf is also absent. According to Heyes's account, the salient reappearance of the Blue Smurf when the occluder is lowered should also impair participants' memory for the immediately preceding event where the ball returns behind the occluder. So Heyes's account predicts that participants should not expect the ball to be behind the occluder and therefore be significantly slower in the P+A- condition than in the P+A+ condition. But they are not.

The same critique extends to Heyes's (2014a) account of the infant study, in which Kovács and colleagues used infants' looking time and found that when in the final stage, the Smurf is back, the occluder is lowered down and *there is no ball*, 7-month-olds look longer in the P-A+ than in the P-A- condition. Kovács and colleagues argue that this is evidence that infants' looking time is influenced by their computation of the Smurf's false expectation that the ball should be there. Heyes on the other hand surmises that in the P-A+ condition, the Smurf was away when the infants last saw the ball leave the scene. So she hypothesizes that far from reflecting infants' computation of the Smurf's false belief, the Smurf's unexpected reappearance impaired the infants' memory of this last event by retroactive interference. As a result, infants looked longer in the P-A+ condition than in the P-A- condition because they had forgotten the ball's last motion and expected the ball to be there.

This account predicts conversely that in the P+A- condition, infants should also forget the last event whereby the ball rolled back behind the occluder. If so, then they should *not* expect the ball to be there and should *not* look longer in the P+A- condition than in the P-A- condition upon finding out that there is no ball. So far as I know (but I may be wrong), this comparison has not been tested. But it would be quite interesting to test Heyes's prediction that infants should not look longer in the P+A- than in the P-A- condition. On the assumption that the mechanisms that underlie the adults' responses are the signature of the mechanisms that underlie the infants' responses,[5] I would expect the infants' looking time to be consistent with the adults' reaction times. Since adults were faster to detect the ball when it was there in the P+A- than in P-A- condition, I would conversely expect the infants to look longer when there is no ball in the P+A- condition than in the P-A- condition.

In this short piece, I have looked at some of the details of Cecilia Heyes's non-mentalistic account of

a pair of developmental findings that have been taken by some psychologists as evidence for false-belief understanding in human infancy. I have proceeded in two steps: first, I have called into question the consistency of her account. Secondly, I have drawn attention to evidence that putatively clashes with her appeal to the phenomenon of retroactive interference.

What is distinctive of Heyes's domain-general approach to human social cognition is that it refreshingly purports to break away from the common assumption that the major alternative to the hypothesis that human infants are endowed with implicit mindreading capacities is that they are endowed with behavior-reading capacities. She explores instead the hypothesis that human infants are genetically prepared to submentalize, i.e. to apply to social stimuli domain-general low-level perceptual processes that simulate the effects of mentalizing in social settings. If submentalizing were part of the genetic tool-kit of human social cognition, then findings based on implicit false-belief tasks in human infants should be explainable by Heyes's low-level perceptual novelty approach. Conversely, failure of the low-level perceptual novelty account to perspicuously explain the infant findings casts doubt on Heyes's hypothesis that humans are genetically prepared to submentalize in the face of social stimuli.[6]

References

Heyes, C. (2012) Grist and mills: on the cultural origins of cultural learning. *Philosophical Transactions of the Royal Society* B2012, 367, 2181-2191.

Heyes, C. (2014a) False belief in infancy: a fresh look. *Developmental Science*, 17:5, 647–659.

Heyes, C. (2014b) Submentalizing: I Am Not Really Reading Your Mind. *Perspectives on Psychological Science*, 9(2) 131–143.

Heyes, C. (2015) Animal mindreading: what's the problem? *Psychonomic Bulletin & Review*, 22:313–327.

Heyes, C. (2017) Apes Submentalise. *Trends in Cognitive Sciences*, 2017; 21(1):1-2.

Heyes, C. and Frith, C. (2014) The cultural evolution of mindreading. *Science*, 344, 1357- 1361.

Kovács, A., Téglás, E. and Endress, A. (2010) The Social Sense: Susceptibility to Others' Beliefs in Human Infants and Adults. *Science*, 330, 1830-1834.

Krupenye, C., Kano, F., Hirata, S., Call, J. & Tomasello, M. (2016) Great apes anticipate that other individuals will act according to false beliefs. *Science*, 354 (6308):110-4.

Krupenye, C., Kano, F., Hirata, S., Call, J. & Tomasello, M. (2017) A test of the submentalizing hypothesis: Apes' performance in a false belief task inanimate control. *Communicative & Integrative Biology*, 10(4)
http://dx.doi.org/10.1080/19420889.2017.1343771

Onishi, C. and Baillargeon, R. (2005) Do 15-month-old infants understand false beliefs? *Science*, 308(5719), 255–258.

Phillips, J., Ong, D.C., Surteees, A.D.R., Xin, Y., Williams, S., Saxes, R. & Franck, M.C. (2015) A Second Look at Automatic Theory of Mind: Reconsidering Kovács, Téglás, and Endress (2010).

Psychological Science, 26(9), 1-15.

Samson, D., Apperly, I.A., Braithwaite, J.J., Andrews,□B.J., Bodley Scott, S.E. (2010) Seeing it their way: Evidence for rapid and involuntary computation of what other people see. *Journal of Experimental Psychological. Human Perception and Performance*, 26(9), 1255–1266.

Santiesteban, I., Catmur, C., Coughlan Hopkins, S., Bird,□G., Heyes, C. (2014) Avatars and arrows: Implicit mentalizing or domain- general processing? *Journal of Experimental Psychological. Human Perception and Performance*, 40, 929–937.

Scott, R.M., & Baillargeon, R. (2014) How fresh a look? A reply to Heyes. *Developmental Science*, 17:5, 660–664.

Strickland, B. and Jacob, P. (2015) http://cognitionandculture.net/blog/brent-stricklands-blog/why-reading-minds-is-not-like-reading-words

***

[1] One alternative mentalistic possibility is that far from being submentalizers, human adults are *super-mentalizers* who find it hard not to interpret their environment mentalistically. If so, then they might immediately interpret an arrow as a social cue intended by a conspecific to draw their attention to a particular dot.

[2] Krupenye et al. (2016) have reported findings, which they interpret as evidence that non-human apes can represent the contents of others' false beliefs. Heyes (2017) has proposed her submentalizing alternative account of the findings by Krupenye and colleagues (2016). Krupenye et al. (2017) have subsequently tested and offered evidence against Heyes's (2017) proposal.

[3] For an earlier critique of Heyes's (2014a) explanation of the infant data based on her perceptual novelty account, cf. Scott and Baillargeon (2014).

[4] Kovács and colleagues' mentalistic interpretation of their adult study based on reaction times has been criticized by Phillips et al. (2015) on grounds independent from Heyes's own interpretation based on retroactive interference. While Phillips and colleagues' non-mentalistic critique does not extend to the infant study, Heyes's interpretation does.

[5] An assumption Heyes (2014a, 2014b) must accept since she applies retroactive interference to both the infant and the adult studies.

[6] I am grateful to Dan Sperber for his comments.