Developmental Science

Developmental Science (2014), pp 1–11 DOI: 10.1111/desc.12196

PAPER

Numerical representations and intuitions of probabilities at 12 months

Ernő Téglás,[1] Alexandra Ibanez-Lillo,[3] Albert Costa[1,2] and Luca L. Bonatti[1,2]

1. Cognitive Development Center, Central European University, Budapest, Hungary
2. ICREA, Barcelona, Spain
3. Department of Information and Communication Technologies, Center for Brain and Cognition, Universitat Pompeu Fabra, Barcelona, Spain

Abstract

Recent research shows that preverbal infants can reason about single-case probabilities without relying on observed frequencies, adapting their predictions to relevant dynamic parameters of the situation (Téglás, Vul, Girotto, Gonzalez, Tenenbaum & Bonatti, 2011; Téglás, Girotto, Gonzalez & Bonatti, 2007). Here we show that intuitions of probabilities may derive from the ability to represent a limited number of possibilities. After watching a scene containing moving objects of two ensembles, 12-month-olds looked longer at an unlikely than at a likely single-case outcome when the objects were within the parallel individuation range. However, they did not do so when the scene contained the same ratio between ensembles but a larger number of objects. At the same time, they could form rational expectations about single-case outcomes in scenes containing the same large number of objects when they could exploit subtle physical parameters induced by the objects' movements and their spatial configuration. Our findings demonstrate that at early stages of development the mental representations involved in probability estimations of future individual situations are powerful and sophisticated, but at the same time they depend on infants' overall cognitive architecture, being constrained by the numerical representations spontaneously induced by the situations.

Introduction

Humans often need to think about what happens next. As Hume famously argued, short of logical consequences, all knowledge about future events, including that arising from the strongest causal relations, is probabilistic. Yet classic research on adult reasoning has raised serious questions about whether and how humans can estimate probabilities (e.g. Tversky & Kahneman, 1974, 1981; Gigerenzer & Hoffrage, 1995; Cosmides & Tooby, 1996). Crucial to our research, despite sharp differences among theories of probabilistic reasoning, the consensus supports the view that humans are good at tracking the frequencies of past events, but are poor at exploiting this ability in order to judge the probability of single, unexperienced events. For example, while adults give reasonable estimates of the distribution of students in different fields of study, they fail to estimate accurately the probability that a single student with certain character traits will major in a particular field (Kahneman & Tversky, 1973). Likewise, when pre-adolescents are told that in a village with a known number of liars the noses of some known proportion of these liars will turn red when they lie, they can estimate the frequency of liars in a random group of people with red noses, but they are unable to estimate the probability that a given person with a red nose is a liar (Zhu & Gigerenzer, 2006). This evidence indicates that people have difficulty reasoning probabilistically about single events.

Still, recent research suggests that in several conditions adults (e.g. Griffiths & Tenenbaum, 2006) reason rationally about single cases, in a way consistent with Bayesian theories. In simple situations, even infants seem to form expectations about single outcomes (e.g. Téglás *et al*., 2011). Importantly, they can form these expectations in novel situations with which they have had no previous experience. This has

Address for correspondence: Luca L. Bonatti, ICREA and Universitat Pompeu Fabra, C. Roc Boronat, 138, Edifici Tanger, 55.110, 08018 Barcelona, Spain; e-mail: lucabonatti@mac.com

been interpreted as revealing that the ability to intuit the probability of single events is not accomplished through a simple extrapolation from frequency, or past experience.[1]

In the first demonstration of this type of proto-probabilistic reasoning (Téglás *et al.*, 2007), infants saw a lottery-like container with one exit point. Three identical objects and a fourth object, different in color and shape, bounced inside the container. Then, while an occluder masked the container, infants saw one single object exiting. In one condition, the exited object belonged to the most numerous class, whereas in another condition infants saw the unique object exit. After the occlusion period, infants looked longer when the single object exited the container than when one of the three identical objects exited. This behavior suggests that the infants formed a probabilistically correct expectation even though they had no previous experience with the single outcome. Infants can also flexibly adapt their predictions about such an outcome according to the relevance of the available information (Téglás *et al.*, 2011). For example, if in the situation described above, an object is close to the exit and a brief occlusion hides it, infants expect this object to exit the container first, regardless of whether it came from the more or less numerous class. However, if the occlusion period is long, infants disregard the object's distance from the exit, which is no longer relevant because the objects continue to move during the occlusion. Instead, they again use the size of the class to form expectations about the single outcome (see also Téglás *et al.*, 2007, Experiment 2).

Taken together, these results suggest that 12-month-old infants have an intuitive sense of probability that (a) does not necessarily depend on previous experiences with the frequencies of the outcomes and (b) can exploit several sources of information to optimize expectations (class size or physical distance). While infants' intuitions of probabilities may be independent of the experience of frequencies, however, they certainly may interact with other domains of cognition. The current research explores how infants' ability to form expectations of probabilities is influenced by the physical and numerical representations of the visual scene. The ability to represent quantities is crucial because, in order to estimate the probability of an outcome, infants have to represent the class distribution of the objects present in the scene. The ability to reason about physical aspects of the scene is crucial too, because changes in the physical configuration of a situation may render some potential inference about class numerosity irrelevant.

Here, we focus on how the representation of quantity may affect infants' intuitions of probabilities about single cases. Ample evidence exists that quantity estimates do not stem from a single uniform system. A well-studied system in adults, often described as the 'subitizing', or the 'object tracking' system, represents and tracks small quantities of objects in parallel (Pylyshyn, 2001, 2007; Scholl & Pylyshyn, 1999). It is precise and fast, but unable to accommodate quantities beyond the limits of the indexes available to track objects individually (about four). A second system can represent large quantities, but only in an imprecise and approximate way. Many studies have described some of its properties in adults and nonhuman animals (for a review, see Feigenson, Dehaene & Spelke, 2004). Several lines of research suggest that the same systems of representation also exist in infancy and are characterized by the same constraints (Feigenson, 2005; Feigenson & Carey, 2005; Xu & Spelke, 2000; Carey, 2009). It is important to our research that infants can also represent ensembles of objects grouped according to some distinguishing feature, such as color or shape (Zosh, Halberda & Feigenson, 2011). Furthermore, when the number of objects within each ensemble remains within the limits of parallel individuation, they can also represent the individual objects within the sets. However, when such limits are exceeded, ensemble representations only contain a global, approximate representation of quantities within each ensemble, perhaps as a continuous variable losing track of the individual objects within the ensembles (Zosh *et al.*, 2011; Feigenson & Halberda, 2004, 2008).

The questions we ask here are the following: How are infants' probability intuitions about single cases affected, if they are affected at all, by these systems of representations; and how adaptive to relevant cues in the scenes are such intuitions? In order to investigate these questions, we presented infants with scenes similar to those tested by Téglás *et al.* (2007), changing both the number and size of the presented objects. All of the scenes contained objects of two categories, but while their ratio remained constant (always 1:3), their absolute number varied. In some of the experiments, the number of objects within each category remained within the parallel individuation limit, whereas in others it exceeded it. If infants can reason about the probability of single future events only when they can

[1]By construction, the 'violation of expectations' method (VOE), widely used in infant studies and used also here, can only detect surprise retroactively, after an outcome has occurred. Thus, on the basis of the data that we and other researchers present, we cannot conclude anything about proactive, anticipatory expectations that infants may form while looking at a scene. Indeed, real anticipation for actions or nontrivial outcomes are rarely found in infancy (Southgate, Johnson, Osborne & Csibra, 2009; Southgate, Johnson, El Karoui & Csibra, 2010). In this paper, we follow the standard practice of referring to VOE as indicating that infants expect an outcome, while remaining neutral about the proactive or retroactive nature of their expectations.

individually represent them, then the absolute cardinality of the ensembles, and not their ratio, will influence their expectations: infants might form expectations of probability only when the future states of affairs, determined by the number of objects currently experienced, can be tracked. If, instead, infants can reason about the probability of single cases independently of their ability to individually track possible states of the world, then ratio, and not absolute cardinalities, should be the most important determining factor. Experiments 1 and 2 test these alternative hypotheses.

Experiment 1

Method

Participants

Twenty healthy full-term 12-month-old infants were retained for analysis (12 girls, mean age 12 months, 17 days). An additional 16 infants were tested but not retained (12 became fussy, one because of the caretaker's intervention, three for material failure; see Procedure). Infants were considered to be fussy if during familiarization or test they gave signs of discomfort while listening to the stimuli, such as making more than sporadic vocal emissions or frequently turning their bodies towards their caretakers or making other clear signs of avoidance of the display or the experimental situation. We also excluded infants whose caretakers actively interfered with the experiment by talking to the infants or inciting them to look at the stimuli during the test phase.[2]

Materials

We generated movies as 3D animations with Maya 6.0 (25 fps), using Maya's simulator of physical movements, and compressed as QuickTime files with Sorenson Squeeze. The movies simulated four solid objects bouncing inside a container, resembling a lottery machine. Their movement was restricted to one frontal plane, so that no object would ever pass behind another one. This procedure ensured that infants could see the full population at any moment during the movie. The collisions between the objects respected solidity and gravity. The container frame covered a 14 × 14 cm area. Each object occupied approximately 0.5% of the area of the container. There were two different types of movies, one for the familiarization and one for the test phase. The familiarization movies presented two objects of each kind bouncing for 14.5 s. After the bouncing period, one object exited the container. After the exit, an occluder progressively covered the container and its content, and the trial ended. The length of the familiarization movies was 19.24 s.

The test movies were like the familiarization movies, but had two crucial differences. First, the frame contained three objects of one category and a single object of the other category. Second, the exit phase and the occlusion phase were presented in reverse order. Thus, at the end of the bouncing period (13 s), the occluder progressively faded over the container. Then, a full occlusion period lasting 2 s followed. Finally an object exited the container. After the exit, the occluder disappeared, revealing the container and its contents once more (Figure 1). The experiment was run with PsyScope X (http://psy.cns.sissa.it), on an Apple DualG5 computer. Movies were presented on a 17-inch screen. A camera hidden behind the screen digitally recorded infants' faces on a separate computer. Recordings were then inspected offline with the software PsyCode (http://psy.cns.sissa.it) to compute looking times.

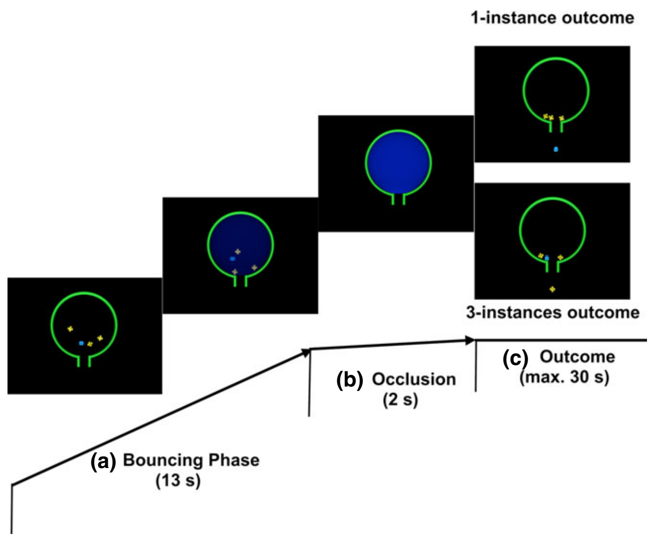

Figure 1 *Structure of the experiments, exemplified with the objects used in Experiment 1. The object size and number changed between experiments, but the structure was identical.*

[2]Although the rejection rates obtained by applying these criteria are substantial, they are not uncommon. Slaughter and Suddendorf (2007) ran a meta-analysis of infant experiments reporting rejection rates ranging from 0 to 68%. These authors show that such variation does not systematically influence experimental outcomes. Indeed, the current rejection rate is comparable to that obtained in our previous work with analog methodology (Téglás *et al.*, 2007, 2011). It is therefore likely to be due to the nature of the paradigm.

Procedure

Infants sat on their caretaker's lap, approximately 80 cm from the screen, in a darkened room. Caretakers wore black opaque glasses during the experiment. They were instructed not to interact with infants, to hold infants at their hips with both hands. They had to let infants move freely. If infants turned their entire body away from their initial posture (facing the screen), caretakers were instructed to count to 5 and then gently turn the infants back towards the center. This procedure gives sufficient time to trigger a trial timeout without modifying infants' natural reactions to the stimuli. Thus, it ensures that caretakers reorient infants towards the screen after any possible timeout, while at the same time it allows the experimenter to continue testing infants in further trials. The experimenter, who was blind to the experimental conditions, monitored infants' behavior from a separate screen. Before each movie, a visual attractor appeared to orient infants' attention towards the center. When infants looked at the center, a movie started playing. To ensure that infants saw every movie in its entirety, the presentation of the stimuli was infant-controlled. Movies were paused when infants looked away from the screen and continued playing once they looked back at it.

During familiarization, infants saw two movies that ended with an occlusion. Looking time was not measured during this phase. Immediately after familiarization, four test trials began, each featuring a different experimental movie. At the end of the occlusion phase, an object exited the container accompanied by a sound. At that point, looking time measurement began. Half of the test trials ended with a 1-instance outcome, and half with a 3-instance outcome. The object categories differed in shape (cube or star) and color (yellow or blue). The combinations of shape, color and outcome produced eight different animations. Each participant saw a different combination of shape and color of the final object in each of the four test movies, and the total combinations of shapes and colors were Latin-squared across participants. Half of the infants saw a 1-instance outcome first, and the other half a 3-instance outcome first. A trial ended when infants looked away for more than 2 consecutive s, or when they looked for more than 30 cumulative s. Looking time was coded offline for analysis. All reported data refer to the offline data coding. Infants were excluded from the analysis if they had cumulative timeouts in two or more trials, or if they became fussy. Trials were excluded from the analysis if infants turned away exactly when the object exited the container, when the experimenter erroneously triggered a timeout before a 2-s look-away period, or when looking times exceeded 2 *SD*s, computed per conditions.

Results and discussion

A repeated measure ANOVA, with Outcome Type (1-instance vs. 3-instance outcome) as the within-participant factor and participants as the random factor, and mean looking time as the dependent variable, found that infants looked longer when the single, unique object exited the container ($M_{3\text{-instances}} = 11.47$ s, $M_{1\text{-instance}} = 14.87$ s; $F(1, 19) = 5.59$, $p \leq .029$; Figure 2). Separate ANOVAs showed that neither color nor shape of the exited object contributed to this difference. This result replicates Téglás *et al.* (2007) with different stimuli, confirming that when viewing a novel scene 12-month-olds expect the most likely outcome, out of all possible ones, even without previous experience with it.

The current finding is compatible with the hypothesis that infants create representations of individual future outcomes, and form expectations about the most likely outcome by considering which cases may end with a 1-instance outcome or a 3-instance outcome. If these possibilities are indeed constructed individually, then the ability to reason about the probability of single cases may not extend to situations involving more numerous possible outcomes. In the second experiment, we investigated whether this is the case. We created movies containing a number of objects far beyond any range that could be individually represented, but with a ratio between categories identical to that of Experiment 1. Because infants at 10 months can already discriminate a 2:3 ratio (Xu & Arriaga, 2007), we assume that 12-month-old infants should have no difficulty at

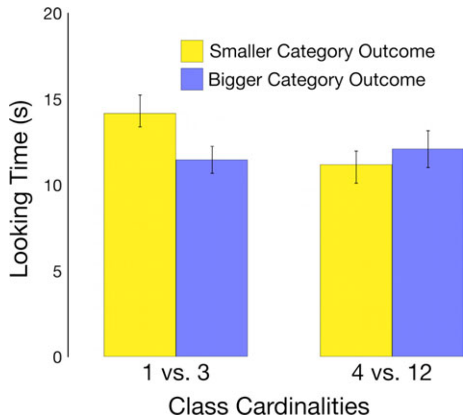

Figure 2 *Results of Experiments 1 and 2. In both experiments, infants saw one object belonging to either the more numerous or the less numerous class exiting the container. The ratio between categories was identical in both experiments, but the absolute number was different. In Experiment 1 both categories were within the subitizing range. In Experiment 2 both categories were beyond the subitizing range.*

discriminating a 1:3 ratio. When the absolute cardinality per category exceeds 3, however, objects within ensembles cannot be represented individually (Feigenson & Halberda, 2004; Feigenson *et al.*, 2004; Feigenson & Carey, 2005). To ensure that in both categories the absolute number of objects surpassed the parallel individuation range, we used, respectively, 4 and 12 objects. If infants can form expectations about the most likely outcome by recruiting any system of quantity representations, then they should be able to form these expectations even when the scene contains ensembles of these larger cardinalities. If, instead, infants can form expectations about single outcomes only when they can consider the possible states of affairs individually, then they may be unable to reason about single events when the scene is composed of ensembles with ratios identical to those in Experiment 1, but with higher cardinalities.

Experiment 2

Method

Participants

Twenty full-term 12-month-olds were retained for analysis (11 girls, mean age 12 months, 19 days). An additional 11 infants were tested but not retained (11 became fussy).

Materials and procedure

We generated movies with the same properties and constraints as in Experiment 1, but with 16 objects. The familiarization movies followed the same event sequence as in Experiment 1, but displayed 8 objects of each category. The test movies followed the same event sequence as in Experiment 1, but contained 4 objects of one category and 12 of the other category. The objects occupied approximately 9% of the container's area (Figure 3A). Otherwise, materials and procedure were identical to Experiment 1.

Results and discussion

Infants did not look differentially at the outcomes ($M_{12\text{-instances}}$ = 13s; $M_{4\text{-instances}}$ = 12.5 s; $F(1, 19) = .12$, $p > .7$; Figure 2). No other effect was significant. We explored whether this null result differed from the infants' behavior in Experiment 1 by pooling the data of the two experiments. We ran a mixed-model ANOVA with Experiment (1 vs. 2) as the between-participant factor, Outcome Type (1-instance vs. 3-instance outcome) as the within-participant factor, and participants as the random factor nested within the Experiment. The only significant effect was an interaction between Experiment and Outcome Type ($F(1, 38) = 5.2$, $p \leq .02$). Post-hoc analyses with the Scheffé method revealed that participants in Experiment 1 looked longer at the impossible outcome ($p \leq .02$), but participants in Experiment 2 did not ($p > .4$). The interaction shows that only infants in Experiment 1 were sensitive to the differences in probabilities of the single outcomes, whereas participants in Experiment 2 were indifferent to them.

It is also possible, however, that dynamic scenes containing 16 moving objects are simply too complex for infants to analyze. Perhaps infants did not react differentially in the test phase because they were overwhelmed by the sheer complexity of the situation. To test this possibility, we ran Experiment 3. We maintained the same number of objects, but introduced information that could induce 12-month-olds to reason about the scene without using quantity representations. Here, we devised a modification of our scenes to elicit a change in infants'

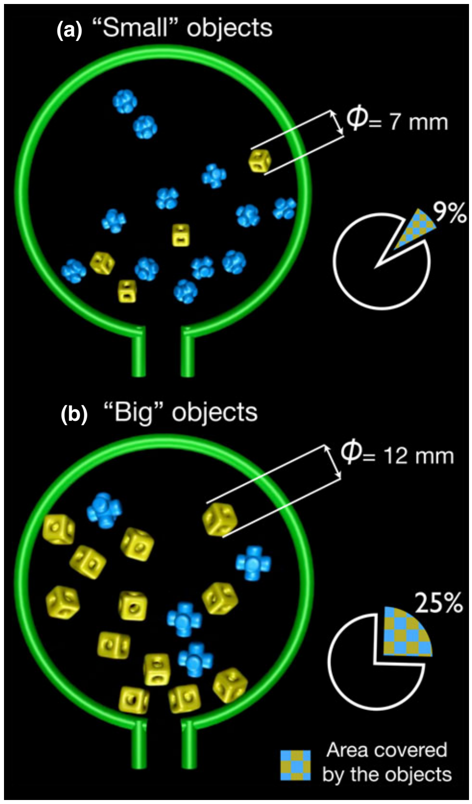

Figure 3 *A comparison of the sizes of the objects in Experiments 2 (small objects) and 3 (big objects). Because of the frequent object collisions, big objects provide density cues that small objects do not provide.*

reasoning by using a physical parameter that best fits with scenes containing many objects: density. We reasoned that in a crowded scene, the size of the objects may alter the perception of the likely outcomes. In Experiment 2, the objects were small and bounced freely inside the container. However, if the objects were larger, and their size and collisions constrained their movements, then the distance of the objects from the exit might become a relevant parameter. In a densely crowded frame, large objects cannot suddenly travel long distances, and infants may therefore consider it natural that an object would leave the container after the occlusion if it were close to the exit before the occlusion, regardless of whether or not it belongs to the most numerous or the less numerous class. If objects are far from the exit before occlusion, however, then class membership should determine what outcome infants expect. Namely, if all objects of the more numerous category are near the exit and the objects of the less numerous category are far from it, the latter could hardly travel down through the thick layer of objects separating them from the exit during the brief occlusion period. By contrast, if the objects of the more numerous category are separated from the exit by the objects of the less numerous category, no 'thick wall' prevents them from reaching the exit. Thus, surprise at the post-occlusion outcome should vary according to the category that is closer to the exit pre-occlusion. Infants may be surprised at a 4-instance exit, but only if the objects of the less numerous category are separated from the exit by the objects of the more numerous category. Such a result would offer more proof of the sophistication with which infants process dynamic scenes to form expectations about their future behavior. At the same time, it would prove that they can process and make inferences about scenes as complex as those presented in Experiment 2 once they are not required to reason about numerical differences.

Experiment 3

Method

Participants

Twenty-four healthy full-term 12-month-old infants were retained for analysis (10 girls, mean age 12 months, 17 days). An additional 15 infants were tested but not retained (seven became fussy, one fell sleep, one was excluded due to caretaker's interaction, one turned in synchrony with the outcomes, and four due to experimental error).

Materials and Procedure

We generated movies containing 16 objects, as in Experiment 2, but the size of each object covered approximately the 1.2% of the visible area of the container, or about 25% of its total area. The increased object size gave an impression of crowdedness and reduced the distance that any object could travel before colliding with other objects. Figure 3 shows a typical frame of the movies presented in Experiments 2 and 3, allowing one to gauge the object dimensions. The familiarization movies were constructed as in Experiment 2, except that the objects were bigger. Their motion trajectories were generated by using the same physical parameters used in Experiments 1 and 2. Because the objects were bigger than those of the previous experiments, the actual number of collisions could not be the same. However, the fact that the physical parameters applied were identical to those of Experiments 1 and 2 guaranteed that the collisions and the general dynamics in this experiment were as natural as those of Experiments 1 and 2, despite the difference in object size. The test movies always contained 4 objects of one category and 12 of the other category. At the beginning of the movies, the objects belonging to the two categories were intermixed; however, the motion of the objects was such that immediately before the occlusion the two categories would separate. Namely, in half of the test movies, immediately before the occlusion, all the objects of the less numerous category were positioned above all the objects of the more numerous category. In the other half of the movies, the spatial arrangement of the objects of the two categories relative to the exit was reversed (Figure 4). Because the objects were first intermixed and separated between categories only at the end of the movies, when they were looking at the screen infants necessarily noticed objects belonging to both categories, and not only those that finally landed in the vicinity of the exit point.

The position of the categories before the occlusion was crossed with the category membership of the exited object during the occlusion. Every participant saw all the combinations of pre-occlusion category position and post-occlusion category membership of the exited object. The other factors were counterbalanced across participants. Every other aspect of the procedure was identical to Experiment 2.

Results and discussion

A repeated measure ANOVA on single trials, with pre-occlusion category position (Close/Distant) and post-occlusion membership of the exited object (4-instance vs.

12-instance outcome) as the within-participant factors and participants as the random factor, revealed a main effect of post-occlusion object category membership and no main effect of category position. Infants looked longer at 4-instance outcomes ($M_{12\text{-instances}} = 13.4$ s, $M_{4\text{-instances}} = 16.2$ s; $F(1, 23) = 5.08$, $p \leq .04$), but looked for equal amounts of time to objects that, before the occlusion, were either far or distant from the exit ($M_{Close} = 14$ s; $M_{Distant} = 15.7$ s; $F(1, 23) = 1.1$, $p > .29$; Figure 5).

However, the result of main interest concerns the interaction between the two factors, because the predicted effect is a longer looking time for the Distant/4-instance outcome with respect to all the other conditions. This interaction was significant ($F(1,22) = 5.4$, $p \leq .03$). Most importantly, Scheffé post-hoc analyses revealed that the Distant/4-instance outcome was significantly different from all other conditions ($p \leq .02$ with the Distant/12-instance and the Close/4-instance conditions; and $p \leq .004$ with the Close/4-instance condition). By contrast, no other comparison was significant, including the comparison Close/12-instance vs. Close/4-instance. In short, the full effect of the interaction was carried by the higher looking time to the Distant/4-instance condition, as predicted.

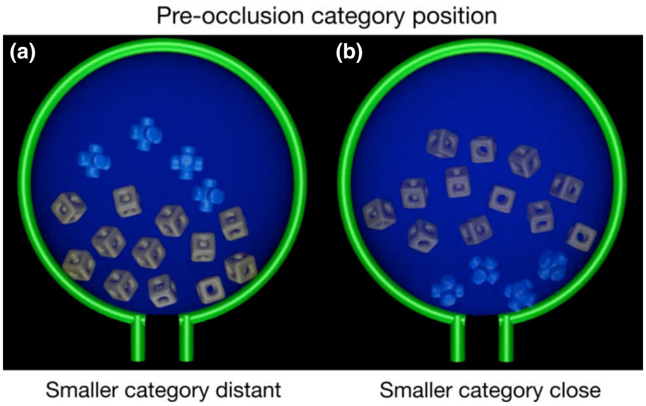

Figure 4 *An example of pre-occlusion object positions in Experiment 3. (a) When the occlusion begins, all objects of the less represented category are distant from the exit, and the objects of the more represented category create a dense layer that is difficult to bypass. (b) When, instead, all objects of the more represented category are distant from the exit, they have no obstacle in percolating down to the exit during occlusion.*

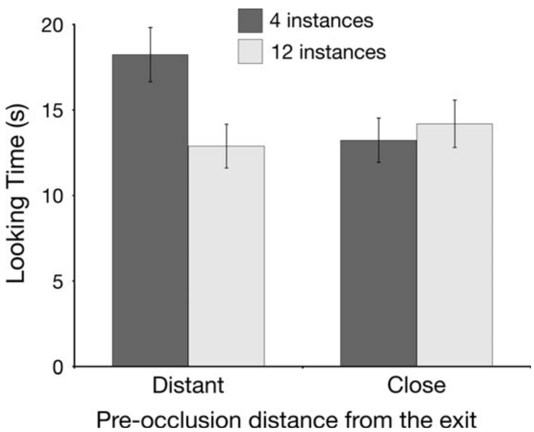

Figure 5 *Results of Experiment 3.*

This result excludes the possibility that the infants' failure in Experiment 2 was due to the general complexity of the scenes. The nature of the interaction also allows us to interpret the main effect of post-occlusion object category membership. This main effect does not indicate that infants could use numerical information from large classes to form expectations about single events. Had they done so, they would have also looked longer at the Close/4-instance condition than at the Close/12-instance condition. This did not occur. Instead, clearly the main effect of post-occlusion category membership was due to the unique condition in which the density information was relevant. Thus, infants' expectations were due entirely to their sensitivity to physical cues such as density. Interestingly, it appears that infants used such cues only when they were relevant, namely, in the Distant/4-instance condition.

There is, however, an alternative interpretation of these results that does not need to rely on density but rather on numerical reasoning. One could argue that infants were only paying attention to the subset of objects near the exit (say four items). When the subset near the exit corresponds to the most numerous category, they naturally predict that the outcome will be one of such objects, and hence they are surprised if this is not so. However, when the subset near the exit corresponds to the less numerous category, chances are that such a subset also includes items of the more numerous category. Hence, no expectation can be formed, since both outcomes are more or less equiprobable. This interpretation would account for the interaction observed in the experiment where infants only showed surprise in the Distant/4-instance condition.

Some evidence suggests that this alternative explanation is unlikely. Younger infants than those tested here appear to be able to track two sets of objects in parallel in displays very similar to the ones used in our experiment (Zosh *et al.*, 2011). Therefore, it seems unlikely that younger infants can attend to the totality of two ensembles simultaneously and that older infants focus only on partial regions of very similar displays. Thus, we believe that our results reveal that infants can exploit a physical cue such as density to form expectations about the next future event, when they cannot be helped by a numerical representation that does not allow them to represent single outcomes individually.

Besides showing that the failure to form expectations in Experiment 2 was not due to the sheer complexity of our stimuli, the current results show that infants can also extract fine-grained information from a situation, such as the length of an occlusion (Téglás *et al.*, 2011) or the density of a display, to form expectations about future events. This sophistication makes their failure to use numerical information when predicting single-case probabilities in scenes containing large quantities of objects all the more surprising. We discuss the relevance of this conclusion below.

General discussion

Increasingly, experimental results suggest that at 12 months infants already make sophisticated use of information to estimate the probability of events. In part, these skills depend on their capacity to track frequencies of experienced events, that is to say, to extract information from the past (e.g. Saffran, Aslin & Newport, 1996). In part, they depend on pure intuitions about non-experienced events (Téglás *et al.*, 2007, 2011). In the current paper, we have provided evidence that when infants are induced to reason about the next individual outcome of a scene that features two different categories – even a scene they have never encountered – they are surprised at improbable outcomes when objects of each category remain within the limit of parallel individuation. We have also shown that when such a limit is overcome, infants do not form similar expectations about single-case outcomes, although they can reason about very subtle physical parameters of a dynamic situation. If the same frame contains many objects that, by virtue of their size and number, may induce an impression of density, such as when one category became separated from the exit by a thick layer of objects of the second category, infants were surprised when they saw one object belonging to the more distant category exiting the frame. Considering these results, it is very surprising that 12-month-olds showed no sign of expecting the most probable outcome when they had to rely on the numerical information provided by the ratio between such categories. This fact is even more surprising considering that well before that age infants can discriminate less extreme ratios (Xu & Spelke, 2000; Xu, 2003; Xu & Arriaga, 2007); they can reason about abstract ratios (McCrink & Wynn, 2007) and perform simple algebraic operations over similarly large quantities (McCrink & Wynn, 2004). Yet, when the same ratio existed between categories remaining within the limits of parallel object individuation, they succeeded. Overall, the results suggest that probability intuitions about single events present the set size signature characteristic of object tracking (Feigenson & Carey, 2005). These intuitions are shaped by the systems of representation of quantities of the ensembles present in a scene. Furthermore, reasoning about single events in situations involving many objects is mostly driven by infants' processing of physical variables, not by numerical or ratio representations.

The results advance our understanding about how infants may form expectations about single events. Cesana-Arlotti, Téglás and Bonatti (2012) proposed that intuitions of probabilities depend on intuitions of possibilities. According to their hypothesis, infants represent the logical space of future possibilities afforded by the current situation. In order to reason about outcomes not yet experienced, they consider what states may occur (for example, that a yellow object will exit the container or that a blue object will exit), represent them individually (as one yellow-object outcome, another yellow-object outcome, another yellow-object outcome, and a blue-object outcome), quantify the possibilities (for example, determine that the cases where a yellow object may exit are 3 times as many as the cases where a blue object may exit), and intuit which one is more likely to occur on the basis of these estimates. Infants may use this procedure to reason forward, constructing future possible states and thus forming expectations about outcomes they have not yet experienced. Or they may use it while reasoning backward, when they try to make sense of a currently experienced outcome that does not fit their intuitions, and build counterfactual states in order to estimate its likelihood. In either case, from the assumption that infants conceive the world as a set of possibilities, together with the idea that possibilities are tracked individually, the existence of natural intuitions about the probability of non-experienced single events can be derived. A consequence of this proposal is that the ability to think about probable or improbable events also depends on how infants represent the future, or the counterfactual possibilities. If 12-month-olds rely on discrete representations of individual possibilities, then keeping track of individual objects or states in a scene may become crucial to reasoning about single-case outcomes. The ability to individuate objects of different classes in parallel may be the most natural representation from which to infer the possible continuation of a scene, and hence gives rise to intuitions of single-case, non-experienced outcomes. By contrast, the representation of large quantities makes room only for gross ratio differences, losing track of individual objects or events, and may not be suitable for representing possible single outcomes. Our results are broadly consistent with this hypothesis. Making the point in the theoretical framework developed by Zosh *et al*. (2011), infants may well be able to represent two ensembles in a scene regardless of the size of the ensembles, but it appears that representing different dynamic ensembles does not suffice to form intuitive and solid expectations of single-case probabilities: infants may need to represent such ensembles as sets of individual objects.

One alternative possibility focuses on the dynamic nature of our stimuli. In each experimental situation, infants see independently moving objects. By contrast, most of the studies in the literature typically use static displays (e.g. Xu & Spelke, 2000; Xu, 2003; Xu & Arriaga, 2007). Perhaps, infants might simply not compute numerical representations when seeing many independently moving objects. While this alternative explanation is possible, we consider it an unlikely account of the current results. While not widespread, the use of dynamic stimuli inducing the formation of numerical representations of large quantities is attested. McCrink and Wynn (2004) used moving objects to simulate additions and subtractions of large quantities, demonstrating that not only do infants younger than those tested here form numerical representations of large quantities of moving objects, but that they can also perform algebraic operations over them. Izard and her collaborators used large quantities of dots moving together, as collections (Izard, Sann, Spelke & Streri, 2009; Coubart, Izard, Spelke, Marie & Streri, 2014). The movements presented to infants were much simpler than those we created, but infants were also much younger. Even in this research, infants as young as 4 days old can form (multimodal) representations of large numerosities after seeing non-static displays. It is thus difficult to explain why infants should succeed at forming a precise numerical representation of four moving objects belonging to two categories, and yet fail at forming an approximate representation of objects of the same categories, with the same ratios, when more objects are present, but only in the case we studied. We would add that it is also difficult to imagine what natural pressure might have favored the selection of an evolutionarily ancient system shared by humans and animals (Feigenson *et al*., 2004; Brannon & Merritt, 2011), such as that of the representation of large approximate quantities, uniquely adapted to static stimuli.

Under certain conditions infants seem to draw inferences about single-case probabilities from large numerosities. Exploring the hypothesis orally presented by Bonatti (2008), Denison and Xu (2010) tested whether 12- to 14-month-olds can make a single choice between two alternatives originating from two populations of objects (lollipops). Infants were familiarized with two jars containing 50 lollipops, in a 4:1 or a 1:4 distribution of pink-to-black lollipops. Then, the experimenter randomly selected two lollipops, one from each jar (which remained statically visible to infants). She placed them in two different opaque cups in such a way that infants could see from which jars the lollipops were coming, but not which lollipop entered the cups. Because they could not see the lollipops in the cups, they had to use the visual information of the distribution in the jars to orient their choice. Infants chose the cups containing the lollipop coming from the jar which contained a higher number of the objects that they preferred, showing that they were able to exploit the 1:4 distribution to orient their choices. Denison and Xu (2010) interpret these results as showing that infants can reason about single-case probabilities with large set sizes. In the current experiments, we do not find traces of similar abilities. Several factors can explain the discrepancies between the results of Denison and Xu (2010) and our own. We only tested spontaneous surprise at a single unexpected event, but we did not require infants to select one alternative. As happens in other cases (e.g. Mehler & Bever, 1967), spontaneous preferences and active choices may not be subject to the same constraints and may recruit different reasoning paths or representations.[3] Furthermore, our experiments require memory representations of the relevant past information, whereas in Denison and Xu, the numerical information was continuously present. Finally, in their experiments Denison and Xu (2010) used a more extreme ratio difference. Such a difference is further increased by the fact that infants could not inspect the full population, but could only see the subset of objects appearing on the front of the jars, thereby making the visible ratio difference more extreme. It is possible that these differences trigger representations that, although not statistical in nature, may be appropriate for guiding choices, such as ‘most’ or ‘few’. That is, infants may create rough representations about what the jars contained, such as ‘Most lollipops are pink’. Afterwards, faced with the need to make a choice, they may decide to select the cup that was coming from the jar with the most pink lollipops. This is a rational strategy

[3] We would like to stress this point: intuitive preferences may not contribute to determine final choices, and conversely, the lack of intuitive preferences does not necessarily imply a lack of motive for a final choice. I may prefer ice-cream to apples, but I decide to grab an apple because I am on a diet. Likewise, I may have no preference for ice-cream over apples, but I may decide to eat ice-cream because it melts quickly and it will be spoiled otherwise. Thus, when making a decision, it is possible to choose an outcome without having an initial, spontaneous preference for one of the outcomes.

that would explain the results even in the absence of any spontaneous intuition of probability.[4] Other differences exist between the design of the current experiments and that of Denison and Xu (2010). Further research is needed to understand the precise nature of the representations of situations involving choices, motivational factors, or explicit action planning, as opposed to the simple spontaneous analysis of a scene, as revealed by the violation of expectation methods used here.

Our results are a partial confirmation of and challenge to Bayesian accounts of human reasoning. By showing that infants modulate their expectations of probabilities according to some reliable cues present in a scene (Experiments 1 and 3), they confirm and extend the findings of Téglás *et al*. (2011) that infants can rationally integrate information in the environment to issue an optimal prediction. The failure to use large number information when this is the only relevant cue (Experiment 2), however, does not fit naturally within a Bayesian framework. It is always possible to accommodate a hierarchical Bayesian account by enriching the background set of articulated theories about the mental representations that infants can exploit to frame their expectations about future events. In this case, however, the weight of the explanation resides in these theories, and not in the Bayesian mechanisms of inference. Our results suggest that it is precisely the interrelation between different levels of mental representations that is responsible for infants' behavior.

Our experiments speak to the complexity, the limits, and the sophistication of the mental representations constructed when, at an early stage of development, infants cope with uncertain situations and form expectations about their future continuation. They were motivated by the hypothesis that intuiting possibilities is the basis for intuiting probabilities, and that a system for enumerating possibilities, independent of systems to track previously experienced events, is involved in shaping such intuitions. We have presented experiments in which very comparable situations, differing only in the number of objects, do or do not elicit intuitions of probabilities. Our data are consistent with the hypothesis. They suggest that probability reasoning may be a by-product of reasoning about possibilities, modulated by the actual number of possibilities infants can represent. Such immediate reasoning may be restricted to situations involving objects that can be individually represented within the sets to which they belong. Speculating on the widely documented errors that adults make when reasoning about single cases, we submit that perhaps they may have their origin in this constraint. Humans may be able to reason probabilistically about single, non-experienced cases, but only when the ancillary systems of representations needed to represent possibilities are appropriate. The representation of quantities we are all accustomed to, with its deceptively undifferentiated aspect, may hide important distinctions that are crucial for such reasoning to be successful.

[4]The experiments by Denison and Xu (2010) are not sufficient to conclude that infants respond on the basis of such intuitions. In this work, information about what infants would do with small categories is lacking. We do not know how they would have behaved had they been tested with jars containing similar ratios, but confined within the subitizing range. It is not impossible that the representations that infants can use to generate intuitions about single future events are not the same as those used to make decisions and guide actions. That is, it is conceivable that infants may solve our tasks with situations involving small, but not large, numerosities, while failing a decision task with small, but not large, numerosities.

Acknowledgements

This work was supported by grants PSI2012-31961 and FVG 'PsyScope XL' to LLB, and by the Hungarian Science Foundation Grant OTKA NK83997. We thank J. Brick, A. Isaja, F. Gandolfo, K. Mehta and L. Filippin for comments and support. We also thank J. Mehler and the LCD Lab at Trieste, Italy, for providing great material and moral support for this research.

References

Bonatti, L.L. (2008). At the origin of rationality: how intuitions of probabilities shape predictions about the future. Paper presented at the 16th International Conference on Infant Studies, Vancouver, Canada.

Brannon, E.M., & Merritt, D.J. (2011). Evolutionary foundations of the approximate number system. In S. Dehaene & E.M. Brannon (Eds.), *Space, time, and number in the brain: Searching for the foundations of mathematical thought* (pp. 207–224). London: Academic Press.

Carey, S. (2009). *The origin of concepts*. New York: Oxford University Press.

Cesana-Arlotti, N., Téglás, E., & Bonatti, L.L. (2012). The probable and the possible at 12 months: intuitive reasoning about the uncertain future. *Advances in Child Development and Behavior*, **43**, 1–25.

Cosmides, L., & Tooby, J. (1996). Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgment under uncertainty. *Cognition*, **58** (1), 1–73.

Coubart, A., Izard, V., Spelke, E.S., Marie, J., & Streri, A. (2014). Dissociation between small and large numerosities in newborn infants. *Developmental Science*, **17** (1), 11–22.

Denison, S., & Xu, F. (2010). Twelve- to 14-month-old infants can predict single-event probability with large set sizes. *Developmental Science*, **13** (5), 798–803.
Feigenson, L. (2005). A double-dissociation in infants' representations of object arrays. *Cognition*, **95** (3), B37–B48.
Feigenson, L., & Carey, S. (2005). On the limits of infants' quantification of small object arrays. *Cognition*, **97** (3), 295–313.
Feigenson, L., Dehaene, S., & Spelke, E.S. (2004). Core systems of number. *Trends in Cognitive Sciences*, **8** (7), 307–314.
Feigenson, L., & Halberda, J. (2004). Infants chunk object arrays into sets of individuals. *Cognition*, **91** (2), 173–190.
Feigenson, L., & Halberda, J. (2008). Conceptual knowledge increases infants' memory capacity. *Proceedings of the National Academy of Sciences of the United States of America*, **105** (29), 9926–9930.
Gigerenzer, G., & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: frequency formats. *Psychological Review*, **102** (4), 684–704.
Griffiths, T.L., & Tenenbaum, J.B. (2006). Optimal predictions in everyday cognition. *Psychological Science*, **17** (9), 767–773.
Izard, V., Sann, C., Spelke, E.S., & Streri, A. (2009). Newborn infants perceive abstract numbers. *Proceedings of the National Academy of Sciences of the United States of America*, **106** (25), 10382–10385.
Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review*, **80** (4), 237–251.
McCrink, K., & Wynn, K. (2004). Large-number addition and subtraction by 9-month-old infants. *Psychological Science*, **15** (11), 776–781.
McCrink, K., & Wynn, K. (2007). Ratio abstraction by 6-month- old infants. *Psychological Science*, **18** (8), 740–745.
Mehler, J., & Bever, T.G. (1967). Cognitive capacity of very young children. *Science*, **158** (797), 141–142.
Pylyshyn, Z.W. (2001). Visual indexes, preconceptual objects, and situated vision. *Cognition, 80 (1–2)*, **12**, 7–158.
Pylyshyn, Z.W. (2007). *Things and places: How the mind connects with the world*. Cambridge, MA: MIT Press.
Saffran, J.R., Aslin, R.N., & Newport, E.L. (1996). Statistical learning by 8-month-old infants. *Science*, **274** (5294), 1926–1928.
Scholl, B.J., & Pylyshyn, Z.W. (1999). Tracking multiple items through occlusion: clues to visual objecthood. *Cognitive Psychology*, **38** (2), 259–290.
Slaughter, V., & Suddendorf, T. (2007). Participant loss due to 'fussiness' in infant visual paradigms: a review of the last 20 years. *Infant Behavior and Development*, **30** (3), 505–514.
Southgate, V., Johnson, M.H., El Karoui, I., & Csibra, G. (2010). Motor system activation reveals infants' on-line prediction of others' goals. *Psychological Science*, **21** (3), 355–359.
Southgate, V., Johnson, M.H., Osborne, T., & Csibra, G. (2009). Predictive motor activation during action observation in human infants. *Biology Letters*, **5** (6), 769–772.
Téglás, E., Girotto, V., Gonzalez, M., & Bonatti, L.L. (2007). Intuitions of probabilities shape expectations about the future at 12 months and beyond. *Proceedings of the National Academy of Sciences of the United States of America*, **104** (48), 19156–19159.
Téglás, E., Vul, E., Girotto, V., Gonzalez, M., Tenenbaum, J.B., & Bonatti, L.L. (2011). Pure reasoning in 12-month-old infants as probabilistic inference. *Science*, **332** (6033), 1054–1059.
Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: heuristics and biases. *Science*, **185** (4157), 1124–1131.
Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, **211** (4481), 453–458.
Xu, F. (2003). Numerosity discrimination in infants: evidence for two systems of representations. *Cognition*, **89** (1), B15–B25.
Xu, F., & Arriaga, R.I. (2007). Number discrimination in 10-month-old infants. *British Journal of Developmental Psychology*, **25** (1), 103–108.
Xu, F., & Spelke, E.S. (2000). Large number discrimination in 6-month-old infants. *Cognition*, **74** (1), B1–B11.
Zhu, L., & Gigerenzer, G. (2006). Children can solve Bayesian problems: the role of representation in mental computation. *Cognition*, **98** (3), 287–308.
Zosh, J.M., Halberda, J., & Feigenson, L. (2011). Memory for multiple visual ensembles in infancy. *Journal of Experimental Psychology: General*, **140** (2), 141–158.

Received: 15 February 2013
Accepted: 18 March 2014