Source Reliability in the Development of Children's Understanding of Causal Systems

Germaine Symons (g.symons@bbk.ac.uk)
Department of Psychological Sciences, Birkbeck College,
Malet Street, London, WC1E 7HX, UK

Professor Andrew Tolmie (a.tolmie@ioe.ac.uk)
Department of Psychology and Human Development, Institute of Education,
20 Bedford Way, WC1H 0AL

Professor Mike Oaksford (m.oaksford@bbk.ac.uk)
Department of Psychological Sciences, Birkbeck College,
Malet Street, London, WC1E 7HX, UK

Abstract

Children have been shown to discriminate between reliable and unreliable sources from as young as 3 or 4 years old using the selective trust paradigm (Koenig, Clément & Harris, 2004). However, it is unclear whether children are discriminating between informants because they have some epistemic awareness regarding the knowledge of the informants, or because they regard the behaviour of the unreliable informants as bizarre. The current experiment manipulates source reliability in a more naturalistic way (science teacher vs. nursery child), and looks at the effects of age (6-7, 8-9, 10-11 years old) on children's' predictions relating to a familiar causal system - cars on an inclined plane, where height, surface friction, and starting point on the slope, as well as the weight of the car, can be changed. Children are either told unintuitive information coming from a differentially reliable source, or no information, regarding the effect of weight. They are then asked to make predictions regarding how far the car travels. Children in the high reliable condition are more likely to change their predictions regarding the effect of weight following information from the more reliable source. This may occur in older children only, but more research needs to be done.

Keywords: source reliability, testimony, social cognition

Introduction

For children, much of the information regarding the world at large comes via a source, such as testimony from the people around them (e.g. parents, teachers, peers) or other forms of culturally transmitted information (e.g. books, the Internet, television). This is particularly so in school. For example, in class, most of the information a child learns about is given to them either by a teacher, or read in a text book, or other related media. However, they also receive information from their peers, parents, and other sources of information, such as television, and the Internet. These sources of information can be more or less reliable. For example, a teacher is a more reliable source of information than a peer (in most cases!). Understanding the development of how children incorporate testimony from differentially reliable sources into their reasoning about the world can be very useful, as it can help to inform teaching practice.

The selective trust paradigm was developed (see Koenig, Clément & Harris, 2004) to assess at what age children became sensitive to the reliability of the source of information, and whether source reliability affects their understanding of events. Children are introduced to accurate and inaccurate informants, identified as such through their labelling of objects, and then asked questions such as "which informant did they prefer" or "did any of the informants do anything wrong". Using this paradigm, children have been shown to take source reliability into account from as young as 3 and 4 years old, where they show a preference for accurate informants (Koenig & Harris, 2005); a preference for reliability over age (Jaswal & Neely, 2006); they can revise their preference if a previously reliable source becomes unreliable (Scofield & Behrend, 2008); take into account relative accuracy (Pasquini, Corriveau, Koenig & Harris, 2007); prefer reliable children over unreliable adults (even though they usually prefer adults as sources of information; Jaswal & Neely, 2006); prefer a consensus among sources (Corriveau, Fusaro & Harris, 2009); and so on.

The fact that children show a preference for the reliable source of information in many different contexts has led to the claim that children, from as young as three or four years old are showing epistemic awareness regarding the knowledge of the informants (for example, see Koenig & Harris, 2007). However, they could be basing their responses purely on the output of the informant, without making any inferences regarding the knowledge of the informant. The oddness of the behaviour of the inaccurate informant, mis-naming common everyday items, may be enough to explain the preference for the accurate informant (Lucas & Lewis, 2010). The selective trust paradigm does not discriminate between the two explanations.

Furthermore, if one wants to gain an understanding of how children incorporate and use information from differentially related sources into their reasoning about the world, such that the understanding can be used to inform teaching practice, then it is desirable to use paradigms that

are familiar to the child, and manipulate source reliability in a naturalistic way. One can then be more confident that the differences observed are likely to exist in real life and not just in the confines of the scientist's lab.

Adults also appear to take source reliability into account - research that takes a Bayesian approach has found that people rate arguments from more reliable sources as being more convincing (e.g. Hahn, Oaksford, & Bayindir, 2005; Hahn, Harris, & Corner, 2009). Here, source reliability is manipulated in a much more naturalistic way. People are asked to evaluate information from sources they are likely to have come across in their everyday lives (such as a research body vs. TV interview in Hahn, Oaksford, & Bayindir, 2005; or information that comes from journal article vs. an advertisement in Hahn, Harris, & Corner, 2009).

Using these more naturalistic paradigms as inspiration, we used a causal system that is familiar to children - cars on an inclined plane, where surface, height, starting point on the slope and weight of the car can change. Not only do children personally experience the effects of these different variables in their everyday lives, but they also learn about them in primary school science classes from a very young age (in the UK, forces and motion are covered in Key Stage 1, which covers ages 5 – 7 years old).

To establish when and how children use information from differentially reliable sources in their reasoning about this system, we gave them (unintuitive) information on how the system works, from a more (a science teacher) or less (a nursery child) reliable source, and asked them to make predictions regarding how far the car travels. That is, we tell them that weight does not affect how far the car travels, and see whether and how this affects their predictions and explanations relating to weight. A science teacher was chosen as the reliable source as even young children (3-4 years) have been shown to differentiate expertise, as long as the nature of the expert is familiar (Lutz & Keil, 2002). Furthermore, children learn about forces and motion in science class (where the motion of objects on inclined planes is often used to illustrate the point), so science teachers should be seen as an expert - more likely to know about a similar causal system. In contrast, a nursery child (in the UK, children in the school nursery class are aged 3-4 years) was chosen as they were younger than all the participants, and therefore likely to be deemed not an expert when providing information regarding the causal system in question.

This was done with three age groups (6-7 years, 8-9 years, 10-11 years), so that we can examine how these predictions and explanations changed over time.

If children preferentially use information from reliable sources to inform their predictions, and their explanations, in their reasoning about the system, it suggests they regard the reliable source as having informed knowledge that pertains to the system under question.

Method

Participants

Three year groups from a middle class Roman Catholic primary school in London. Year 2: N=20, mean age = 7.16 years (SD = 0.32, range = 6.56 – 7.48); Year 4: N=20, mean age = 9.12 years (SD = 0.28, range = 8.54 – 9.47); and Year 6: N=19, mean age = 11.26 years (SD = 0.26, range = 10.88 – 11.84). Months reported as percentage of year. The data from one Year 6 child was removed as the child failed to properly participate in the experiment.

Design

This study employed a 3(age) x 3(source reliability) x 3(time of testing) mixed model design. Age and source reliability were between subject factors, and time of testing was a within subject factor.

The children were from Year 2 (aged 6-7 years), Year 4 (aged 8-9 years), and Year 6 (10-11 years). There were three source reliability groups. Children were either told relevant information about the system from a science teacher, or a nursery child, or they were given no information. The relevant information the children were told is as follows: "weight does not make a difference to how far the car travels". Most children (and adults) think that weight does affect how far the car travels, and their causal predictions reflect this. Data was collected, 1) before the children were given any information about the game, 2) after the children had either received the relevant information, reportedly from the different sources, or they heard no information, and 3) after children had seen for themselves that the relevant information was true.

The children were asked to make predictions regarding how far they think the car will travel (prediction), how sure they were (degree of confidence), and why they thought that (explanation). They did this for the high and low position for each of the variables in turn (in total they provided responses for 8 different set ups - high/low for height, surface friction, starting point, and weight). The high/low question order was alternated where approximately half of the questions started with the high position of a particular variable, and half started with the low position.

Materials

The main apparatus, the cars on slopes game (see Fig. 1), consisted of three slopes in a row, with different surface frictions (smooth, medium and rough surfaces). Each slope could be raised to three heights (high, medium and low height) and start at three points on the slope (high, medium and low starting point). The car varied in weight (light, medium and heavy), altered by adding to the car one, two or three little beanbags. There was a fourth slope which always remained in the standard position - medium height, medium surface friction, and medium starting point. This was accompanied by the medium weight car that remained on the side of the track as a reminder to the children as to

where the car would land, when the apparatus was set in the standard position.

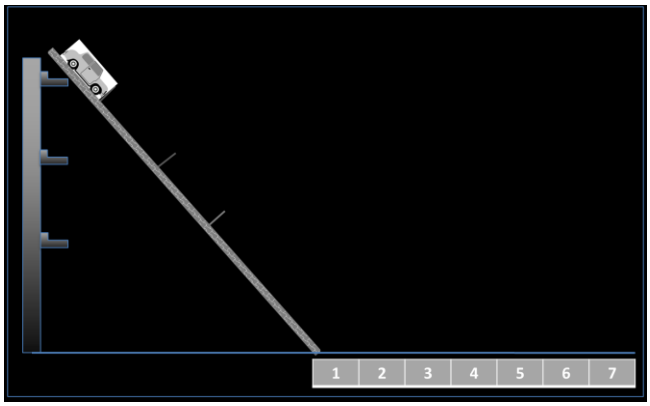


Figure 1: The cars on slopes game

The British Picture Vocabulary Scale (BPVS), and a verbal fluency task (where children were asked to name as many animals as they could in one minute) were used to get a measure of children's language skills.

A stick scale with seven sticks of ascending height was used to get an indication of how sure the children were of their predictions.

Procedure

In the first session, children participated in the two language tasks, the BVPS and the verbal fluency task. They were then told how the cars on slopes game worked, and got to see how far the car would travel in the standard position. After that they played freely on the game for six goes. Baseline data was then collected, where children were firstly shown the standard position, and then the high and low set up for each of the variables in turn. When making their responses, they were reminded that everything but the variable in question stays the same. They were asked how far they thought the car would travel at that particular position, how sure they were, and why did they think that. This concluded the first session.

At the beginning of the second session, the experimenter either told the child that a (science teacher/nursery child) told her that weight did not make a difference to how far the car traveled, and asked what they thought, or just asked the child what (s)he thought the effect of weight was. After this, the experimenter reminded them of where the car would land when the game was on the standard set-up. Then they were shown different set ups again and asked how far they thought the car would travel, how sure they were and why did they think that. This same set of data was collected again after children did a fair test on weight of the car and saw that the car landed in the same box regardless of how heavy it was. Finally to check our source reliability manipulation worked, children were asked to say on a scale of 1 to 10, how likely they thought a science teacher and nursery child would be right if you asked them a question.

Results

Only the prediction data for weight was analysed and reported in this paper.

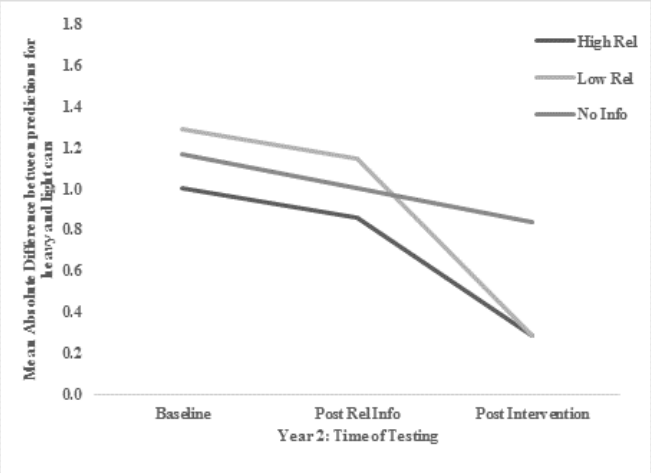
Nearly all children appeared to think that weight had an effect on how far the car travelled at baseline (Yr 2 – 95%; Yr 4 – 90%; Yr 6 – 78%), although the nature of the effect depended on age. Year 2 and 4 predicted that the lighter car would travel further, and year 6 predicted that the heavier car would travel further, $\chi^2(2) = 6.61$, $p = 0.037$. This difference has been observed previously possibly suggesting a declining salience of the horizontal dimension as the children get older (Hast, 2014).

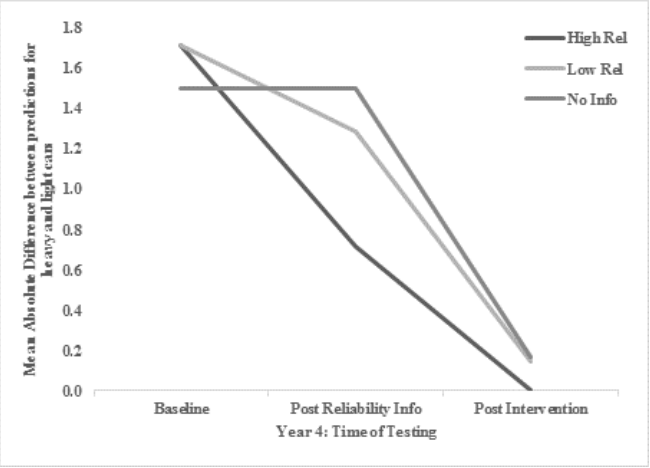
To establish that our source reliability manipulation worked, we asked children on a scale of 1 to 10, how likely they thought a science teacher, and nursery child, would be right if you asked them a question. All children rated the science teacher as more reliable than a nursery child (Yr 2: Sci. Teacher = 0.97, Nur. Child = 0.39; Yr 4: Sci. Teacher = 0.94, Nur. Child = 0.41; Yr 6: Sci. Teacher = 0.91, Nur. Child = 0.44). This difference was statistically significant, $F(2,55) = 562.87$, $p<0.001$. There were no effects of age $F(2,55) = 0.007$, $p = 0.99$.

The absolute difference between the predicted distance travelled for heavy and light car was calculated ("difference score") - if children come to believe that weight does not have an effect on how far the car travels, this difference should decrease to 0.

Figure 2 shows mean absolute difference between children's predictions for the heavy and light car at baseline, either after children have received relevant information from differentially reliable sources or they received no information, and after they had intervened on the system and seen for themselves that the relevant information was true, for each age group. As can be seen Year 4 and 6 children appear to show a reduction in prediction difference after they received relevant information regarding the effect of weight, and that this reduction appears to be greatest for the high reliable source. In contrast, the Year 4 and 6 children who did not receive any relevant information did not change their prediction at all. However, Year 2 children do not appear to discriminate in the same way, with children in the high and low reliable source conditions, *and* in the no information conditions all slightly reducing their predictions after they either receive information from differentially reliable sources or they received no information. Furthermore, although most children decreased their prediction differences to 0 after they had seen that the relevant information was true, some Year 2 children did not. Instead, they persisted in making predictions that suggested they think weight has an effect on how far the car travels.

The data was then analysed using a 3 (age) x 3 (source reliability) x 3 (time of testing) mixed model ANOVA, with age and source reliability as between subjects variables, and time of testing as a within subjects variable.





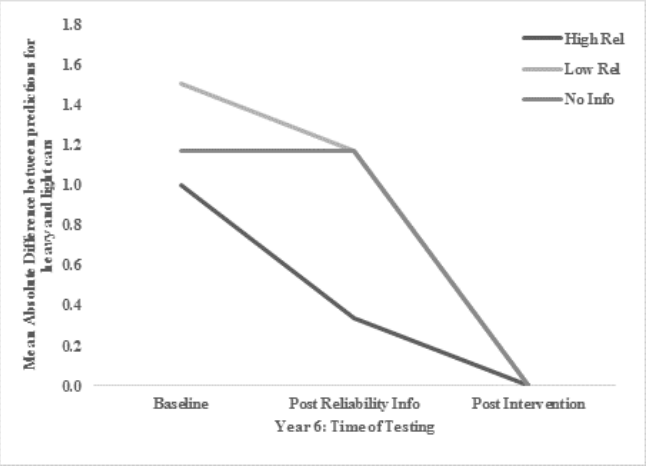


Figure 2: shows the mean absolute difference between predictions for heavy and light cars for each year group, in the low and high reliability and no information conditions, at each time of testing.

There were significant differences in the difference score between times of testing $F(2,98) = 53.85$, $p < 0.001$, $\eta_p^2 = 0.52$. The difference score decreased after children were given the information from a differentially reliable source ($p = 0.018$).

It also decreased after children had intervened on the system and witnessed the fact that the relevant information, weight does not affect how far the car travels, was true ($p < 0.001$).

There was no effect of age ($p = 0.18$, $\eta_p^2 = 0.067$). However, there was an interaction between age and time of testing, $F(4, 98) = 2.54$, $p = 0.045$, $\eta_p^2 = 0.17$ (see Figure 3). To decompose this interaction, we used a Bonferroni corrected alpha level of 0.017. Post hoc between subject one-way ANOVAs show that there is no difference between the three Year groups at baseline, $F(2,55) = 3.29$, $p = 0.045$, $\eta^2 = 0.11$ and post reliability information, $F(2,55) = 0.50$, $p = 0.61$, $\eta^2 = 0.02$; but there was a difference between year groups post intervention, $F(2,55) = 4.67$, $p = 0.013$, $\eta^2 = 0.15$. Year 2 did not decrease the difference between high and low weight predictions as much as Year 4 and Year 6 (who approached 0). Pairwise comparisons showed a significant difference between Year 2 and Year 6 ($p = 0.017$), and an almost significant difference between Year 2 and Year 4 ($p = 0.076$) at this level.

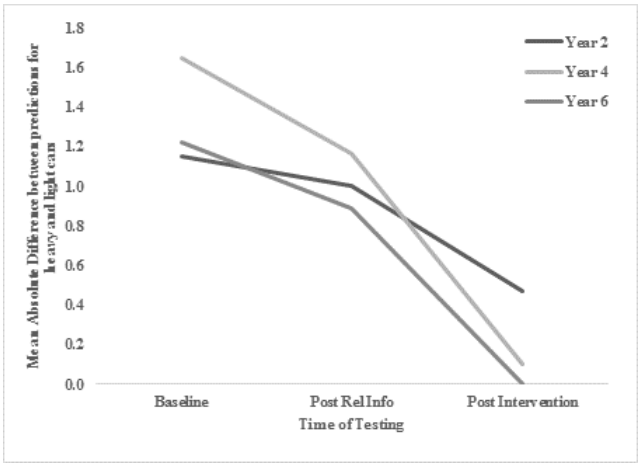


Figure 3: shows the mean absolute difference for between predictions for heavy and light cars for each year group, at each time of testing.

The main effect of source reliability narrowly missed significance, $F(2, 49) = 2.82$, $p = 0.069$, $\eta_p^2 = 0.10$. Although children in the high reliability condition decreased their prediction difference more often, compared with the other conditions, the mean difference score did not reflect this as the variance between predictions of distance travelled for heavy and light was quite large (range = 1-4), and cancelled out the effect of the high reliability source.

To avoid this problem, the frequency with which children decreased, or did not decrease their predictions regarding the effect of weight was calculated, collapsing across age. As can be seen in Table 1, children are more likely to decrease their prediction regarding the effect of weight when they receive information from a high reliable source,

compared with when they receive information from a low reliable source, or received no information. According to the likelihood ratio statistic, this association is statistically significant, $G^2(2) = 6.02$, p = 0.049. To further understand the nature of the association, we partitioned the likelihood ratio statistic, comparing the frequency of reduction (or not) in the low reliability and no information conditions and discovered no significant association, $G_1^2(1) = 0.76$, p = 0.38. Given this, the low reliability and no information conditions were collapsed into one and compared with the high reliability condition, which showed a significant association, $G_2^2(1) = 5.26$, p = 0.022. This suggests that children who received information from a high reliability source were more likely to decrease their predictions regarding the effect of weight than children who either received information from a low reliability source, or no information.

Table 1: Frequency with which children's predictions regarding the effect of weight decreased, or did not decrease.

Source Reliability	Reduction	No Reduction
High	12	8
Low	7	13
No Info	4	14

Discussion

Overall, many children, on hearing relevant (unintuitive) information that pertains to the system under question will change their prediction to conform to the received information. They decreased the difference between predictions regarding how far the car would travel for the heavy and light car, following being told that weight does not affect distance travelled. When the children then witness evidence that the information is 'true' (observing that the light, medium and heavy weight car all land in the same box) nearly all children altered their predictions to conform to what they had just observed.

This reduction following receipt of information from differentially reliable sources appears to be driven by children in the high reliability source condition. These children, who received the information that weight did not affect how far the car travelled from "a science teacher" were more likely to decrease the difference in their predictions for how far the heavy and light car would travel, compared with receiving information from "a nursery child" or receiving no information. This suggests that at least some of the children have some kind of epistemic awareness of what a science teacher (and a nursery child) might know, and understand the implications as it relates to the causal system under question.

Furthermore, it may be that there are age related differences in how likely it is that a child incorporates the information they hear from differentially reliable sources into their reasoning about the world. Although not significant, the Year 2 children (aged 6-7) do not appear to be paying as much attention to the reliability of the source as the older children. Increasing the sample size (there were only 6-7 children per condition) may shed some light on whether this trend in the data is real.

The source reliability effect was very small, occurred even though the information regarding source reliability was delivered by the experimenter (…a *science teacher/nursery* child told me that…"). In future studies it would be useful to manipulate source reliability in a more realistic way, possibly by videoing (people who look like) science teachers and nursery children talking about what they think is the effect of weight. Making the reliability of the different sources more conspicuous (particularly for younger children) may strengthen the size of the source reliability effect. Using video to demonstrate reliability is a standard procedure in the selective trust literature (e.g. Corriveau & Harris, 2009; Jaswal & Neely, 2006; Pasquini et al., 2007).

As well as not appearing to pay attention to source reliability, Year 2 children also appeared to ignore observational evidence, where some of them continued to make predictions that conform with their intuitive understanding of how the causal system works (that weight has an effect). This may be because some younger children are less able to inhibit their intuitions regarding the effect of weight. It is well known that the younger children are the more difficult they find it to inhibit their intuitive responses (Dempster & Corkhill, 1999). Furthermore, even experts appear to have an intuitive response that they then need to inhibit, when faced with problems that do not align with one's intuitive beliefs about the world. They just happen to be better at inhibiting intuitive responses than novices (Masson, Potvin, Riopel, & Foisy, 2014), so it would be no surprise that the younger children also find it difficult.

When asked, even the youngest children appear to think that science teachers are more likely to get a question correct, compared with a nursery child, so these children clearly have some understanding of source reliability and they would no doubt respond similarly to children in the selective trust experiments (e.g. Koenig & Harris, 2005; Jaswal & Neely, 2006; Scofield & Behrend, 2008), if asked questions regarding who they would trust to provide reliable information. However, this does not appear to cause many of them to amend their predictions to concur with information from the reliable vs. unreliable source. As such, it is unclear whether the younger children are showing epistemic awareness regarding the knowledge of the informants (see Koenig & Harris, 2007). It could be that they do have some kind of epistemic awareness relating to the knowledge of the reliable sources, but the implications of the information were not clear to them. Alternatively, it may be that just hearing the pertinent information from a reliable source was not enough to override their intuitions, even if they did understand the implications of the information (bear in mind that even adults commonly think

that weight has an effect on how far the car travels – the intuitions must be very strong). Further research needs to be done to tease out these possibilities.

In future, increasing the impact of the source of information may be useful in understanding the nature of the effect. In this study the children were given second hand information by the experimenter, putatively coming from a (un)reliable source. It would be interesting to see whether changing the mode of delivery of the pertinent information (maybe with a video of a science teacher/child conveying the information) would make a difference to the strength of the effect.

In conclusion, it is clear that at least older primary school children are capable of discriminating between high and low reliable sources, and adjusting their reasoning about the world at large to conform to the information they have heard. However, while younger children could accurately judge reliability of the sources, they failed to fully incorporate this information from these sources into their reasoning about the causal system at hand. The reasons for this are unclear and further research needs to be done.

References

Corriveau, K., & Harris, P. L. (2009). Choosing your informant: weighing familiarity and recent accuracy. *Developmental Science*, *12*, 426-437.

Dempster, F. N., & Corkill, A. J. (1999). Interference and inhibition in cognition and behavior: Unifying themes for educational psychology. *Educational Psychology Review*, 11, 1-87.

Hahn, U., Oaksford, M., & Bayindir, H. (2005). How convinced should we be by negative evidence? *Proceedings of the Annual Conference of the Cognitive Science Society, 27*, 887-892.

Hahn, Ulrike, Harris, A. J., & Corner, A. (2009). Argument content and argument source: An exploration. *Informal Logic*, *29*, 337-367.

Jaswal, V. K., & Neely, L. A. (2006). Adults Don't Always Know Best Preschoolers Use Past Reliability Over Age When Learning New Words. *Psychological Science*, *17*, 757-758.

Koenig, M. A., Clément, F., & Harris, P. (2004). Trust in testimony: Children's use of true and false statements. *Psychological Science*, *15*, 694–698.

Koenig, M. A., & Harris, P. L. (2005). Preschoolers mistrust ignorant and inaccurate speakers. *Child Development*, *76*, 1261-1277.

Koenig, M. A., & Harris, P. (2007). The Basis of Epistemic Trust: Reliable Testimony or Reliable Sources *Episteme*, *4*, 264-284.

Lucas, A. J., & Lewis, C. (2010). Should We Trust Experiments on Trust. *Human Development*, *53*, 167-172.

Lutz, D. J. and Keil, F. C. (2002), Early Understanding of the Division of Cognitive Labor. *Child Development, 73*, 1073–1084.

Masson, S., Potvin, P., & Riopel, M. (2014). Differences in brain activation between novices and experts in science during a task involving a common misconception in electricity. *Mind, Brain, and Education, 8,* 44–55.

Pasquini, E. S., Corriveau, K. H., Koenig, M., & Harris, P. L. (2007). Preschoolers monitor the relative accuracy of informants *Developmental Psychology*, *43*, 1216-1226.

Scofield, J., & Behrend, D. A. (2008). Learning words from reliable and unreliable speakers. *Cognitive Development*, *23*, 278-290.