Infants ask for help when they know they don’t know

Louise Goupil\textsuperscript{a,1,2}, Margaux Romand-Monnier\textsuperscript{a,1}, and Sid Kouider\textsuperscript{a,2}
were not given this opportunity and could only choose a location by themselves (control group; \( n = 40 \)). This manipulation enabled us to test whether infants can monitor and communicate their own uncertainty. Indeed, if infants can monitor their own knowledge state, they should use the AFH option (i.e., opt-out) when they have forgotten the toy location, thereby avoiding mistakes and improving their performance \((22, 23)\). Furthermore, if infants can monitor the strength of their memory trace, they should use the AFH option more often at higher levels of uncertainty (i.e., for longer delays and impossible trials).

We first examined the overall performance by computing mean accuracy for the pointing task (Fig. 1B, Left). Infants pointed more often to the correct location \([\text{mean accuracy } 61\%; \; t(77) = 4.91; \; P < .001; \; 2 \text{ infants asked for help on every trial and did not provide any pointing response; consequently, they were excluded from all further analysis.}]\). This is the case for both the experimental group \([\text{mean accuracy } 66\%; \; t(37) = 4.8 \; ; \; P < .001]\) and the control group \([\text{mean accuracy } 56\%; \; t(39) = 2.2 \; ; \; P < .05]\). Crucially, consistent with our hypothesis, the experimental group performed better than the control group \([\text{Fig. 1B; } t(76) = 2.21; \; P = .03; \; \text{see also Fig. S1 for the distribution of this effect}].\)

These results suggest that infants used the AFH option strategically to improve their performance.

However, it remains possible that infants in the experimental group performed better because of a general increase in motivation. In particular, the procedure may have been more stimulating for infants in the experimental group, as they could interact with their parent. Notably, if the effect was due to a general increase in motivation, one should observe a higher rate of correct responses in the experimental group compared with the control group. By contrast, if infants genuinely monitor their own uncertainty, they should specifically ask for help to avoid making mistakes. In this case, one should observe a lower rate of incorrect responses and a similar rate of correct responses in the experimental group compared with the control group. To disentangle these two hypotheses, one computed separately the proportion of correct responses over the total number of trials \((\text{i.e., [correct trials/(correct trials + incorrect trials + no response trials + AFH trials in the experimental group)]})\) versus \([\text{incorrect trials/(correct trials + incorrect trials + no response trials + AFH trials in the experimental group)]}\). \* \(P < 0.05\); \* \(P < 0.01\); \* \(P < 0.001\). All error bars indicate SEMs.

Fig. 1. (A) Experimental procedure. Infants watched as a toy was conspicuously hidden under one of two opaque boxes in full view (possible trials) or behind a curtain (impossible trials). For possible trials, the two boxes were then occluded behind the curtain for a variable delay \((3, 6, 9, \text{ or } 12 \text{ s})\). Then, infants were presented with the two boxes again and taught to indicate where they remembered the toy to be by pointing toward its location. The chosen box was then pushed forward for the infant to recover the toy in the case of a correct response, or discover that there was no toy in the case of an incorrect response. Crucially, in a training phase, infants in the experimental group were familiarized with the option of asking their caregiver for help \((\text{Materials and Methods})\). By contrast, infants in the control group were not taught the AFH option. Thus, during the rest of the experiment, infants in the experimental group had the opportunity to decide whether they should respond by themselves (i.e., point toward one of the boxes) or acknowledge uncertainty (i.e., ask their caregiver to provide them with the forgotten information), whereas infants in the control group had no other option but to answer by themselves. (B, Left) Mean accuracy of the pointing responses \((\text{i.e., correct responses/(correct + incorrect responses)})\) for each group (control group in blue and experimental group in green). The red dotted line illustrates chance level. (B, Right) The proportion of correct and incorrect responses was computed for each participant by dividing the number of correct/incorrect pointing responses by the total number of trials \((\text{i.e., [correct trials/(correct trials + incorrect trials + no response trials + AFH trials in the experimental group)]})\) versus \([\text{incorrect trials/(correct trials + incorrect trials + no response trials + AFH trials in the experimental group)]}\). \* \(P < 0.05\); \* \(P < 0.01\); \* \(P < 0.001\). All error bars indicate SEMs.
revealed that the performance improvement in the experimental group was primarily due to infants producing a lower rate of incorrect responses compared with infants in the control group \( t(76) = 3.4; P < .01 \). Here, the proportion of correct responses remained equivalent across the two groups \( t(76) = .7; P > .9 \). This interaction between group and response accuracy \( F(1,76) = 4.6; P < .02 \) shows that infants in the experimental group selectively asked for help to avoid making incorrect responses.

The analysis above compared infants familiarized with the AFH option with infants who were not given this opportunity. However, a closer inspection of the individual data in the experimental group revealed important interindividual differences in the use of the AFH option. Indeed, a total of 14 infants out of 42 never asked for help. Importantly, these infants performed at an accuracy rate (56%) that was similar to the control group (56%; \( t(52) = 1; P > .02 \) and worse than infants who asked for help in the experimental group (72%; \( t(36) = 2.33; P < .01 \) (Fig. S2). Like these, infants who belonged to the experimental group but never asked for help displayed the same rate of correct and incorrect responses as the control group (all \( t < 1 \); Fig. S2). This observation confirms that infants who asked for help in the experimental group used this option to avoid making mistakes.

We then tested whether task difficulty had an impact on the probability of asking for help. Indeed, if infants were monitoring their uncertainty about the toy location, they should have asked for help more often as the memorization delay increased. This analysis restricted to the participants in the experimental group, who asked for help in at least one trial per condition \( n = 21 \). An ANOVA revealed that the probability of asking for help as high or impossible than for possible trials \( F(2,4; F(1,22) = 24.22; P < .01 \). Furthermore, if infants were able to produce an AFH response increased with increasing delays \( F(2,2) = 4.62; P < .05 \). Thus, infants tended to ask for help when they felt uncertain about the toy location. However, we next considered the possibility that infants simply learned during the training phase to avoid impossible trials by asking for help (Materials and Methods). If this was the case, the group differences observed should be restricted to impossible trials, and both groups should perform similarly on possible trials. By contrast, if infants genuinely monitor their uncertainty, they should be able to generalize the AFH strategy to possible trials and increase their performance accordingly. To test this, we computed mean accuracy for possible trials in isolation. This analysis revealed that even when restricting our analysis to possible trials, performance as higher in the experimental group compared with the control group (69% versus 57%; \( t(76) = 2.43; P < .01 \)). This indicates that infants did not simply avoid impossible trials but rather generalized the use of the AFH option to possible trials to improve their performance.

Finally, we examined the proportion of correct and incorrect responses over the total number of trials, computed separately for the possible and impossible conditions (Fig. 2C). We performed a mixed linear regression on the proportion of responses, using group, accuracy, and task difficulty (possible vs. impossible) as predictors and subject as a random variable. Critically, we observed a three-way interaction (likelihood ratio tests for model comparison: \( N_{\text{Subjects}} = 77, N_{\text{Observations}} = 156, \chi^2 = 8.94, P < .01 \)) but not for impossible trials \( P > .4 \). In the impossible condition, only a main effect of group as observed \( N_{\text{Objects}} = 69, N_{\text{Observations}} = 138, \chi^2 = 5.8, P < .01 \). This pattern as due to the fact that infants in the experimental group avoided impossible trials regardless of accuracy. By contrast, the pattern in the possible condition reflected the fact that the experimental group produced fewer errors than the control group \( t(76) = 3.34; P < .01 \). Here, the proportion of correct responses did not vary across the two groups \( t(76) = 1.4; P > .3 \). These results confirm that infants used the AFH option strategically to avoid making errors even in possible trials.

Discussion

When given the opportunity to decide whether they should respond by themselves or avoid responding by asking for help, 2- to 3-year-old children are able to strategically adapt their behavior. That is, they selectively seek help to avoid making errors and to avoid difficult choices. In the comparative literature, these adaptive “opt-out” behaviors have been taken as evidence for metacognitive uncertainty monitoring in several species (22, 23, 27). However, some authors have argued that such behavioral patterns could also be explained by associative or reinforcement learning mechanisms (29, 33). For instance, they suggest that difficult trials are simply avoided because individuals learn that the probability of obtaining a reward is lower for those trials (29, 33). Whether or not this associative interpretation can be ruled out in comparative research, in which animals are extensively trained, remains a controversial issue (23, 31). However, in the present study, an associative account seems unlikely because infants only received a few trials (i.e., a maximum of 10 trials for each level of task difficulty), leaving little room for associative learning. Moreover, the proportion of AFH responses did not increase across time (effect of trial rank on the proportion of AFH responses: \( F(1,2) = .22; P > .6 \), ruling out an associative interpretation in terms of reinforcement learning.

Another issue raised in the comparative literature concerns the fact that when the opt-out alternative is available simultaneously, infants might take place...
The experiment started with a warm-up phase during which the infants in the control group were not taught that they could ask for help, and even though their caregiver remained unresponsive, they did observe a few spontaneous “AFH-like” responses in this group [mean number of AFH responses in the control group: 6.6; in the experimental group: 1.42; t(39) = 3; P < .05; Fig. S3]. Ho ever, when the frequency at which infants looked to and addressed their parent in the control group, they found absolutely no increase in task difficulty (Fig. S3A), and excluding those trials did not impact performance (Fig. S3B). Thus, infants in the control group did not orient selectively toward their parents when they were more likely to have forgotten the toy location. In turn, this finding confirms that infants in the experimental group did not automatically turn to their parents when the infant’s uncertainty was not expressed. Rather, our results are consistent with the idea that infants in the experimental group learned that they could communicate with their caregiver to obtain some help, when they felt that they were likely to make an error.

The fact that the infants in the control group did not spontaneously ask for help here when they were uncertain indicates that they needed to be instructed that the AFH option was available in order for them to use it in a strategic manner. Still, 35% of the infants in the experimental group did not take advantage of the AFH option. This raises the question of how some infants ask for help here. We noted that there might be impossibility that this difference in behavior reflects differences in metacognitive ability. Notably, children have often been found to overestimate their own performances (1, 12, 13). Thus, one tempting interpretation is that some infants never asked for help because they always felt confident that they could respond correctly on their own. However, several alternative interpretations remain. In particular, we noticed that the infants did not ask for help in the experimental group tended to be less language rich (e.g., using smaller vocabulary size compared to infants in the control group) and did not ask for help in not without the presence of their caregivers. This indicates that the difference in behavior is not due to differences in language development but rather to differences in the way infants use the AFH option. Thus, it might also be that other factors, such as executive functions and attachment, determined whether or not infants asked for help in this experiment. This is important in future research, since one cannot rule out the possibility that the infants in the control group might have actually asked for help when no response came to their mind (e.g., to seek comfort), whereas their caregivers might have observed a similar tendency in the control group. In fact, although infants in the control group did not ask for help, they could ask for help, even though their caregivers remained unresponsive, and it is possible that this difference in behavior reflects differences in metacognitive ability.

Materials and Methods

Participants. Eighty healthy full-term infants were included in the final analysis (mean age, 20.17 mo; age range, 19–21.06 mo). Half of them participated in the study as the control group (n = 40; mean = 20.08 mo; SEM = 0.09; range, 19–20.97 mo; 19 females), and the other half as the experimental group (n = 40; mean = 20.26 mo; SEM = 0.09; range, 19.17–21.06 mo; 19 females). An additional 51 infants (N_{experimental} = 22; N_{control} = 29) were tested but not included in the sample because of fussiness (8), procedure error (5), failure to point to the boxes to indicate a choice in the training phase (21), participation in less than two test trials (5), refusal to take part in the experiment (9), or caregiver interference (3). The study was approved by the regional ethical committee for biomedical research (CERES; Consell General d’Investigacions Germans en Salut) and informed consent was obtained from the parents before the experiment. All infants were given a diploma for taking part in the study. Infants’ vocabulary was evaluated with a French adaptation of the MacArthur–Bates Communicative Development Inventory (38), which allowed us to verify that there were no differences in vocabulary size between the two groups (t(69) = 0.2; P > 0.8; nine questionnaires were not returned).

Materials and Apparatus. The apparatus consisted of two identical boxes (12 × 12 × 13 cm), each placed on a piece of black cardboard (32 × 31.5 cm). Two wooden toys and two cups were dedicated to the warm-up phase. Ten unique plastic characters were dedicated to the experiment. They were stored on a table out of the infants’ view and randomly sampled to be presented individually over the course of 4 training trials and 10 experimental trials. In both groups, the infant was seated in a high chair facing the testing table. The experimenter and the parent sat on the other side of the table, opposite the child (Fig. 1A). An opaque black curtain (20 × 60 cm) split the table (70 × 60 × 73 cm) in two. Preceding the session, the parent was instructed to keep his or her gazed on the infant and not to interfere with the infant in any way, and to refrain from moving his or her own head and body from talking during the trials, except when the task required them to do so. The entire scene was recorded from two perspectives, behind the experimenter and behind the infant, to ensure the neutrality of the parent and experimenter.

Procedure. The experiment started with a warm-up phase during which the infant and their caregiver played with the experimenter. As soon as the infant started to feel comfortable, a training phase began. It consisted of four trials, for which the location of the toy was pseudorandomized. In the first two trials, similar in both the experimental and control group, infants saw the experimenter hide a toy under one of two opaque boxes. After a delay during which the boxes were hidden behind a curtain, the experimenter asked them to point to indicate where they remembered the toy to be. As soon as the infant produced a clear response, the selected box was pushed forward to allow him or her to recover the toy. This was followed by two impossible trials in which the toy was hidden beneath one of two opaque boxes out of the infant’s view (e.g., behind the curtain). Infants from the experimental group were taught to ask for help when they did not know the location of the toy. To do so, infants’ pointing responses in these trials were ignored, and the experimenter turned to the caregivers and asked them if they knew where the toy was. Caregivers were instructed to wait for their child to look at them in the eyes before helping them by pushing the correct box forward and saying “Here it is, look.” Importantly, infants from the control group were not taught this option. To match the two groups, their pointing responses were also systematically ignored in these trials. After asking the infant a second time about the location of the toy, the experimenter simply pushed the correct box forward. The testing phase (10 trials) was identical across the two groups and similar to the training phase, except that there were now five levels of difficulty: possible trials with 3, 6, 9, or 12 s of memorization delay, and impossible trials. The order of presentation was pseudorandomized using a Latin square across the 10 conditions (two sides and five levels of difficulty). The same randomization was used in both groups: Infants in the experimental and control groups were thus matched.
for order of presentation in both phases of the experiment. The side on which the parent sat was also randomized. The experiment stopped after the infant had completed 10 experimental trials (corresponding to the 10 experimental conditions described above) or became too fussy to continue (in which case they did not complete every experimental condition).

**Data Collection and Analysis.** Responses were coded from video recordings by two independent observers (M.R.M. and a naive coder), who were blind to the conditions (location of the toy and delay). Four different types of responses were identified: pointing to the left, pointing to the right, asking for help, or no response (i.e., trials for which the infant did not produce a pointing or AFH response). To compute performance, only pointing responses were used: If the infant pointed toward the box under which the toy was hidden, the response was considered correct; if the infant pointed toward the opposite box, the response was considered incorrect. Therefore, AFH responses did not count as a correct or incorrect response and, just like “no response” trials, were not included in the computation of mean accuracy. Notably, the proportion of “no response” trials was not significantly different between the control and experimental groups [t(76) = 0.5; P > 0.60]. Coders agreed on 570 of the 641 responses collected (88.92%). Trials with discrepancies between the two codings (n = 71) were recorded by a third coder (L.G.) blind to the experimental conditions. The naive coder’s data were used for all of the analyses, except for trials with a disagreement between the two main coders, in which data from the third coder were used. The naive coder also blindly coded parents’ and experimenter’s behavior, to ensure their neutrality and that no external information was available to influence infants’ choices. Trials with experimental errors (n = 32) or parental interferences (n = 4) were discarded.

**ACKNOWLEDGMENTS.** We thank A.-C. Fievet and C. Girard for support with data collection, and F. Goffinet (Hôpital Cochin) for his support of our research. We thank A. Vlassova, J. Sackur, L. Barbosa, and T. Andrillon for discussions on the manuscript. This research was supported by funding from the European Research Council (ERC-DynaMind and ERC-Metaware) and the Agence National de la Recherche (ANR-10-LABX-0087 and ANR-10-IDEX-0001-02).

10. Begus K, Gilga T, Southgate V (2014) Infants learn what they want to learn: Responding to discrepancies between the two codings (n = 71) were recorded by a third coder (L.G.) blind to the experimental conditions. The naive coder’s data were used for all of the analyses, except for trials with a disagreement between the two main coders, in which data from the third coder were used. The naive coder also blindly coded parents’ and experimenter’s behavior, to ensure their neutrality and that no external information was available to influence infants’ choices. Trials with experimental errors (n = 32) or parental interferences (n = 4) were discarded.

**ACKNOWLEDGMENTS.** We thank A.-C. Fievet and C. Girard for support with data collection, and F. Goffinet (Hôpital Cochin) for his support of our research. We thank A. Vlassova, J. Sackur, L. Barbosa, and T. Andrillon for discussions on the manuscript. This research was supported by funding from the European Research Council (ERC-DynaMind and ERC-Metaware) and the Agence National de la Recherche (ANR-10-LABX-0087 and ANR-10-IDEX-0001-02).

10. Begus K, Gilga T, Southgate V (2014) Infants learn what they want to learn: Responding to discrepancies between the two codings (n = 71) were recorded by a third coder (L.G.) blind to the experimental conditions. The naive coder’s data were used for all of the analyses, except for trials with a disagreement between the two main coders, in which data from the third coder were used. The naive coder also blindly coded parents’ and experimenter’s behavior, to ensure their neutrality and that no external information was available to influence infants’ choices. Trials with experimental errors (n = 32) or parental interferences (n = 4) were discarded.

**ACKNOWLEDGMENTS.** We thank A.-C. Fievet and C. Girard for support with data collection, and F. Goffinet (Hôpital Cochin) for his support of our research. We thank A. Vlassova, J. Sackur, L. Barbosa, and T. Andrillon for discussions on the manuscript. This research was supported by funding from the European Research Council (ERC-DynaMind and ERC-Metaware) and the Agence National de la Recherche (ANR-10-LABX-0087 and ANR-10-IDEX-0001-02).

10. Begus K, Gilga T, Southgate V (2014) Infants learn what they want to learn: Responding to discrepancies between the two codings (n = 71) were recorded by a third coder (L.G.) blind to the experimental conditions. The naive coder’s data were used for all of the analyses, except for trials with a disagreement between the two main coders, in which data from the third coder were used. The naive coder also blindly coded parents’ and experimenter’s behavior, to ensure their neutrality and that no external information was available to influence infants’ choices. Trials with experimental errors (n = 32) or parental interferences (n = 4) were discarded.

**ACKNOWLEDGMENTS.** We thank A.-C. Fievet and C. Girard for support with data collection, and F. Goffinet (Hôpital Cochin) for his support of our research. We thank A. Vlassova, J. Sackur, L. Barbosa, and T. Andrillon for discussions on the manuscript. This research was supported by funding from the European Research Council (ERC-DynaMind and ERC-Metaware) and the Agence National de la Recherche (ANR-10-LABX-0087 and ANR-10-IDEX-0001-02).