Being the third wheel: Toddlers use bystander learning to acquire cue-specific valence knowledge

Bridget L. Callaghan a,b,*, Tricia Choy a, Kaitlin O’Sullivan c, Emma Routhier d, Nora Cabrera d, Victoria Goode d, Tovah Klein d, Nim Tottenham a

a Department of Psychology, Columbia University, New York, NY 10027, USA
b Department of Psychology, University of California, Los Angeles, Los Angeles, CA 90095, USA
c Department of Emergency Medicine, School of Medicine, Georgetown University, Washington, DC 20007, USA
d Barnard Center for Toddler Development, Barnard College, Columbia University, New York, NY 10027, USA

Article history:
Received 10 September 2021
Revised 24 January 2022
Available online 8 March 2022

Keywords:
Threat conditioning
Discrimination
Child development
Emotional eavesdropping
Observational learning
Memory retention

Abstract

Observing others is an important means of gathering information by proxy regarding safety and danger, a form of learning that is available as early as infancy. In two experiments, we examined the specificity and retention of emotional eavesdropping (i.e., bystander learning) on cue-specific discriminant learning during toddlerhood. After witnessing one adult admonish another for playing with Toy A (with no admonishment for Toy B), toddlers learned to choose Toy B for themselves regardless of whether they were tested immediately or 2 weeks later (Experiment 1). However, if asked to make a toy choice for someone else (i.e., when toddlers’ personal risk was lower), approximately half the toddlers instead selected Toy A (Experiment 2). However, such choices were accompanied by toddlers’ social monitoring of the adults, suggesting that toddlers may have been attempting to safely gain (via surrogacy) more information about risk contingencies. These findings suggest that toddlers can learn to discriminate valence in a cue-specific manner through social observation.

© 2022 Elsevier Inc. All rights reserved.
Introduction

Theoretical models of emotional learning are overwhelmingly based on firsthand learning experiences—for example, classical conditioning paradigms that pair aversive outcomes (e.g., shock, loud noises) with previously neutral stimuli (e.g., tones). Although such firsthand experiential conditioning models the acquisition of some fears, in the real world, learning occurs through a variety of indirect pathways, including social observation. By observing others’ learning about threat-cueing stimuli, it has been shown that adults can circumvent the need to directly subject themselves to negative outcomes yet acquire valuable emotional information (Olsson, Feldman-Hall, Haaker, & Hensler, 2018). Thus, learning via observation has adaptive value that takes advantage of humans’ social nature.

Observing others learn about potential danger in the world is especially useful very early in life at a time when there are fewer opportunities for firsthand interaction with events/objects. It has been well established that by the first year of life infants’ behavior is guided in a valence-specific manner (i.e., positive or negative emotional tones) by observing how an adult reacts to events or objects (Feinman & Lewis, 1983; Gunnar & Stone, 1984; Hornik, Risenhoover, & Gunnar 1987; Klinnert, Emde, Butterfield, & Campos, 1986; Sorce, Emde, Campos, & Klinnert, 1985), with some added benefit if this adult is the parent or the parent is near (Klinnert et al., 1986; Zarbatany & Lamb, 1985). Moreover, studies have shown that these effects can be cue specific as early as infancy rather than reflecting behavioral alterations as a function of global change in infant mood (Feinman & Lewis, 1983; Sorce et al., 1985).

Not only can social information guide current behavior, but learning and remembering predictive cues for negative outcomes via observation also has been observed during development (Silvers et al., 2021), with emergence of these skills seen as early as infancy. Infants and toddlers can learn to attach emotional valence to predictive cues provided by others (Gerull & Rapee 2002; Gewirtz & Peláez-Nogueras, 1992; Patzwald, Curley, Hauf, & Elsner 2018) even if the actors were observed in videos (Mumme & Fernald, 2003). For example, by 14–18 months of age toddlers were able to use an actor’s attentional cues (gaze and action) to regulate their behavior toward specific toys, playing more with a toy that was associated with the actor’s happy reaction than with the actor’s disgust reaction (Repacholi, 1998). Infants and toddlers aged 11–20 months also know whether an emotional reaction is related to a toy at all (Moses, Baldwin, Rosicky, & Tidball, 2001). Meltzoff, Waismeyer, and Gopnik (2012) used the term “observational causal learning” to describe the ability to learn and infer cause–effect associations by observing others’ causal acts. Together, these studies demonstrate that the capacity for observational causal learning becomes increasingly sophisticated over the first several years of life.

Typically, in studies of observational learning, the emotional information is directly communicated to infants by their social partner. As described by Repacholi and Meltzoff (2007), young children also engage in indirect social learning by eavesdropping on emotional interactions between social partners (e.g., watching a parent scold a sibling for drawing on the wall, watching two parents argue over a household preference). That is, they can engage in observational causal learning even when the emotional message is intended for another person. In a seminal emotional eavesdropping study (Repacholi & Meltzoff, 2007), 18-month-old toddlers observed two actors engage in an emotional exchange (either neutral or negative in a between-participants design) regarding a toy. The bystanding 18-month-olds then used this ‘overheard’ information to guide their object-directed actions, with toddlers who observed a negative response playing less with the toy than those who observed a neutral response, but only if they were being watched by the emoter (Repacholi, Meltzoff, Hennings, & Ruba, 2016). In other words, toddlers aged 18 months and older used a “second-order” observational causal learning process to guide their own behavior in subsequent trials. This learning seemed to occur not because toddlers changed their own emotion but rather through observation; they indirectly learned about how one actor (i.e., the emoter) felt about another actor’s use of an object. By 24 months of age, the ability of toddlers to engage in observational social learning is well established, with studies demonstrating efficient social learning through observing adult–adult interactions by 24–48 months of age.
The current study

The current study extends this line of research by asking whether emotional eavesdropping can produce cue-specific discriminant learning in toddlers (24–46 months of age). Notably, although this age range includes children in the toddler and preschool periods, for the sake of simplicity we use the term toddler to refer to our sample throughout this article. That is, although toddlers have been shown to learn by being bystanders, this effect has been demonstrated via presentations of single objects in isolation in a between-participants design. Here, we asked whether toddlers would be able to use bystander learning to distinguish between two simultaneously presented objects—one for which they have witnessed a negative social exchange and one for which they have witnessed a neutral social exchange (using a within-participants design).

Presenting toddlers with two stimuli answers an important question that could not be answered by past research. Namely, if presented with only one stimulus during an angry adult interaction, do toddlers learn that the emoter dislikes only one of the toys or might toddlers engage their priors about angry adults and anticipate a general negative outcome in a nonspecific way (i.e., a trait-level judgment)? On the other hand, presentation of two stimuli (only one of which creates the angry adult interaction) offers an opportunity to assess whether toddlers are applying a generalized "bad" prediction or they can discriminate those stimuli that instigated the angry interaction from those that did not. It remains unclear whether toddlers can learn via this complicated form of observational learning (when multiple toys could be the target of an emoter's affect). Generalized learning and cue-specific learning are distinct and rely on different neurobiology (Asok, Kandel, & Rayman, 2018), and discriminant cue-specific learning, being more of a sophisticated process than generalized learning, typically tends to develop later (Nuttall, Valentino, Comas, McNeill, & Stey, 2014; Valentino et al., 2014) and has different consequences for toddlers' behavior. For example, learning that an adult has an aversion for one stimulus but not another should provide a more accurate representation of potential sources of positive and negative emotional events in an ever-changing social environment.

As such, in the current study, we employed a within-participants design to test whether toddlers could achieve complicated discriminatory (cue-specific) observational learning via emotional eavesdropping. In addition, we tested whether this learning varied as a function of time (immediate vs. several weeks' delay). It is established that retention of experiential memories improves dramatically across the first several years of life (e.g., see Hayne & Imuta, 2011, for a review). For example, operant memory improves linearly across 6–18 months of age, with the oldest infants/toddlers retaining the memory for more than 13 weeks (Hartshorn et al., 1998), and similar retention intervals have been reported using deferred imitation (Bauer & Dow, 1994). In addition, discriminatory observational learning (although not through emotional eavesdropping) can be retained for at least 1 h in 14-month-olds (Hertenstein & Campos, 2004). Associative memories were retained for at least 24 h in 36-month-olds, but accuracy was at chance level 1 week after learning at that age (Saragosa-Harris et al., 2021). In terms of emotional memory, 35- to 65-month-olds were shown to exhibit good immediate recall of an individual from a video clip's emotions, with performance depending on the own emotion understanding of the toddler/preschooler/kindergartener (Channell & Barth, 2013). In another study, 34- to 64-month-old toddlers/preschoolers/kindergarteners were shown to have good emotional memory recall over 24 h, but performance was enhanced after napping (Kurdziel, Kent, & Spencer, 2018). Given the evidence that older infants can retain experiential memories for at least a week, and that children up to 65 months of age (kindergarteners) have good immediate and 24-h delay memory, in the current study we tested the retention of cue-specific discriminant learning via emotional eavesdropping over a period of approximately 2 weeks in toddlers (24–46 months of age). This age range was selected because we expected that observational learning would be a more challenging task than the tasks used in previous memory studies in younger infants and toddlers (Bauer & Dow, 1994; Hartshorn et al., 1998) and that there may be variability in 2-week memory retention across this period of development (Saragosa-Harris et al., 2021).
In the current study, each toddler observed an adult (emoter) expressing neutral affect toward another adult (experimenter) in reference to one toy and expressing annoyed affect in reference to a second similar toy. We assessed annoyed versus neutral affect (within-participants) because we were specifically interested in threat learning via observation, and annoyance is a relatively threatening emotion to which toddlers would have routine access from parents (e.g., in response to misbehavior). Following toddlers’ bystander observation, we asked whether toddlers used that information over time (immediately to approximately 2 weeks after learning using a between-participants comparison) to guide their own behavior toward the target toys (Experiment 1) and to guide another person’s behavior toward the toys (Experiment 2).

Hypotheses

Experiment 1. There were three hypotheses in Experiment 1. First, we hypothesized that toddlers’ positive affect would be lowest when the emoter was expressing annoyed affect, relative to the other training phases. Second, we hypothesized that toddlers would make more choices toward the safe toy and play longer overall with the safe toy and that any toddlers who did choose the risky toy would take longer to do so than those who chose the safe toy. The third hypothesis (which was exploratory) was that toddlers who expected the emoter to get annoyed (i.e., who had a memory of the negative event) would monitor the adults more during periods of interaction with the risky toy. That is, we might observe a difference in social monitoring behavior between the risky and safe toys even for toddlers who did not inhibit their play with the risky toy, which would suggest that toddlers expected different outcomes for the two toys.

Experiment 2. In Experiment 2, we hypothesized that more toddlers would choose the risky toy than the safe toy for a surrogate actor (i.e., they could use the surrogate actor to obtain more information about the actions of the emoter). We also hypothesized that if toddlers chose the risky toy for a surrogate to gather more emotional information about the actors, they would also engage in more social monitoring of the actors than toddlers who chose the safe toy for the surrogate.

Experiment 1

Method

Participants

We used a convenience sample of $N = 40$ toddlers who were enrolled at the university’s preschool for at least one term during the year 2015. One male toddler (37 months old) in the delay condition was trained but refused to return for the delay test. Thus, our final sample consisted of $N = 39$ toddlers aged 26–42 months, with 1 toddler being an outlier in age at 47 months (mean age = 33 months, 41% boys and 59% girls; see Table 1 for sample demographics broken down by train–test delay). The number of students enrolled in the preschool in 2015 whose parents consented to their participation dictated this sample size. The university’s institutional review board approved the protocol. Parents provided written consent for their toddlers’ participation.

Design

A mixed within/between-participants design was used in the experiment. The within-participants condition was emoter’s affect (neutral vs. annoyed) expressed toward the experimenter regarding one of the two target toys (monkey or fish) (see Fig. 1). To ensure that there were no carryover effects of the annoyed affect into the neutral condition, we fixed the order of emoter’s affect as follows: neutral first, followed by annoyed. The target toys associated with emoter’s neutral versus annoyed affect were counterbalanced between toddlers. The same actors posed as the experimenter and the emoter throughout the entire experiment. The between-participants condition was the train–test interval (immediate vs. delay) to which toddlers were randomly assigned. Toddlers in the delay group were tested 14–20 days after training (mode = 14 days).
Table 1
Demographic information across Experiments 1 and 2.

<table>
<thead>
<tr>
<th>Experimental condition</th>
<th>n</th>
<th>Sex (% male/ female)</th>
<th>Average delay in days between training and test (range)</th>
<th>Mean age in years at training (range)</th>
<th>Toddlers' race/ethnicity (n)</th>
<th>Parental marital status (%)</th>
<th>Maternal education (%)</th>
<th>Paternal education (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experiment 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Immediate</td>
<td>19</td>
<td>53/47</td>
<td>0 (0)</td>
<td>2.84 (2.4–3.9)</td>
<td>Caucasian (14), Asian American (4), Latino/a (1)</td>
<td>Married (95), separated (5), never married (0)</td>
<td>Bachelor’s (26), graduate (74)</td>
<td>Bachelor’s (47), graduate (53)</td>
</tr>
<tr>
<td>Delay</td>
<td>20</td>
<td>30/70</td>
<td>15.5 (14–20)</td>
<td>2.56 (2.2–3.3)</td>
<td>Caucasian (19), Asian American (1)</td>
<td>Married (95), separated (5), never married (0)</td>
<td>Bachelor’s (25), graduate (75)</td>
<td>Bachelor’s (21), graduate (79)</td>
</tr>
<tr>
<td>Experiment 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Immediate</td>
<td>17</td>
<td>41/59</td>
<td>0 (0)</td>
<td>2.71 (2.0–3.0)</td>
<td>Caucasian (13), Asian American (1), Latino/a (1), African American (2)</td>
<td>Married (94), separated (0), never married (6)</td>
<td>Bachelor’s (35), graduate (65)</td>
<td>High school (18), bachelor’s (53), graduate (29)</td>
</tr>
<tr>
<td>Delay</td>
<td>47</td>
<td>51/49</td>
<td>14.2 (10–22)</td>
<td>2.34 (2.0–3.0)</td>
<td>Caucasian (32), Asian American (8), Latino/a (4), African American (1), Native American (1), Black (1)</td>
<td>Married (94), separated (4), never married (2)</td>
<td>High school (2), bachelor’s (30), graduate (68)</td>
<td>Bachelor’s (40), graduate (60)</td>
</tr>
</tbody>
</table>
Materials
Two adult actors (experimenter and emoter) and a familiar teacher from the preschool were involved in the experiment. The teacher remained neutral and out of view of the toddler during the experiment. Testing occurred at the preschool during regular school hours. The rectangular testing room had a door at either end; the toddler entered one door from the schoolroom, and the emoter entered through the second door. The room contained a small table with two toddler-sized chairs positioned on either side (the experimenter sat facing the toddler entry door). The teacher sat behind the toddler and the emoter sat to the left of the experimenter, both on adult-sized chairs. To the right of the experimenter was an opaque toy box, which contained the warm-up and cool-down toys as well as the two target toys (see Fig. 1 for a photograph and description of the two target toys). Two video cameras, one aimed at the toddler's face and another wide-angle view of the room, were used to record the session.

Procedure
This study began during the fourth week of the term when toddlers were adjusted to parental absence and to the routine of the center. The experimenter attended the preschool classes on two separate occasions a week before the study began. On those “familiarization” days, the teachers at the preschool introduced the experimenter by name and said that she was here to “watch the toddlers play.”

There were four distinct parts to the study: (a) warm-up, (b) target toy training, (c) testing (either immediate or at a delay; see Fig. 2B for a graphical summary of those phases), and (d) cool-down (see online supplementary material for a script of the training and testing sessions). To start the study, the teacher approached a toddler in the classroom and told him or her that the experimenter had some new toys, asking whether the toddler would like to play with them. If the toddler agreed, the teacher
and toddler would walk to the adjoining testing room and the teacher would seat the toddler on the chair.

Warm-up. After the toddler was seated, the experimenter pulled out the warm-up toys (Slinky and kaleidoscope) one by one and allowed the toddler to play with each toy for approximately 20 s unless the toddler’s attention waned earlier. The experimenter encouraged the toddler to play with the toys using phrases such as “You can play,” “Here you go,” and “It does this.”

Training phase: Target toy demonstration. Two novel toys were designed by the research group for the purposes of this study and acted as the target toys (see Fig. 1B). In short, the target toys were similarly constructed but were different in theme (monkey in jungle vs. fish in sea). Each toy consisted of a box containing a hidden noisemaker. A plush toy (monkey or fish) sat on top of the box, and pressing down
on the plush toy activated the buzzer (which made a “honk” sound for the monkey and a “boing” sound for the fish) (see Fig. 1A). The experimenter introduced the target toy demonstration phase by saying, “Now I have some very special toys for you to play with, I made them myself, so we have to be careful. I am going to show you what they do, then it will be your turn to play.” The experimenter then took the first toy, described it (see Fig. 1C), and demonstrated the noise action before putting away the first toy, taking out the second toy, describing it, and demonstrating its noise. The order of toy presentation was counterbalanced between toddlers.

**Training phase: Emotional exchange.** After the actions of the two target toys had been demonstrated to the toddler, the emoter knocked on the door and entered the room. The experimenter introduced the emoter to the toddler, and the emoter sat down. The experimenter then demonstrated each of the toys to the emoter. The emoter displayed neutral affect in response to the first toy (safe toy) and displayed annoyed affect in response to the second toy (risky toy). The experimenter then remarked “Oh, I thought it [the risky toy] was interesting” before moving on to the test phase.

**Testing phase.** After the emoter’s remark about the risky toy, the toddler was tested. In the immediate condition, both target toys were placed in front of the toddler on the table and the toddler was told that it was now his or her turn to play (Fig. 2B). The emoter and experimenter then adopted neutral facial expressions and looked toward the center of the table between the two toys for 20 s. Neither the emoter nor the experimenter communicated with the toddler during the play period. In the delay condition, the experimenter then said to the toddler, “I will bring these toys back another day for you to play with.” Approximately 2 weeks later, the experimenter returned and the warm-up and target toy demonstration phases were repeated, with the change that the experimenter inserted “remember” language into the demonstration (e.g., “Remember last time I showed you these toys …”). After the toy demonstration, the emoter then entered the room, and the toddler was told it was now his or her turn to play (the procedure for the immediate test was then followed).

**Cool-down.** At the end of the testing period (for the immediate condition), and at the end of the emoter’s display of annoyed affect (for the delay condition), the experimenter brought out the two cool-down toys (ball-in-tube and bells-in-tube), encouraging the toddler to play with them (maximum of 20 s with each toy). The experimenter interacted with the toddler during the cool-down session. For the immediate condition, the emoter excused herself before the cool-down toys were brought out. For the delay condition, the emoter remained in the room for the cool-down toys, but began reading a book (i.e., was no longer paying attention to the toddler or the experimenter).

**Scoring of toddlers’ behavior**

Toddlers’ overt emotional expressions across the course of the training phase (target toy demonstration and neutral and negative emotional exchanges) were video-recorded and coded. We separated the training phase of the study into four separate periods: (a) experimenter demonstrates target toys, (b) emoter enters the room, (c) emoter displays neutral affect, and (d) emoter displays annoyed affect. Toddlers’ learning was assessed through three behavioral indices measured during the testing phase: (a) toddlers’ choice of toy to play with, (b) latency to make the choice, and (c) duration of engagement with each toy. In addition, we also measured toddlers’ social monitoring of the actors when toddlers were engaged with each toy as a proxy for their expectations of the actors’ emotions. Trained observers who were blind to the aims, design, and group assignment (immediate vs. delay) of toddlers scored the adults’ and toddlers’ emotions and toddlers’ behaviors (see Supplemental Fig. 1a in supplementary material for a visual depiction of the coding schedule in Experiment 1).

**Emotional expression across each training phase.** During each stage of the training phase (target toy demonstration, emoter entering the room, emoter displaying neutral affect, and emoter displaying annoyed affect), observers assigned each toddler a single score for positive affect based on the toddler’s general positive tone during that stage. The positive affect scale ranged from 0 to 3 (0 = absence of positive affect, 1 = interest, 2 = slight smile, 3 = broad smile or laughter).
Choice of toy to engage with first. Toddlers' choice of what toy to first engage with was considered to be the most unbiased indication of their learning because it occurred prior to any potential feedback from the emoter about toddlers' behaviors. To be counted as a choice, toddlers needed to touch a toy within the 20-s toddler testing phase. Toddlers' behavior for this variable therefore was coded as “safe choice,” “risky choice,” or “no choice.” Although few toddlers were in the “no choice” group, we opted to include them in the analyses because their complete behavioral inhibition, overriding their initial motivation to play with the toys and instead sitting in the room for 20 s not playing, deserved to be modeled.

Latency to make a choice. We calculated the length of time in seconds it took toddlers to make a choice. Latencies were positively skewed and were log(base 10 + 1) transformed (because there was 1 participant with a latency of zero). Toddlers who were coded as “no choice” did not receive a latency score.

Duration of engagement with each toy. We considered that toddlers either may show a sustained aversion to the risky toy across the 20-s test period or may initially choose the safe toy but progress to the risky toy after building confidence that the emoter would not react to their behaviors (transitory aversion). To test for sustained aversion, we calculated the duration of time in seconds that toddlers spent engaged with each toy across the 20-s testing phase. Toddlers did not need to be in direct contact with the toy to be counted as engaged with it. For example, if toddlers pressed the buzzer and then lifted their hand while looking at the toy before then touching it again, that entire engagement period would be coded continuously. Toddlers could also be engaged with both toys simultaneously—for example, if they touched both toys at the same time. The analyses that looked at the duration of engagement with each toy controlled for latency to engage with any toy because these are not independent measures. Toddlers who were coded as “no choice” did not receive a duration of engagement score.

Social monitoring tied to toy engagement. From a camera that was focused on toddlers’ face, we coded bouts of fixation on the experimenter and emoter (social monitoring). We considered bouts of fixation only if they immediately preceded, occurred during, or immediately followed a toy engagement period (see Supplemental Fig. 1a). “Immediate” was defined as moving from the past point of fixation to the social target fixation with no fixations in between. Importantly, although bouts of monitoring were triggered by engagement with a toy (e.g., pointing at or touching a toy and then looking at the emoter/experimenter), monitoring periods could extend beyond the toy engagement period. For that reason, monitoring tied to toy engagements could be longer or shorter than the overall length of the toy engagements themselves. Toddlers who were coded as “no choice” did not receive a social monitoring score.

Statistics and analysis

Data were analyzed with parametric tests unless there was uneven distribution of toddlers between the groups (e.g., when looking at behavior as a function of initial toy choice). For parametric tests, in conditions where the assumption of sphericity was violated, we report the Greenhouse–Geisser adjusted $p$ values and the nominal degrees of freedom. Because toddlers were from the same class at the preschool, and because we did not design the study to have the power to test differences in behavior across months of age, we did not consider toddlers’ age in our analyses. However, see Appendix in e supplementary material for primary choice analyses stratified by toddlers’ age. Notably, the age of toddlers was not associated with any of the outcome variables in these supplemental analyses. Sex of toddlers was entered as a covariate in all analyses.

Results of main analyses

Toddlers’ emotional expression across each of the four periods of the training phase

A repeated-measures analysis of variance (ANOVA) was performed to examine the change in toddlers’ positive affect across each of the four training phases (within-participants variable): (a) experimenter demonstrates target toys, (b) emoter enters the room, (c) emoter displays neutral affect, and (d) emoter displays annoyed affect. The train–test interval (immediate vs. delay) was a between-
participants variable, and sex was a covariate. There was a significant within-participants effect of the training phase, $F(3, 105) = 17.45, p < .0001, \eta^2_p = .33$, indicating that positive affect changed across the course of training (see Fig. 2A). However, there was no interaction between training phase and train–test interval, $F(3, 105) = 0.86, p = .466, \eta^2_p = .02$, or between training phase and sex, $F(3, 105) = 2.52, p = .078, \eta^2_p = .07$. There was a between-participants effect of sex, $F(1, 35) = 4.79, p = .035, \eta^2_p = .12$, whereby girls expressed overall lower positive affect than boys. The between-participants effect of train–test interval was not significant, $F(1, 35) = 0.27, p = .607, \eta^2_p = .01$. Because of the significant within-participants effect of training phase on positive affect, we performed pairwise comparisons of the estimated marginal means between the training phase where the emoter expressed annoyed affect and all other training phases (with Bonferroni adjustment for multiple comparisons, three contrasts: $z = .017$). Toddlers’ positive affect when the emoter displayed annoyed affect was lower than when the experimenter demonstrated the target toys, $p < .0001$, confidence interval (CI) $[−1.68, −0.73]$, but was not different from when the emoter displayed neutral affect, $p = .037$, CI $[−1.03, −0.02]$, and when the emoter entered the room, $p = 1$, CI $[−0.55, 0.33]$.

First toy choice

We examined toddlers’ first toy choice as an indicator of their preferences before they could assess whether the emoter would respond to their behaviors. We performed a multinomial logistic regression to examine whether there was an overall bias in choice behavior toward the safe choice, risky choice, or no choice. The model was significant, suggesting that choices were not evenly distributed, $\chi^2(2) = 20.22, p < .0001$ (see Fig. 3A). Post hoc tests revealed that more toddlers chose the safe toy than the risky toy, $\beta = 0.82$, Wald $\chi^2(1) = 5.15, p = .023$, and toddlers chose the safe toy and the risky toy more than making no choice, $\beta = 2.12$, Wald $\chi^2(1) = 12.04, p = .001$, and $\beta = 1.30$, Wald $\chi^2(1) = 3.97, p = .046$, respectively. We also tested whether there was a bias in toddlers’ choice behavior as a function of train–test interval or sex (although a post hoc power analysis revealed that we were adequately powered at 80% to observe only a large effect size for these analyses). Toddlers’ first toy choice did not differ between the immediate and delay groups, $\chi^2(2) = 0.44, p = .803$, Cramer’s $V = .11$ (but see Fig. 3B for choice behavior broken down by immediate and delay groups for illustration), nor did it differ between boys and girls, $\chi^2(2) = 1.17, p = .558$, Cramer’s $V = .17$.

Latency to choose

We assessed differences in choice latency between the risky toy and safe toy. There were 3 toddlers who did not engage with any toy across the test period and were excluded from this analysis, leaving

![Fig. 3](image-url)
There was no main effect of train–test interval, whether latency to make a choice differed as a function of train–test interval, controlling for sex.

There was no significant within-participants effect of engagement duration as a function of the train–test interval. We used repeated-measures ANOVA, controlling for latency to make a choice, and train–test interval as a function of first toy choice. There was a significant effect of first toy choice on the mean rank difference in engagement duration between the safe toy and risky toy, \( U = 80, p = .049, d = −0.66 \), whereby toddlers who made the safe choice first spent relatively longer engaging with the safe toy than with the risky toy (mean difference score = 1.21 s, evidence for sustained aversion). However, the few toddlers who made the risky choice first also spent relatively longer engaging with the risky toy than with the safe toy (mean difference score = −2.75 s). Splitting the data by first toy choice (risky choice vs. safe choice), we then used one-sample \( t \) tests (difference from zero) to assess whether there was a bias toward engaging longer with either toy across the entire testing session. The one-sample \( t \) tests revealed that although there was a general tendency for toddlers to play longer with their first toy choice, neither those initially choosing the safe toy, \( t(24) = 0.94, p = .355, CI [−1.43, 3.85] \), nor those initially choosing the risky toy, \( t(10) = −1.82, p = .10, CI [−6.12, 0.62] \), showed a significant bias toward playing longer with either toy (evidence for transitory aversion to the risky toy). In other words, regardless of the first toy chosen, most toddlers ended up playing with both toys during the 20-s test period.

We also tested whether there was a general within-participants difference in duration of toddler engagement with the safe and risky toys across the entire 20-s test period across all toddlers and as a function of the train–test interval. We used repeated-measures ANOVA, controlling for latency to make a choice and sex. There was no significant within-participants effect of engagement duration for the safe and risky toys, \( F(1, 32) = 1.08, p = .307, \eta^2_p = .03 \), suggesting that play times for both toys were comparable across the 20-s test phase. There was also no between-participants effect of latency to make a choice, \( F(3, 32) = 2.43, p = .129, \eta^2_p = .07 \), train–test interval, \( F(1, 32) = 0.04, p = .841, \eta^2_p = .00 \), or sex, \( F(1, 32) = 0.02, p = .895, \eta^2_p = .00 \), on duration of engagement with the safe and risky toys. Moreover, all the interactions between duration of engagement and train–test interval, \( F(1, 32) = 0.94, p = .339, \eta^2_p = .03 \), between duration of engagement and sex, \( F(1, 32) = 0.58, p = .452, \eta^2_p = .02 \), and between duration of engagement and latency to make a choice, \( F(1, 32) = 0.96, p = .336, \eta^2_p = .03 \), were not significant.

Social monitoring during toy engagement

We assessed differences in social monitoring (i.e., total looking time) toward the adult actors while toddlers engaged with the risky toy and safe toy. Three toddlers who did not engage with any toy across the test period were excluded from this analysis, leaving \( N = 36 \). First, we used a repeated-measures ANOVA with the within–participants variables of (a) actor (looking to experimenter vs. emitter) and (b) toy engagement (engagement with safe or risky toy) and the dependent measure of total looking time. There were no effects of actor, \( F(1, 35) = 1.18, p = .285, \eta^2_p = .03 \), or toy engagement, \( F(1, 35) = 0.37, p = .547, \eta^2_p = .01 \), and no significant interaction between actor and toy engagement, \( F(1, 35) = 2.63, p = .114, \eta^2_p = .07 \).

Next, we examined social monitoring (i.e., total looking time to actors) during toy engagement as a function of train–test interval. Because the previous analysis showed no effect of actor on total looking
time, here we averaged social monitoring across actors (emoter and experimenter) during engagement with each toy, producing two social monitoring values for each toddler (one during engagement with the risky toy and one during engagement with the safe toy). A repeated-measures ANOVA was used with the within-participants variable of toy (risky vs. safe) and the between-participants variable of test–train interval (immediate vs. delay), controlling for sex, and the dependent measure of social monitoring (total looking time). There was no main effect of toy, $F(1, 33) = 0.20, p = .655, \eta^2_p = .01$, train–test interval, $F(1, 33) = 0.49, p = .491, \eta^2_p = .01$, or sex, $F(1, 33) = 0.32, p = .576, \eta^2_p = .01$, on total looking time. There were also no interactions between toy and train–test Interval, $F(1, 33) = 0.36, p = .550, \eta^2_p = .01$, or between toy and sex, $F(1, 33) = 0.02, p = .897, \eta^2_p = .001$, on total looking time.

Finally, we assessed whether toddlers' initial toy choice predicted social monitoring in a manner concordant with their toy choice. That is, would toddlers who initially selected the safe toy exhibit more social monitoring while engaged with the safe toy (and vice versa for toddlers who initially selected the risky toy)? We again collapsed social monitoring values across actor because the initial analysis revealed no effect of actor on social monitoring values. Because more toddlers chose the safe toy than the risky toy, we performed a nonparametric Mann–Whitney $U$ test to assess whether the dependent measure of difference in social monitoring duration (difference score of monitoring during safe vs. risky toy engagement) depended on initial toy choice (independent variable). Because there were no effects of train–test interval and sex in social monitoring reported above, the difference score was created with the raw social monitoring values rather than residualized values from the repeated-measures ANOVA. There was a significant effect of initial toy choice on the difference in social monitoring during safe toy play relative to risky toy play, $U = 80, p = .049, d = 0.70$, whereby toddlers who initially chose the safe toy engaged in more social monitoring during their play with the safe toy and toddlers who initially chose the risky toy engaged in more social monitoring during their play with the risky toy (see Fig. 4). However, post hoc one-sample $t$ tests showed that these biases did not significantly differ from zero (for toddlers initially choosing the safe toy, $t(24) = 1.74, p = .096, CI [−0.09, 0.99]$; for toddlers initially choosing the risky toy, $t(10) = −1.49, p = .167, CI [−1.44, 0.28]$).

**Discussion**

The data from Experiment 1 support our hypothesis that toddlers can learn via emotional eavesdropping to discriminate between two stimuli in a cue-specific manner. That is, toddlers can discrim-
inate those stimuli that instigated a negative interaction between two adults from those that did not. More toddlers picked the safe toy than the risky toy as their first choice, indicating that toddlers were not merely associating negativity with the emoter but rather were using the emoter’s affect to discriminate between the two objects. Importantly, because the train–test interval was not associated with toddlers’ choice, the data do not support the view that cue-specific observationally learned memory is forgotten in toddlers after 2 weeks. Of the few toddlers who actually did choose the risky toy first, they did not differ in play initiation time. When considering the entire 20-s test period, toddlers did not play longer with the safe toy than with risky toy, and most toddlers ended up exploring both toy options, although there was a tendency to play longer with whichever toy they selected first and for more social monitoring during their play with that toy. However, toddlers did not exhibit more social monitoring during their play with the risky toy overall, contrary to our hypothesis.

Toddlers in Experiment 1 clearly differentiated between the safe and risky toys (evidenced through their initial choice of toy to play with). However, it was also clear that although there was a tendency for toddlers to play more overall with their initially chosen toy, they did eventually venture toward and explore the other toy during the 20-s test period. We speculated that for the majority of toddlers, they may have been initially cautious, choosing the safe option for themselves while monitoring the adults. Then, after they were sure that the emoter was not interested in their play, they may have moved to explore the risky toy. If true, we reasoned that toddlers may be more willing to choose the risky toy for a surrogate actor than for themselves (i.e., they could use the surrogate actor to obtain more information about the actions of the emoter) and would then engage in more social monitoring of the actors. We tested those hypotheses in Experiment 2.

**Experiment 2**

**Method**

**Participants**

Our initial sample consisted of \( N = 69 \) toddlers who were enrolled at the university’s preschool. There were 5 exclusions from that initial sample (refusal at test: \( n = 3 \) girls, 25–34 months old, delay condition; experimenter error: \( n = 1 \) girl, 34 months old; noncompliance: \( n = 1 \) girl, 34 months old). Hence, our final sample consisted of \( N = 64 \) toddlers aged 22–40 months (mean age = 30 months; 48% boys and 52% girls; see Table 1 for sample demographics broken down by train–test interval). The mean age of this sample (30 months) was slightly younger than that in Experiment 1 (33 months), \( t(101) = 3.4, p = .001 \) (see Appendix in supplementary material for analyses testing effects of age on outcome variables, all of which were not significant). Because there was no prior examination of emotional eavesdropping on toddlers’ decision making for social surrogates, we had no basis to define a sample size for Experiment 2. Thus, we decided to enroll all children (for whom parents provided informed consent) from fall 2015 to fall 2017, which we expected to be approximately double the sample size of that achieved in Experiment 1. The university’s institutional review board approved the protocol.

**Design**

The same mixed within/between-participants design used in Experiment 1 was used in Experiment 2. Toddlers in the delay group were tested 10–22 days after training (mode = 14 days). Toddlers were semirandomly assigned to the immediate and delay intervals; because of scheduling difficulties (an upcoming holiday), many more toddlers than usual needed to go through the training phase of the study on the same day, leaving no time for immediate testing on that day (immediate testing effectively doubled the length of the session). Thus, those toddlers went through the training phase and then were placed in the delay group and were tested after the holiday. This resulted in a smaller sample size for the immediate interval (\( n = 17 \)) than for the delay interval (\( n = 47 \)).

**Materials**

The same materials used in Experiment 1 were also used in Experiment 2.
Procedure

The same procedure used in Experiment 1 was also used in Experiment 2 except for one major change to the testing phase: Toddlers made a choice for a social surrogate (the experimenter) rather than for themselves. During the testing phase, after the emoter expressed annoyed affect to the second toy, toddlers were tested while the emoter remained in the room. In the immediate condition, the experimenter placed both target toys on her own lap, in front of the toddler but out of the toddler's reach. The experimenter then told the toddler that they would play one more time and asked the toddler to choose a toy for the experimenter to play with (surrogate choice; see Fig. 5B). The toddler was given approximately 20 s to make his or her choice (either verbally or by pointing). Once the toddler had made the choice, the experimenter demonstrated that toy again, making the noise action another two times, before then performing the same demonstrations on the second toy (the toy the toddler did not choose). The emoter maintained neutral affect during the testing phase.

Fig. 5. (A) Estimated marginal means (and standard errors) of toddlers’ positive affect during each training phase in Experiment 2. The training phase is written on the upper side of the x-axis. The asterisk (*) indicates a significant difference between groups. Toddlers’ positive affect was higher during the target toy demonstration phase than during the emoter display of annoyed affect phase. Gray shading depicts the critical phase of training when the emoter displays annoyed affect. (B) The emoter’s display of neutral and annoyed affect balloons out to visual depictions of the room setup during those phases. Panel B also shows the setup of the testing phase, which is separated from the training phase by an arrow indicating an immediate or delayed test. The toddler sat opposite the experimenter and emoter with a table (gray box) between them. The monkey and fish indicate the toys that were used as the conditioned stimuli. The teacher sat behind the toddler. The words spoken by the bolded character are written underneath each box in Panel B. The words for the critical experimental manipulation (emoter’s display of annoyed affect) are also bolded for emphasis.
In the delay condition, after the emoter had expressed annoyed affect to the second toy, the experimenter remarked “Oh, I thought it [the risky toy] was interesting” and told the toddler “I will bring these toys back another day for you to play with.” Approximately 2 weeks later, the experimenter returned and the warm-up and target toy demonstration phases were demonstrated again, with the change that the experimenter inserted “remember” language into the demonstration (e.g., “Remember last time I showed you these toys . . .”). After the toy demonstration, the emoter entered the room and the experimenter told the toddler that she would play one more time and asked the toddler to choose a toy for her to play with (surrogate choice; the procedure for the immediate test was then followed). See Appendix in supplementary material for manipulation checks of the emotional expressions of the experimenter and emoter during each phase of the study.

Scoring of toddlers’ behavior

As in Experiment 1, toddlers’ overt emotional expressions across the course of the training phase of the study (target toy demonstration and neutral and negative emotional exchanges) were video-recorded and coded. Learning was assessed through two behavioral indices that were measured during the training phase: (a) toddlers’ choice of toy for the surrogate to play with and (b) their latency to make the choice for the surrogate. In addition, we measured toddlers’ social monitoring of adult actors during the surrogate play phase. Trained observers who were blind to the aims, design, and group assignment (immediate vs. delay) of toddlers scored adults’ and toddlers’ emotions and toddlers’ behaviors (see Supplemental Fig. 1b for a graphical depiction of the coding schedule in Experiment 2).

Choice of toy for surrogate. Toddlers needed to indicate their choice for the surrogate verbally (e.g., saying “fish” or “monkey”), by touching a toy, or by reaching for/pointing to a toy. Toddlers’ behavior for the surrogate choice variable therefore was coded as “safe choice,” “risky choice,” or “no choice.”

Latency to make a choice for the surrogate. We calculated the length of time in seconds it took toddlers to make that choice. Latencies were positively skewed and were log(base 10) transformed. Toddlers who were coded as “no choice” did not receive a latency score.

Social monitoring tied to surrogate’s toy engagement. Using the camera that was focused on toddlers’ face, we coded bouts of fixation on the experimenter and emoter during the surrogate’s play with each of the toys (safe and risky). Toddlers who were coded as “no choice” did not receive a social monitoring score.

Statistics and analysis

Data were analyzed with parametric tests unless there was uneven distribution between the groups. For parametric tests, in conditions where the assumption of sphericity was violated, we report the Greenhouse–Geiser adjusted p values and the nominal degrees of freedom. Because toddlers were within a close age range and were from the same class at the preschool, and because we did not expect to have the power to test differences in behavior across months of age, we did not consider toddler age in our analyses. However, see Appendix in supplementary material for primary analyses stratified by age. Sex of toddlers was entered as a covariate in all analyses.

Results of main analyses

Toddlers’ emotional expression across each of the four periods of the training phase

A repeated-measures ANOVA was performed to examine the change in toddlers’ positive affect across each of the four training phases (within-participants variable): (a) experimenter demonstrates target toys, (b) emoter enters the room, (c) emoter displays neutral affect, and (d) emoter displays annoyed affect (see Fig. 5A). The train–test interval (immediate vs. delay) was a between-participants variable, and sex was a covariate. There were n = 3 participants whose facial affect could not be coded due to their face repeatedly leaving the camera frame, leaving a final N = 61 for this
analysis. There was a significant within-participants effect of training phase, $F(3, 174) = 22.48, p < .0001, \eta^2_p = .28$; however, there was no interaction between training phase and train–test interval, $F(3, 174) = 2.30, p = .069, \eta^2_p = .04$, or between training phase and sex, $F(3, 174) = 0.03, p = .98, \eta^2_p = .00$. There was a between-participants effect of train–test interval, $F(1, 58) = 29.92, p < .0001, \eta^2_p = .34$, whereby toddlers tested at the delay had lower overall positive affect than toddlers tested immediately after training. The between-participants effect of sex was not significant, $F(1, 58) = 0.17, p = .686, \eta^2_p = .00$. Because of the significant within-participants effect of training phase on toddlers’ positive emotion expression, we performed pairwise comparisons of the estimated marginal means between the training phase where the emoter expressed annoyed affect and all other training phases (with Bonferroni adjustment for multiple comparisons, three contrasts: $\alpha = .017$). Toddlers’ positive affect when the emoter displayed annoyed affect was lower than when the experimenter was demonstrating the target toys, $p < .0001, CI [−1.14, −0.51]$, but was not different from when the emoter entered the room, $p = .058, CI [−0.01, 0.46]$, or when the emoter was expressing neutral affect, $p = 1, CI [−0.25, 0.23]$. In other words, like in Experiment 1, positive affect dropped when the emoter first entered the room and did not recover throughout the rest of the training period.

### Choice for the surrogate

We performed a multinomial logistic regression to examine whether there was an overall bias in choice for the surrogate toward the safe choice, risky choice, or no choice. The model was significant, suggesting that choices for the surrogate were not evenly distributed, $\chi^2(2) = 17.58, p = .0002$ (see Fig. 6A). Post hoc tests revealed that there was no difference in the number of toddlers who chose the safe toy and the risky toy for the surrogate, $\beta = 0.12, Wald \chi^2(1) = 0.16, p = .691$. However, toddlers chose either the safe toy or the risky toy more often than they made no choice for the surrogate, $\beta = 1.35, Wald \chi^2(1) = 10.13, p = .001$, and $\beta = 1.46, Wald \chi^2(1) = 12.02, p = .001$, respectively. We also tested whether there was a difference in toddlers’ choice as a function of train–test interval or sex (although a post hoc power analysis revealed that we were adequately powered at 80% to observe only a medium to large effect size for these analyses). Toddlers’ choice of toy for the surrogate did not differ between the immediate and delay groups, $\chi^2(2) = 0.73, p = .695$, Cramer’s $V = .11$ (but see Fig. 6B for choice behavior broken down by the immediate and delay groups for illustration), nor did it differ between boys and girls, $\chi^2(2) = 2.69, p = .261$, Cramer’s $V = .21$.

Our hypothesis that toddlers would choose the risky toy for the surrogate more frequently than the safe toy was not supported in Experiment 2. Thus, we performed exploratory analyses to examine individual difference variables that might help to explain toddlers’ choice for the surrogate (i.e., to

---

**Fig. 6.** (A) Percent of toddlers in Experiment 2 making each of the choices listed on the $x$ axis (safe choice, risky choice, and no choice) for the surrogate. The asterisk (*) indicates a significant difference between groups. Toddlers made safe and risky choices equally frequently, and they made each of those choices more frequently than no choice, n.s, not significant. (B) Percent of toddlers in Experiment 2 making each of the choices listed on the $x$ axis for the surrogate, broken down by immediate and delay conditions. Although there was no significant effect of train–test interval on toddlers’ choices, the data are shown broken down by train–test interval to illustrate the consistency in responding across time.
assess whether there was any evidence that toddlers were not choosing randomly). We explored whether toddlers’ level of positive affect during the critical experimental manipulation (emoter display of annoyance) differed between toddlers who went on to choose the risky or safe toy for the surrogate. We speculated that toddlers who went on to choose the safe toy might not have been as affected by the emotional eavesdropping procedure and therefore might not have felt the need to use the surrogate to test whether the emoter would display negative affect. Consistent with that post hoc hypothesis, toddlers’ positive affect during the emoter’s display of annoyed affect was lower in toddlers who went on to choose the risky toy than in toddlers who went on to choose the safe toy, \( t(52) = 3.45, p = .001, d = 0.94 \) (see Supplemental Fig. 2 in supplementary material).

Latency to make a choice for the surrogate

The 7 toddlers who did not make a choice for the surrogate were excluded from this analysis (final \( N = 57 \)). Because the distributions of toddlers choosing the risky and safe toys for the surrogate were not different, we performed an ANOVA to test whether the mean latency to make a choice for the surrogate differed as a function of the choice made (risky vs. safe) or the train–test interval (immediate vs. delay), controlling for sex. There was no main effect of choice for the surrogate, \( F(1, 52) = 0.38, p = .539, \eta^2_p = .01 \), train–test interval, \( F(1, 52) = 1.91, p = .173, \eta^2_p = .04 \), or sex, \( F(1, 52) = 0.27, p = .605, \eta^2_p = .01 \), on latency to make a choice for the surrogate, and the interaction between choice for surrogate and train–test interval was also not significant, \( F(1, 52) = 0.21, p = .652, \eta^2_p = .00 \).

Social monitoring tied to surrogate’s toy interaction

Of the \( N = 57 \) toddlers who made a toy choice for the surrogate, \( n = 2 \) had their eyes outside of the camera frame and could not be coded for social monitoring, resulting in a final \( N = 55 \) for this analysis. We examined whether there were any within-participants differences in toddlers’ social monitoring (total looking time to each actor) while the surrogate was interacting with the safe and risky toys. We performed a repeated-measures ANOVA, with social monitoring to each actor (surrogate vs. emoter) during surrogate’s toy interaction (safe vs. risky) as the repeated measures and toddlers’ choice for the surrogate and the train–test interval as the independent variables, controlling for sex. There was a within-participants main effect of the actor, \( F(1, 50) = 12.44, p = .001, \eta^2_p = .20 \), indicating that toddlers looked more at the surrogate than at the emoter. There was also a main effect of surrogate interaction with the safe and risky toys, \( F(1, 50) = 4.06, p = .049, \eta^2_p = .08 \), indicating that toddlers looked more overall when the surrogate was interacting with the risky toy than when she was interacting with the safe toy. There was also a significant interaction between surrogate’s toy interaction and toddlers’ choice for surrogate, \( F(1, 50) = 4.58, p = .037, \eta^2_p = .08 \), as well as a significant three-way interaction among surrogate’s toy interaction, toddlers’ choice for surrogate, and train–test interval, \( F(1, 50) = 4.45, p = .040, \eta^2_p = .08 \). No other main effects or interactions were significant, largest \( F(1, 50) = 3.56, p = .065, \eta^2_p = .07 \).

To further understand the two interaction effects we saw, we averaged social monitoring across the actors (surrogate and emoter) for each child and ran two repeated-measures ANOVAs: one for the immediate train–test interval and one for the delay train–test interval. Social monitoring during surrogate interaction with the risky and safe toys was the repeated dependent measure, and toddlers’ choice for surrogate was the between-participants variable, controlling for sex. For the immediate group, there was a significant interaction between surrogate’s toy interaction and toddlers’ choice for surrogate on toddlers’ social monitoring behavior, \( F(1, 13) = 5.12, p = .041, \eta^2_p = .283 \). Paired-samples post hoc \( t \) tests (with Bonferroni adjustment for two comparisons: \( z = .025 \)) revealed that toddlers who chose the risky toy for the surrogate engaged in more social monitoring when the surrogate was interacting with that risky toy than when she was interacting with the safe toy, \( t(6) = 0.39, p = .008, CI = [−1.93, −0.44] \) (see Fig. 7A). In contrast, toddlers who chose the safe toy for the surrogate did not alter their looking behavior as a function of what toy the surrogate was interacting with, \( t(8) = 0.57, p = .583, CI = [−0.85, 1.41] \). The within-participants effect of surrogate’s toy interaction on toddlers’ social monitoring was not significant, \( F(1, 13) = 1.19, p = .296, \eta^2_p = .08 \), nor was the between-participants main effect of toddlers’ choice for the surrogate, \( F(1, 13) = 0.13, p = .726, \eta^2_p = .01 \), or sex, \( F(1, 13) = 0.25, p = .625, \eta^2_p = .02 \). The interaction between sex and surrogate’s toy interaction was also not significant, \( F(1, 13) = 0.06, p = .809, \eta^2_p = .01 \).
For the delay group, there was a significant between-participants effect of toddlers’ choice for the surrogate, $F(1, 36) = 8.79, p = .005, \eta_p^2 = .20$, demonstrating that toddlers who chose the risky toy for the surrogate spent more time socially monitoring overall (i.e., regardless of the toy the surrogate was interacting with) than toddlers who chose the safe toy for the surrogate (see Fig. 7B). There was also a significant main effect of sex, $F(1, 36) = 8.81, p = .005, \eta_p^2 = .20$, demonstrating that girls spent more time socially monitoring overall than boys. However, there was no effect of surrogate toy interaction, $F(1, 36) = 1.72, p = .199, \eta_p^2 = .05$, and no interactions between surrogate’s toy interaction and toddlers’ choice for surrogate, $F(1, 36) = 0.02, p = .891, \eta_p^2 = .00$, or between surrogate’s toy interaction and sex, $F(1, 36) = 2.56, p = .118, \eta_p^2 = .07$.

Discussion

The data from Experiment 2 did not support our hypothesis that toddlers would be more likely to choose the risky toy for the surrogate; instead, there was a relatively even split between toddlers choosing the risky toy and the safe toy for the surrogate. One way to interpret these findings is that toddlers’ choices for the surrogate were random. However, our exploratory analyses suggest that toddlers’ behavior in Experiment 2 may, in fact, have been motivated. Specifically, we observed two behaviors in toddlers that differed as a function of the toy they chose for the surrogate, suggesting that these choices were functional. First, we observed that toddlers who chose the risky toy for the surrogate expressed lower positive affect during the training phase where the emoter displayed annoyed affect than toddlers who went on to choose the safe toy for the surrogate. Second, toddlers who chose the risky toy for the surrogate also engaged in more social monitoring after their choice was made relative to toddlers who made the safe choice for the surrogate. It is possible that toddlers who ultimately chose the risky toy for the surrogate might have been more affected (or confused) by the observed emotional interaction. If so, those toddlers may have been motivated to discover more about the interaction, thereby choosing the risky toy for the surrogate and then monitoring the adults when the surrogate interacted with it. In other words, although there was no group-level bias for toddlers to choose the safe or risky toy for the surrogate in Experiment 2, the differences in emotional expres-

Fig. 7. Estimated marginal means (and standard errors) of toddlers’ social monitoring (averaged between the emoter and experimenter) during surrogate interaction with the safe and risky toys across the entire test phase in Experiment 2. Surrogate toy interactions are represented by different colored bars (safe: white; risky: gray). Toddlers’ choices for the surrogate are listed along the x-axis (risky on the left and safe on the right). Because there was a significant interaction with the train–test interval, the data are shown broken down by train–test interval, with the immediate group in the left panel (A) and the delay group in the right panel (B). The asterisk (*) indicates a significant difference between groups. For the immediate condition (A), toddlers who made the risky choice for the surrogate engaged in more social monitoring when the surrogate was interacting with the risky toy than when she was interacting with the safe toy. For toddlers in the delay condition (B), toddlers who made the risky choice for the surrogate engaged in more social monitoring overall than those who made the safe choice for the surrogate regardless of what toy the surrogate was interacting with.

sion during learning, and in social monitoring during surrogate’s toy interaction, suggest that toddlers were not choosing randomly. These interpretations will remain speculative until confirmed in future studies, especially considering that the unbalanced sample size of toddlers choosing safe versus risky toys for themselves prevented these same exploratory analyses in Experiment 1.

Importantly, the social monitoring of toddlers in Experiment 2 was the only outcome that differed as a function of the train–test interval. Toddlers in the immediate group who chose the risky toy for the surrogate spent more time social monitoring when the surrogate was interacting with the risky toy than when she was interacting with the safe toy. In other words, those toddlers engaged in discriminative social monitoring, suggesting that they had a specific memory of the interaction. A related interpretation of these findings is that toddlers paid attention to whatever toy the experimenter played with first and then decreased their attention if the second toy was the safe toy. For toddlers who chose the risky toy for the surrogate, because the second toy the experimenter played with was the safe toy, they would have reduced their monitoring. For toddlers who chose the safe toy for the surrogate, because the risky toy was the second toy the experimenter played with, this would result in a pattern of sustained attention across both toys. Such an interpretation would make sense of similar monitoring levels across risky toy engagement in the risky choice group and both toy engagements in the safe choice group. Regardless, both interpretations support that toddlers were likely engaging in discriminative learning. In contrast, at the delay toddlers who chose the risky toy for the surrogate did not show discriminative social monitoring, although they did engage in more social monitoring than toddlers who chose the safe toy for the surrogate. The different social monitoring behaviors at the delay for toddlers choosing the risky toy suggest that their memory of the prior adult interaction may have become less specific over time or that there was an increase in their general wariness over time. Future studies will be required to differentiate between these possibilities.

General discussion

Here we used a within-participants variation of the emotional eavesdropping task (Repacholi & Meltzoff, 2007) to assess learning and long-term retention of discriminant observational information during toddlerhood. Extending the past literature, which had shown that toddlers could learn to associate an emotional reaction to a single toy via observing an adult–adult interaction (Repacholi & Meltzoff, 2007; Repacholi et al., 2016; Skinner, Meltzoff, & Olson, 2017), in Experiment 1 we demonstrated that toddlers could also engage in more sophisticated discriminant learning through the same process. Specifically, toddlers chose to interact with safe toys over risky toys after observing an adult exhibit diverging emotional reactions (annoyed vs. neutral) to another adult in reference to two different toys. Moreover, toddlers appeared to retain a specific memory of that interaction for at least up to 2 weeks after learning (although our power to detect forgetting may have been limited). Thus, the results from Experiment 1 suggest that toddlers are adept at using indirectly observed evidence to discriminate potential risk to themselves versus safety and to alter their behavior accordingly. However, the fact that most toddlers engaged with both toys within the 20-s test window, and that differences in overall engagement between safe and risky toys were small (<2 s in most cases), suggests that toddlers were driven to explore both toy options even when initially apprehensive about the risky toy.

In Experiment 2, when the stakes were supposedly lower for toddlers (i.e., choosing for a surrogate rather than for themselves), the group bias toward choosing the safe option was no longer apparent. Instead, toddlers appeared to engage in different information-seeking strategies depending on the option they chose for the surrogate: Toddlers who chose the risky toy for the surrogate engaged in more social monitoring of the adults than toddlers who chose the safe toy for the surrogate. This suggests that, rather than choosing randomly, toddlers’ choices toward the safe and risky toys in Experiment 2 reflect differing motivations. One possibility for future observational learning studies will be to parametrically manipulate toddlers’ perceived risk to themselves and to examine whether choice behaviors are modulated in concert. Recruiting groups of toddlers who differ in their risk susceptibility/tolerance is another approach that could yield complementary information. For example, toddlers who have had prior experience with adversarial adult interactions may exhibit more inhibited play with the risky toy and may make safer choices also for a surrogate. In addition, toddlers’ level of cognitive control and risk-taking behavior in other tasks may predict the choices toddlers make in this task.
Beyond demonstrating that toddlers could learn to discriminate between two predictive cues via emotional eavesdropping, another aim of the experiments discussed in this article was to establish how long such learning persisted to influence behavior, although it must be noted that we had a limited sample size in which to detect temporal differences in behavior. As mentioned in the Introduction, toddlers who learned contingencies from emotional eavesdropping would do well to retain that information and use it to regulate their behavior across time. In Experiment 1, we observed that toddlers retained discriminative observational memory for at least 2 weeks. This extends earlier work on observational learning (that used paradigms different from emotional eavesdropping), which has shown persistence for at least 1 h in 14-month-olds (Hertenstein & Campos, 2004). Moreover, these data are consistent with studies of direct experiential learning in 18-month-olds that show memory persistence up to 13 weeks postlearning (Hartshorn et al., 1998). It will be important for future studies on emotional eavesdropping to replicate the current findings and then extend the retention interval beyond 2 weeks to map retention functions in toddlers for this form of observationally learned information.

Although memories appeared to be retained for at least 2 weeks in Experiment 1, the specificity of memories waned over that time frame in Experiment 2. Immediately after learning, toddlers who chose the risky toy for the surrogate exhibited more social monitoring when the surrogate interacted with the risky toy than when she interacted with the safe toy. In contrast, 2 weeks after learning, toddlers who chose the risky toy for the surrogate exhibited more social monitoring overall (i.e., whether she was interacting with the risky or safe toy). In other words, looking behavior differed systematically according to what toy toddlers chose for the surrogate, suggesting a specific memory. Yet, the looking behavior associated with the toy choice was specific at the immediate interval and not at the delay interval, perhaps suggesting some cue generalization at the delay. In one direct experiential learning study, it was shown that 1- and 2-year-old toddlers exhibited memory generalization between cues 1 week after learning, but this generalization did not occur at the expense of their memory for the specific original cues, with toddlers performing accurately in recognizing the original cues (Bauer & Dow, 1994).

In these two experiments, children were 2.2–3.9 years of age. As mentioned in the Introduction, observational social learning is well established by 24 months of age; thus, we did not expect toddlers' chronological age to influence learning outcomes in this study. Indeed, we did not see toddlers' age being associated with any of the behaviors assessed. However, it is important to note the potential for individual differences in behaviors reported on here that could be associated with variations in developmental milestones occurring across this stage of development. One potential milestone worth investigating in this respect is language development given that it is associated with emotion regulation capabilities in young children (Bendezú et al., 2018), and another is emotional understanding given that it has been associated with preschoolers' accuracy in recalling emotionally salient information (Channell & Barth, 2013). Future studies should measure such milestones when investigating observational social learning during this age period.

A final point to note in these experiments is that toddlers were learning about social risk from two relatively unknown adults (there was some familiarity with the experimenter, and the emoter was novel). This is similar to the original sequential version of the paradigm where toddlers observe an interaction between two unknown adults (Repacholi & Meltzoff, 2007). Other studies have shown that toddlers (18–24 months old) learn equally well when a same-age peer versus an adult is the target of the emoter's outburst (Seehagen, Schneider, Miebach, Frigge, & Zmyj, 2017). As such, it would appear that this form of observational learning is robust to actor identity and age. This has important implications for who toddlers may look to as models for appropriate behavior in the world, suggesting that information may be learned from a range of known and unknown individuals. Nevertheless, it would be interesting to compare how observational learning may change in the emotional eavesdropping paradigm for a range of different social partners with high ecological validity for toddlers (e.g., between two parents, between a parent and a sibling).

**Limitations**

Although there are many interesting outcomes reported in this article, several limitations should be acknowledged. First, to ensure that there was no emotional carryover between the risky and safe toys, we fixed the order of emoter-expressed affect (neutral first and annoyed second). This design choice
leaves open the possibility that toddlers simply chose the most recent toy, resulting in the biased behavior seen in Experiment 1. However, the fact that a group bias toward the safe toy was not observed in Experiment 2 (where the same fixed order of expressed affect occurred) argues against that interpretation. Another limitation in Experiment 1 was the small sample of toddlers choosing the risky toy for themselves, precluding exploratory examination of behaviors associated with that choice, which might have been informative for understanding the motivations underlying the choice. We also had an unexpectedly lower sample size for the immediate group in Experiment 2, which occurred due to scheduling complications. Yet, the fact that we observed nearly identical distributions of choices for the immediate and delay toddlers in Experiment 2 suggests that sample size was not biasing the results. In addition, the mean age of toddlers was younger in Experiment 2 \((M = 29\) months) than in Experiment 1 \((M = 33\) months). However, because age was not associated with choosing the safe or risky toy across the two experiments, we do not think that age differences between the experiments are critical to their interpretation. Across both experiments, we also had a relatively low sample size within which to detect temporal performance differences. Whereas in Experiment 2 we did see an effect of train–test interval on performance, in Experiment 1 we did not, which might suggest either that memories were retained over 2 weeks or that we simply could not detect a difference if one existed. Future studies should continue to look at the retention of such observationally learned associations across time. Another design choice that acts as a limitation in this study is that we examined only annoyed versus neutral emoter-expressed affect rather than examining how toddlers respond to a range of different emotions (e.g., sadness, happiness, fear, anger). Although it would have been interesting to include a positive interaction, annoyance versus neutral provided a tractable number of conditions to test within this study.

Conclusions

The findings reported here demonstrate that toddlers are skilled at using observational discriminative learning to regulate their own behavior and can retain that information over 2 weeks. We also found suggestive (but not conclusive) support for the idea that toddlers may manipulate the behavior of others to gain more information about unexpected social events. Whether the differential use of observationally learned information on the “traditional” variant (choice for self) and “low-stakes” variant (choice for surrogate) of the emotional eavesdropping procedure is predictive of later behavior, particularly anxiety, is an open question for future research. Moreover, exactly how long such observationally learned information lasts (beyond 2 weeks after learning) to influence toddler behavior needs to be examined.

Acknowledgments

This research was supported by the National Institute of Mental Health (R01MH091864 [to NT] and 1K99MH113821 [to BLC]), the Dana Foundation (to NT), an American Australian Association Fellowship (to BLC), and the National Health and Medical Research Center (to BLC). The authors acknowledge the assistance of the teachers and staff at the university’s daycare.

Author contributions

N. Tottenham, T. Klein, and B. Callaghan developed the study concept, designed the study, and interpreted the data. B. Callaghan, T. Choy, K. O’Sullivan, E. Routhier, N. Cabrera, and V. Goode collected the data. B. Callaghan developed the coding rubric. E. Routhier blind coded the data. B. Callaghan analyzed the data and prepared the manuscript. N. Tottenham and T. Klein provided critical revisions. All authors read and approved the manuscript before submission.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jecp.2022.105391.
References