To catch a liar: The effects of truthful and deceptive testimony on inferential learning

Robert Montague[1], Daniel J. Navarro[2], Amy Perfors[2], Russell Warner[1], & Patrick Shafto[1]
[1] Department of Psychological and Brain Sciences, University of Louisville
[2] School of Psychology, University of Adelaide

Abstract

Much of what people learn is based on the testimony of others, but not all testimony is helpful. This study explores how people deceive and how they deal with deceptive information in the context of a conceptual learning task. Participants play a game in which a learner infers the location of a rectangle based on the testimony of an informant, who is either helpful or deceptive. We investigate the behavior of both informants and learners in this scenario. On the informant level, we demonstrate that people provide different information depending on whether they are helpful or deceptive. Although deceptive informants do lie outright, they more often opt to mislead. From the learner's perspective, we show that people do choose to verify information but no more often when the informant is deceptive. Despite this, we also find that learners are capable of accurately identifying who is deceptive and who is helpful. We conclude by examining common strategies used in the two conditions and their implications in real-world settings.

Keywords: Lies; Pedagogical reasoning; Testimony.

Introduction

Much of what people learn is based on the testimony of others. Unfortunately, not all informants are well intentioned. For example, law enforcement officials must often reconstruct a series of events based only on information obtained from witnesses and suspects, some of whom may attempt to deceive the authorities. For deceptive informants, there are many possible strategies: do they deceive by lying outright or by providing unhelpful information? Similarly for learners, do people attempt to verify a potential deceiver's testimony? How do the results of this interaction affect people's ability to recognize deception?

Most of the research in this area has focused on people's ability to recognize deception using its verbal and nonverbal characteristics (Brandt, Miller, & Hocking, 1982; DePaulo, Stone, & Lassiter, 1985; DeTruck & Miller, 1985; Littlepage & Pineault, 1985; Buller, Strzyzewski, & Hunsaker, 1991; Burgoon, Buller, & Floyd, 2001; Bond & DePaulo, 2006). A typical deception detection study involves participants watching video clips of people either truthfully or falsely describing an experience or opinion. Based only on that information, participants are asked to distinguish the honest statements from the lies. Most people, including police officers, judges, and psychiatrists, perform at close to chance levels (Ekman & O'Sullivan, 1991). These results typically ignore the content of the information and focus on superficial cues to deception in one-off situations.

We are interested in how deceptive interactions play out over time. For example, guilty criminal suspects often try to convince the police of their innocence. In response, police officers verify as many details of their testimony as possible and then decide whether to continue to focus on the suspect or pursue other leads. Analogously, consider a game in which a learner tries to learn a rectangular concept based on clues provided by an informant, who may be deceptive or helpful (see Figure 1). Like the police, the learner can ask the informant or gather information on their own. Like the suspect, the informant can lie or tell the truth, providing helpful or evasive information. To optimize learning, the learner must infer whether the informant is helpful or deceptive in order to decide whether to continue asking for help or gather information on his own. Learners can also verify their information by investigating different points independently.

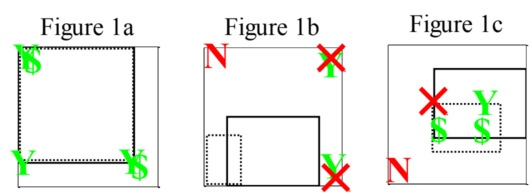

Figure 1: The solid black line represents the rectangle presented to the informant. The task of the informant is to enable the learner to guess the rectangle by providing interior (Y) and exterior (N) points. The learner can also independently explore certain points to verify ($ indicates a verified interior point, X a verified exterior point). The broken black line represents the learner's attempt in that trial to reproduce the informant's rectangle. Figure 1a shows a successful cooperative strategy with verified corner hints. Figure 1b shows deceptive hints that were exposed as lies. Figure 1c shows how hints can be truthful but unhelpful.

Prior studies of inferential learning suggest that an informant's intentions affect what information they provide. In a study using the game described above, informants were instructed to provide helpful examples for their learners (Shafto & Goodman, 2008). These informants provided interior points that marked the rectangle's corners or exterior points that marked its boundaries significantly more often than chance would predict. This suggests that helpful informants do not provide information at random but rather tailor their examples to the concepts they are trying to communicate.

This purposeful sampling is not lost on learners. Children as young as five years old interpret information differently depending on how it is communicated. In a 2009 study, one function of a novel toy was demonstrated by a teacher, by apparent accident, or not at all (Bonawitz, Shafto, Gweon, Chang, Katz, & Shulz, 2009). Children in the latter two conditions explored the toy's other functions much more often than children who saw the purposeful demonstration, suggesting that children in the pedagogical condition inferred that there were no other functions to be found.

How data are sampled also matters when learning words. Xu and Tenenbaum (2007) presented both children and adults with names for novel objects. In one condition, the experimenter picked two other objects as examples; in the other, the participant chose the examples themselves. Participants who were given only one example by the experimenter tended to apply the object's name more broadly. Again, these results suggest that learners appreciate the significance of purposefully sampled information.

Learners can also reason effectively about information from deceptive sources at an early age. In a 2009 study, children were asked to find a piece of candy that had been hidden in one of two boxes (Mascaro & Sperber, 2009). The children watched while a puppet looked in both boxes. After the experimenter explained that the puppet always tells lies, the puppet told the child where the candy was. Four year-old children usually indicated that the candy was actually in the other box. If the children are merely warned that the puppet does not want them to find the candy, comparable performance does not emerge until age six.

In this paper, we contrast learning from teachers and deceivers. Informants in both Bonawitz et al. (2009) and Xu and Tenenbaum (2007) were presented as knowledgeable and helpful. Different issues may arise when the informant has ambiguous intentions. In this paper, we present people with situations similar to those described in Figure 1. Because participants interact by computer, most cues to deception are unavailable to learners and they must rely instead on their informant's testimony. We contrast the information provided by helpful and deceptive informants, focusing on three key questions. First, how does the information provided by helpful and deceptive informants differ? Second, how often do learners verify the information provided, and does this differ based on whether the informant is deceptive or not? Third, how effectively do people recognize when they are being deceived, given only the content of the testimony?

Methods

Participants

One hundred and sixteen students at the University of Louisville participated in exchange for course credit.

Procedure

Pairs of participants were randomly assigned to a cooperative (n = 29) or competitive (n = 29) condition. Each individual was then randomly assigned to the role of informant or learner. Participants were seated at computers on opposite sides of a solid screen. They were then told that they were part of a police investigation of the Rectangle Gang, who got their name from their habit of splitting up the loot from a robbery into bags and burying them in rectangular plots of land.

In both conditions, the learner was told that they would play the role of investigator. They were also told that an informant, a former member of the gang who knew the locations and dimensions of all the rectangular plots of land where money had been hidden, was part of the investigation. Informants were told that they would play the role of the knowledgeable informant.

When the investigator asked for a hint, informants were shown a blue rectangle in a white field. This rectangle indicated the area where the money was buried. Informants were instructed to click on a point in the field, mark it as either inside or outside the rectangle, and then send it to the learner.

The instructions to the informant differed for the cooperative and competitive conditions. Cooperative informants were instructed to be as helpful as possible and that their score would depend on how accurately the learner could reproduce the rectangle's location. In contrast, competitive informants were told to prevent the learner from discovering the true location of the rectangle and that the less accurately the learner could reproduce the rectangle's location, the higher their score would be. Although the accuracy of the learners' drawings was recorded, the game score was not visible during the experiment.

Learners in both conditions were instructed to find the edges of the rectangular plot in which money was buried. A blank white field on the computer screen represented the search area. Learners could either ask the informant for a hint or make exploratory "digs" by clicking anywhere in the field. If they chose to dig, they would "find" either a green $ (representing a point inside the rectangle) or a red X (an exterior point). If they asked for a hint, the informant's response would appear on their field as either a green Y (indicating a place where the informant said that money is buried) or a red N (indicating a place where the informant said that there was nothing). Learners did not know whether their informant was cooperative or competitive but were warned of both possibilities. Learners had five opportunities to gather evidence in any combination of digs or hints for a given field. They were then asked to infer the location of the rectangular plot by clicking and dragging with the mouse. A total of twenty fields were presented.

In the final part of the experiment, learners were asked to rate their informants on a scale from 1 (extremely deceptive) to 21 (extremely helpful). Learners were told that choosing 11 (the midpoint) meant that they did not have an opinion either way. Learners were then asked for a written justification for their rating. A review of these justifications indicated that in two cases one or other of the participants had not understood their instructions. Those data were excluded. In addition, a review of the data indicated that some cooperative informants gave more deceptive hints than helpful ones and some learners made more than five errors (including a verified exterior point or excluding a verified interior point) when drawing their rectangles. The data from these four games were also excluded, leaving twenty-six games in each condition.

Results

Our first question was about how informant type affects the kind of information given. To explore this, we define the information's helpfulness based on how accurately a learner could use it to reproduce the informant's rectangles. For each field, the distance in pixels from each vertex of the informant's rectangle to the corresponding vertex of the learner's rectangle was calculated. These distances were summed to provide a measurement of the learner's error for that rectangle. As Figure 2 demonstrates, learners in the cooperative condition had lower error than those in the competitive condition (cooperative: M = 594.30 pixels, SD = 266.27; competitive: M = 825.12 pixels, SD = 168.78, $t(50) = 3.73$, $p < 0.001$). This suggests that cooperative informants may have provided more helpful information.

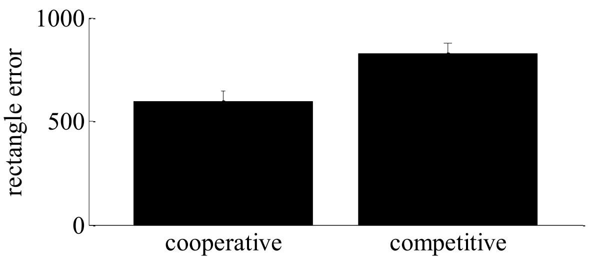

Figure 2: Mean rectangle error by condition. Participants made more errors when faced with a deceptive informant (competitive condition). Error bars represent ±1 SEM.

What information was provided? Comparing deceptive and helpful informants

Why did learners in the cooperative condition perform better? One possibility is that they simply requested more information. However, learners in the cooperative condition did not ask for significantly more hints than those in the competitive condition (cooperative: M = 22.81, SD = 18.43; competitive: M = 17.08, SD = 14.79; $t(50) = 1.24$, p = 0.22).

Another possibility is that informants in the deceptive condition told more lies. In this game, a lie is either an interior point that the informant labeled as exterior or vice versa. The number of lies in each game was divided by the total number of hints. As Figure 3 shows, competitive informants lied more than cooperative informants (M = 0.41, SD = 0.34 vs. M = 0.067, SD = 0.14, $t(50) = 4.70$, $p <$ 0.001). Moreover, the proportion of lies told was correlated with rectangle error, suggesting that lying to learners did affect their ability to ascertain the truth (r = 0.40, $p < 0.01$).

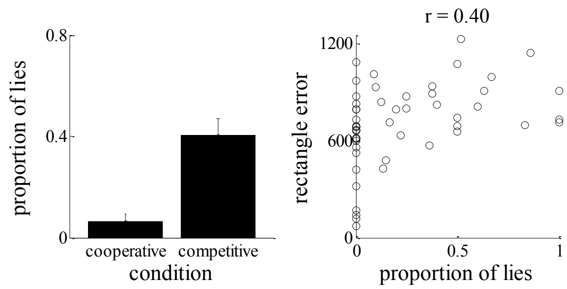

Figure 3: Left: Informants lied about 40% of the time in the competitive condition but only rarely in the cooperative condition. Error bars represent ±1 SEM. Right: Increased lying is related to higher error in guessing the rectangle.

Just as in real life, it was possible to mislead even while giving apparently truthful information. For instance, providing a negative example far away from the rectangle's edge may be technically accurate but gives a misleading sense of the rectangle's boundaries. How often did deceptive informants rely on this type of misinformation? We explored this by calculating what percent of hints were exterior points at least 35 pixels away from the rectangle's edges. As Figure 4 shows, competitive informants provided many more of these (competitive: M = 0.31, SD = 0.32; competitive: M = 0.10, SD = 0.12; $t(50) = 3.06$, $p < .01$). Figure 4 also shows the proportion of interior points that were at least 35 pixels inside. Interestingly, cooperative informants provided these clues much more often (cooperative: M = 0.36 SD = 0.30; competitive: M = 0.12, SD = 0.17; $t(50) = 3.54$, $p < .01$), which may be related to the fact that learners were allowed five hints (considerably more than the two required to mark the corners). Neither the exterior nor the interior hints were correlated with rectangle error (exterior: r = 0.23, p = 0.10; interior: r = -0.10, p = 0.48).

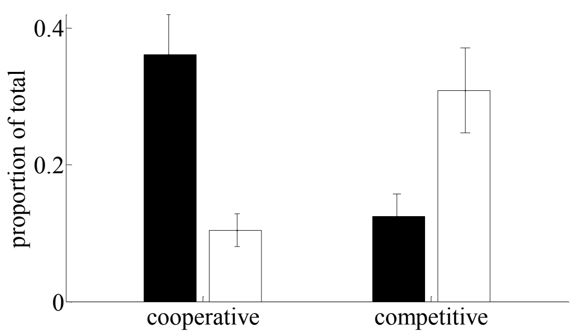

Figure 4: Cooperative informants gave many more unhelpful interior points (black), while competitive informants gave many more unhelpful exterior points (white). Error bars represent ±1 SEM.

The results suggest that correctly labeled points marking the inside corners of the rectangles (as in Figure 1a) were especially helpful. Eighty-five percent of the hints given to the five most successful learners in the cooperative condition (those more than one standard deviation below the mean error) were points of this type. The proportion of these hints is the best predictor of rectangle error (r = -0.63, p < 0.001).

Other kinds of information seem to have been of relatively little use. In contrast to the steady stream of corner hints given to the most successful learners, the four worst learners in the competitive condition (those more than one standard deviation above the mean error) received a mixture of different hints. Although deceptively labeled exterior points made up 52% of the hints, correctly labeled interior points marking corners (11%), correctly labeled exterior points marking edges (9%), and correctly labeled interior points close to the center of the rectangle (11%) were also represented.

How does the information provided by cooperative and competitive informants differ? Higher error in the competitive condition was associated with fewer corner hints and more deceptive hints. Evasive hints were also more common. This suggests that competitive informants successfully misled their learners by providing unhelpful information.

How often did learners verify information?

Our second question was whether learners would verify the information given to them, and whether the tendency to verify would be affected by whether the informant was helpful or deceptive. Because they had the option of searching for evidence independently, learners could request a hint and then "check" the hint by digging in the same place. This checking strategy could be used to verify truthful hints and catch lies. One might expect that learners in the competitive condition would check their hints more often, but this was not the case. The mean proportion of hints checked was the same for both conditions and was not correlated with rectangle error (cooperative: M = 0.28, SD = 0.31; competitive: M = 0.29, SD = 0.32; r = -0.10, p = 0.46). This result seems to reflect a reluctance among learners in the competitive condition to check their hints, possibly because they had only five opportunities to gather information about a given field. Checking information would have meant one less opportunity to gather additional information. However, considering that games consisted of twenty fields with five chances to gather information in each, learners actually had ample opportunity to establish whether their informants were cooperative or competitive, suggesting that learners in the competitive condition did not see a need to check their informants' hints.

Interestingly, 19 of the 52 learners – more than a third – did not check a single hint. Did checking help the remaining learners to guess the rectangles' dimensions more accurately? As Figure 5 shows, we found that checking hints was weakly associated with higher accuracy in the cooperative but not in the competitive condition (ANOVA: 2 (condition: cooperative vs. competitive) x 2 (checking: checkers vs. non-checkers); F(1,48) = 3.27, p = 0.077). Still, the mean proportion of checks was significantly correlated with rectangle error in the cooperative condition (r = -0.51, p < 0.01).

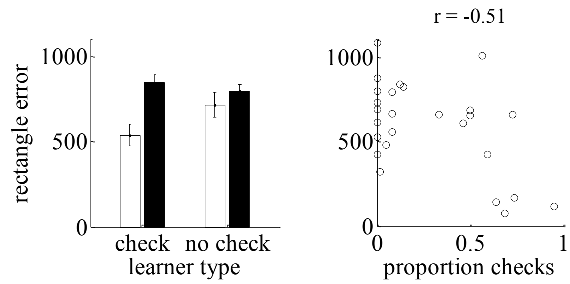

Figure 5: Left: A marginal interaction between condition (white = cooperative, black = competitive) and learner type suggests that checking was only helpful for learners in the cooperative condition. Error bars represent ±1 SEM. Right: increased checking is related to lower error in the cooperative condition.

How often do learners verify the information provided, and does this differ based on whether the informant is deceptive or not? The data suggest that learners checked about 30% of their hints in both conditions on average, although individual behaviors varied widely. Not all learners attempted to verify information from a potentially deceptive informant.

How well do learners recognize deception?

Our third question was whether people were capable of recognizing deception even in the absence of superficial cues like affect or tone. As Figure 6 demonstrates, cooperative informants received significantly higher ratings than competitive informants (cooperative: M = 16.64, SD = 3.82; competitive: M = 9.48, SD = 5.43, t(49) = 5.43, p < .001).

How were people able to identify the deceptive informants? One possibility is that they made use of the verifiable information to check the hints offered by the informant. However, Figure 6 also shows that checking did not affect ratings in either condition (ANOVA: 2 (condition: cooperative vs. competitive) x 2 (checking: checkers vs. non-checkers); F(1,47) = 25.59, p < 0.001).

The data do suggest that different types of checked hints may communicate different information. Intuitively, a verified truthful hint should boost learners' confidence in their informants, while an exposed deceptive hint (caught lie) should lower it. We define verified truthful hints and caught lies as proportions of the number of requests. Both types of hints were correlated with informant ratings in the expected directions (verified truthful: r = 0.56, p < 0.001; caught lies: r = -0.76, p < 0.001). Figure 6 shows the relationship between the proportion of verified truthful hints minus the proportion of caught lies (check difference) and informant rating (r = 0.79, p < 0.001). This suggests that learners may be willing to give inconsistent informants the benefit of the doubt as long as the majority of the information provided is truthful.

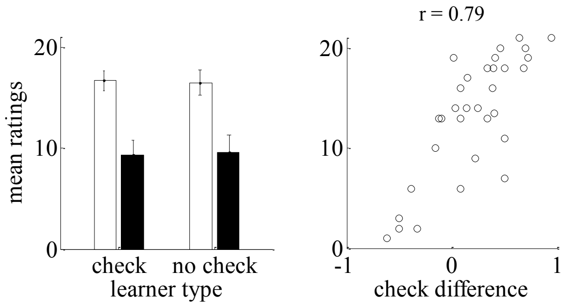

Figure 6: Left: Cooperative learners (white) rated their informants as more helpful whether they checked hints or not. Right: Ratings were strongly associated with the proportion of verified truthful hints minus the proportion of caught lies. This suggests that the ability to identify deception is more dependent on the results of checking rather than the frequency.

Another clue to the informants' intentions may have been the information itself. The proportion of corner hints was correlated with informant rating, as was the proportion of unhelpful exterior hints (corners: $r = 0.50$, $p < 0.001$; unhelpful: $r = -0.28$, $p < 0.05$). These correlations were strengthened by including other types of hints. Cooperative learners received mostly corner hints and unhelpful interior hints ($M = 0.69$, $SD = 0.27$). Taken together, the proportion of truthful interior hints was correlated with informant rating ($r = 0.76$, $p < 0.001$). In contrast, competitive learners received mostly distant exterior points ($M = 0.61$, $SD = 0.28$). Fifty-one percent of these hints were accurately labeled but evasive. The rest were lies. Taken together, the proportion of distant exterior points was negatively correlated with rectangle error ($r = -0.77$, $p < 0.001$).

These results suggest that learners made inferences about the deceptiveness or helpfulness of an informant in order to guide their overall strategy. Although the mean number of requests did not vary by condition, it was correlated with informant rating ($r = 0.42$, $p < 0.01$). This suggests that learners who believed that their informants were helpful asked for more hints. Furthermore, these learners tended to interpret positive examples as corner hints, as predicted by Shafto and Goodman (2008). This result also agrees with the findings of Bonawitz et al. (2009) regarding exploration. Learners who believed that their hints were purposefully sampled rarely searched beyond the boundaries suggested by positive examples.

How effectively do people recognize when they are being deceived? Learners had little difficulty distinguishing cooperative and competitive informants, even when they did not verify their information. This suggests that learners may have based their inferences on the type and frequency of hints they received.

General Discussion

In this experiment we addressed three fundamental questions about how people behave in situations involving deception. We found that deceptive informants give a different pattern of data than helpful informants and that the most effective liars combine outright lying, misleading truths, and helpful hints. Surprisingly, we also found that learners did not verify more often when faced with deceptive agents and that increased verification was only associated with improved performance in the cooperative condition. However, learners were very good at recognizing deceptive agents, even when they did not verify their information.

One intriguing implication of our results is that long-term deception may have as much to do with evasion as actual lies. Although competitive informants lied more often than their cooperative counterparts, they still provided more accurately labeled hints than lies (probably because outright lies would be easier to spot). However, unlike the corner hints characteristic of cooperative informants, accurately labeled hints from competitive informants tended to mark distant exterior points. This suggests that competitive informants understood what kinds of hints would be most helpful and deliberately avoided providing them.

These findings suggest a possible strategy for identifying deceptive sources of information. Consider a police officer interviewing a suspected murderer. The officer need not structure the interrogation around catching the suspect in a lie. A cautious suspect may never tell one. An alternative strategy would be to compare the suspect's testimony to the "ideal" testimony: information that the officer would find most helpful in terms of cracking the case. Discrepancies would suggest an uncooperative suspect, assuming that the suspect has some knowledge of the crime.

Our results also illustrate how people's expectations guide their inferences. Earlier research has shown that learners interpret purposefully sampled examples more narrowly than randomly sampled information (Bonawitz et al., 2009; Xu & Tenenbaum, 2007). In other words, learners have expectations about information from helpful sources. These expectations may help explain why informant ratings were related to the proportion of verified truthful hints and caught lies. Informants whose hints matched the learners' preconceptions of what helpful hints would be were given relatively high ratings, while informants whose hints did not meet the standard were assumed to be unhelpful.

These findings suggest how the hypothetical murder suspect may successfully deceive his interviewing officer. The officer expects testimony that is both truthful and helpful. The murder suspect, assuming that he is guilty and wants to get away with his crime, should provide a mixture of information: enough helpful information to maintain the officer's trust but enough deceptive and evasive information to confound the investigation. The suspect could build his credibility further by providing truthful, helpful information that the officer can verify. Our research suggests that these verifications may even offset the damage caused by the occasional lie.

Because the design of the game used in this study allows us to clearly define abstract concepts such as evasion and trust, it can be used to explore other aspects of deception. Future possibilities include variations in which the informant decides whether or not to provide information. The findings would be most relevant to commercial transactions. For example, a person trying to sell a used car chooses what aspects of the car's history to share with potential buyers, who then decide whether or not to make the purchase. Ideally, these results would used to construct computational models of human deception that would generalize to more naturalistic settings.

Acknowledgements

Daniel J. Navarro was supported by ARC grant DP0773794.

References

Bonawitz, E., Shafto, P., Gweon, H., Chang, I., Katz, S., & Schulz, L. (2009). The double-edged sword of pedagogy: Modeling the effect of pedagogical contexts on preschoolers' exploratory play. *Proceedings of the 31st annual conference of the Cognitive Science Society* (pp. 1575-1581). Austin, TX: Cognitive Science Society.

Bond, C., & DePaulo, B. (2006). Accuracy of deception judgments. *Personality & Social Psychology Review, 10,* 214-234.

Brandt, D., Miller, G., & Hocking, J. (1980). The truth-deception attribution: effects of familiarity on the ability of observers to detect deception. *Human Communication Research, 6,* 99-110.

Buller, D., Strzyzewski, K., & Hunsaker, F. (1991). Interpersonal deception: II. The inferiority of conversational participants as deception detectors. *Communication Monographs, 58,* 25-40.

Burgoon, J., Buller, D., & Floyd, K. (2001). Does participation affect deception success? A test of the interactivity principle. *Human Communication Research, 27,* 503-534.

DePaulo, B., Stone, J., & Lassiter, G. (1985). Telling ingratiating lies: effects of target sex and target attractiveness on verbal and nonverbal deceptive success. *Journal of Personality and Psychology, 48,* 1191-1203.

DeTruck, M., & Miller, G. (1985). Deception and arousal: isolating the behavioral correlates of deception. *Human Communications Research, 12,* 181-201.

Eckman, P., and O'Sullivan, M. (1991). Who can catch a liar? *American Psychologist, 46,* 913-920.

Littlepage, G., & Pineault, M. (1985). Detection of deception of planned and spontaneous communications. *Journal of Social Psychology, 125,* 195-201.

Mascaro, O., & Sperber, D. (2009). The moral, epistemic, and mindreading components of children's vigilance towards deception. *Cognition, 112,* 367-380.

Shafto, P., & Goodman, N. (2008). Teaching games: Statistical sampling assumptions for pedagogical situations. *Proceedings of the 30th annual conference of the Cognitive Science Society* (pp. 1632-1637). Austin, TX: Cognitive Science Society.

Xu, F., & Tenenbaum, J. (2007). Sensitivity to sampling in Bayesian word learning. *Developmental Science, 10,* 288-297.