Surprise! Predicting Infant Visual Attention in a Socially Assistive Robot Contingent Learning Paradigm

Lauren Klein[1], Laurent Itti[2], Beth A. Smith[3], Marcelo Rosales[3], Stefanos Nikolaidis[4], Maja J. Matarić[1]

Abstract— **Early intervention to address developmental disability in infants has the potential to promote improved outcomes in neurodevelopmental structure and function [1]. Researchers are starting to explore Socially Assistive Robotics (SAR) as a tool for delivering early interventions that are synergistic with and enhance human-administered therapy. For SAR to be effective, the robot must be able to consistently attract the attention of the infant in order to engage the infant in a desired activity. This work presents the analysis of eye gaze tracking data from five 6-8 month old infants interacting with a Nao robot that kicked its leg as a contingent reward for infant leg movement. We evaluate a Bayesian model of low-level surprise on video data from the infants' head-mounted camera and on the timing of robot behaviors as a predictor of infant visual attention. The results demonstrate that over 67% of infant gaze locations were in areas the model evaluated to be more surprising than average. We also present an initial exploration using surprise to predict the extent to which the robot attracts infant visual attention during specific intervals in the study. This work is the first to validate the surprise model on infants; our results indicate the potential for using surprise to inform robot behaviors that attract infant attention during SAR interactions.**

Fig. 1. Experiment setup of the SAR leg movement study; the infant is seated across from the Nao robot and is wearing leg and arm motion trackers and an eye tracker. This work uses data from the study to explore Bayesian surprise as a potential predictor of infant visual attention.

I. INTRODUCTION

Exploratory movement and motor babbling are believed to provide necessary practice for infants to learn to control their bodies, and are essential for both cognitive and motor development. Some infants, including infants at risk for developmental disabilities, have greater difficulty with producing or adjusting movements as compared to infants with typical development. A lack of appropriate motor exploration and practice can contribute to impairments in strength, proprioception, and coordination. Researchers estimate that about 9% of infants in the United States are eligible for early intervention [2]. While intensive early intervention has the potential to be more effective at promoting positive neurodevelopmental outcomes than less intense, later intervention, the current standard of care is the latter [2][1].

*This work was supported by the National Science Foundation under grants NSF CBET-1706964 and NSF CCF-1317433

[1]Lauren Klein and Maja Matarić are with the Interaction Lab, Department of Computer Science, University of Southern California, Los Angeles, CA 90089, USA {kleinl,mataric}@usc.edu

[2]Laurent Itti is with the iLab, Department of Computer Science, University of Southern California itti@usc.edu

[3]Beth Smith and Marcelo Rosales are with the Infant Neuromotor Control Laboratory, Division of Biokinesiology and Physical Therapy, University of Southern California beth.smith@pt.usc.edu,mrrosale@usc.edu

[4]Stefanos Nikolaidis is with the Interactive and Collaborative Autonomous Robotic Systems Lab, Department of Computer Science, University of Southern California nikolaid@usc.edu

Our recent work introduced Socially Assistive Robotics (SAR) as a potential tool for providing accessible interventions to promote learning of motor skills for infants [3]. Research into SAR for infants aims to provide support that is complementary to human-administered therapy. Robots can provide contingent rewards to infants to encourage specific motor movements or behaviors. Other work in SAR for infants has explored adapting the difficulty level of the activity to the infant [4] and demonstrating sign language to deaf infants [5]. A key component of these activities has been the use of visual stimuli presented via robot behaviors.

As SAR interactions for infants rely on visual stimuli to teach and reinforce behaviors, robots must be able to reliably attract infants' visual attention. Therefore, robots need a model of infant attention in order to select optimal actions to perform. According to an accepted mental model presented by Cohen [6], infants fixate longer on stimuli they do not understand, or that take longer to fit to their own mental model. Therefore, it is possible that surprising stimuli may be more difficult for infants to model and could support increased visual attention.

In this work, we explore whether surprise can be used to predict infant visual attention. Specifically, we use a Bayesian model by Itti and Baldi [7] to model the surprise induced by the robot's kicking behaviors as well as the surrounding environment and its effect on the infant's gaze. The model was originally tested and validated with adults watching videos. In this work, our goals were: 1) to determine the extent to which that model can be used to predict infant gaze behaviors; and 2) to identify areas of improvement for generalizing that model to infants during SAR interactions, in order to inform future work employing surprising or novel robotic stimuli to encourage infant attention to the robot. Toward this goal, we used recorded video data from a head-mounted camera and gaze tracker on the infant as well as robot actions during a SAR leg movement study described by Fitter et al. [8]. The setup of this study is shown in Fig. 1. Analysis of the data demonstrates that over 67% of infant gaze locations were in areas that the model evaluated as having higher than average surprise values. Additionally, the surprise induced by the robot's kicking behaviors was predictive of the gaze behaviors of 2 out of 5 infants. These results indicate the potential for using surprise to inform robot behaviors that attract infant attention, but also suggest that future models of surprise may need to be personalized to individual infants. This work is part of a larger effort to inform the design of robot behaviors for future SAR interventions for infants.

II. Related Work

Past research has identified multiple modes through which robots can acquire human attention. Robot eye gaze has been successful in acquiring visual attention across various settings. Johnson et al. showed that robot eye gaze can direct user attention toward specific material during a tutoring interaction [9], and Admoni et al. showed a similar effect for directing attention toward a specific area during robot-human handoffs [10]. In addition, Ito et al. demonstrated that people pay more attention to a robot if it displays mutual gaze [11].

Displaying unexpected behaviors has also been useful in attracting attention. Yu et al. found that participants displayed longer fixations on a robot during a teaching activity if the robot generated random movements than if it continuously followed the human's direction of gaze, as the human participant would try to get the robot's attention [12]. Admoni et al. showed that if the robot deliberately paused before letting go of an object during robot-human handoffs, the human would look to the head of the robot for direction [10]. This effect of unexpected, or perhaps surprising behaviors supports our investigation of surprise as a predictor of visual attention.

Finally, time may influence the effect of surprise on attention given to a robot. Bruce et al. found that time of day was one of two tested factors determining whether a human would stop to talk to a robot [13]. In addition, Yu et al. determined that patterns of human eye gaze shifts are time-sensitive with respect to robot movements [12]. This would support exploring surprise as a predictor of gaze, as surprise depends on both spatial and temporal information.

Much of the past work in attention, especially robot-generated gaze behaviors, has involved higher level or top-down stimuli for acquiring attention. In contrast, the work we present focuses specifically on low-level, bottom-up stimuli. There also exists work that explores demonstrating robot attention to the human, such as in Bruce et al.'s work [13]; however, our work focuses on human attention on the robot.

Research has shown that infants are visually interested in items and events which are salient [14]. This effect extends to the SAR setting: in initial testing of infant behavior reactions to the Nao robot's kicking, light, and sound behaviors, prior work in our group confirmed that infants were more likely to look at the robot when it was moving [3], and a spinning light was successfully used to attract infant gaze in order to calibrate the gaze tracker during data collection. However, these findings did not take into account the timing of kicks within a given phase of the data collection to explain why some kicks were looked at while others were not, though a number of kicks were ignored by the infants. It is possible that during SAR activities for infants, robot action selection may require an understanding of both spatial and temporal factors influencing infant visual attention.

The currently favored model of infant attention was introduced by Cohen and includes two phases: attention getting and attention holding [6]. Infants fixate on a stimulus during the attention holding phase. Research suggests that infants will fixate on a stimulus until they form a mental understanding that matches the stimulus. In addition, infants may direct their attention more quickly to stimuli they previously found interesting. This suggests that more complex or surprising stimuli may be more successful at acquiring and holding infant visual attention. To evaluate whether this applies in a SAR interaction with infants, we investigated a Bayesian surprise model as a predictor of infant visual attention.

III. Bayesian Surprise Model

The Bayesian surprise model, described by Itti and Baldi [7][15], provides a method for computing the amount of low-level surprise generated by incoming data over both space and time. We present here a summary of the previously developed model. The model computes probability P(M) representing the extent to which an observer believes in a given hypothesis or model, M, in a model space $\mathcal{M}$. As new data observation D is introduced, the belief in model M changes to P(M|D).

Surprise is defined as the distance between posterior distribution and prior distribution of beliefs over models. This distance, and therefore the amount of surprise, is calculated using the Kullback-Leibler (KL) divergence:

$$S(D,\mathcal{M}) = KL(P(M|D), P(M)) = \int_{\mathcal{M}} P(M|D) log \frac{P(M|D)}{P(M)} dM \quad (1)$$

Incoming data are modeled using Poisson distributions M(λ), as these model the firing patterns of neurons in the brain with firing rate λ. In order to keep P(M) and P(M|D) in the same functional form for a Poisson-distributed D, P(M) is calculated using the Gamma probability density:

$$P(M(\lambda)) = \gamma(\lambda;\alpha,\beta) = \frac{\beta^{\alpha}\lambda^{\alpha-1}e^{-\beta\lambda}}{\Gamma(\alpha)} \quad (2)$$

with shape $\alpha > 0$, inverse scale $\beta > 0$, and Euler Gamma function Γ. To calculate the posterior Gamma density $\gamma(\lambda, \alpha', \beta')$, the shape and inverse scale are updated as:

$$\alpha' = \zeta\alpha + \bar{\lambda} \quad (3)$$

$$\beta' = \zeta\beta + 1 \quad (4)$$

where ζ, the "forgetting factor", limits the extent of the belief in the prior by preserving its mean α/β but increasing its variance α/β^2. This defines the time scale of the model. Itti and Baldi used a ζ value of 0.7 to evaluate the surprise of video data during user studies [7]. Source code to evaluate the Bayesian surprise model on video data and on 1-dimensional signals can be found at `http://ilab.usc.edu/toolkit/`.

The Bayesian surprise model was tested in a user study with adults aged 23-32, with normal vision [15]. Each participant watched 25 minutes of video footage. Eye movement traces, or saccades, were recorded with a gaze tracker and analyzed. The values of pixel patches in each video frame were passed into the surprise model to calculate a matrix of surprise values for each frame.

Measuring the distance between histograms of the participants' actual saccade endpoints and histograms of randomly generated saccade endpoints produced a KL divergence of approximately 0.241 [15]. The distribution of human saccade endpoints was shifted toward more surprising values than the random distribution, indicating that adults gazed toward locations that were more surprising than randomly selected locations. Further analysis showed that over 72% of saccades were targeted toward areas of the video that were more surprising than average, suggesting that adults are attracted to surprising locations of video footage [15]. The success of this model in predicting adult gaze location for video footage motivated our exploration of surprise as a predictor of infant visual attention during a SAR interaction.

IV. User Study: Infant Leg Motion Training

The SAR interaction discussed in this paper used contingent kicking motion from the Nao humanoid robot to encourage leg movement from two male and three female 6-8 month old typically developing infants [8]. Each participant was seated in a chair across from the robot, as shown in Fig. 1. In front of the infant and the robot, a pink toy ball with a bell was suspended at a height that was reachable by kicking. A parent was seated next to the infant at all times. Each infant wore a head-mounted eye tracker and inertial sensors within bands on their wrists and ankles. [8]. Prior work found that wearing those sensors has a negligible effect on infant leg movement frequency and that the sensors provide accurate infant movement data [16]. At the beginning and end of the activity, the robot sat motionless for 2 minutes to assess the baseline movement level of the infant.

After the initial baseline phase, the infant entered an 8 minute contingency phase. During this phase, each acceleration above 3 m/s^2 of the infant's leg was immediately followed by the robot kicking behavior. Three types of behaviors were introduced, in three separate stages:

1) Robot Kicking: The robot kicked its pink ball.
2) Robot Kicking and Lights: The robot kicked the ball, and the LED lights on the robot flashed in various colors.
3) Robot Kicking and Sound: The robot kicked the ball and emitted a pre-recorded infant babbling noise.

Each of the stages lasted 2.5 minutes. The behaviors were counterbalanced and randomly assigned to each infant to mitigate against ordering effects. This study procedure was approved by the University of Southern California Institutional Review Board under protocol #HS-14-00911.

The results from this study showed that 9 of 12 of the infants learned the contingency. The researchers defined learning the contingency as demonstrating threshold leg accelerations during the contingency phase at 1.5 times their baseline frequency. In addition, 9 of 12 infants were classified as imitating the robot during some phase of the study. These results highlight the success of visual stimuli in the form of robot behavior in motivating infant movement.

During this interaction, timing and distraction played important roles in infant behavior. The researchers noted that the onset of imitation occurred most often in the later two stages of the contingency phase. As each stage was only 2.5 minutes, it was essential for the robot to reliably acquire visual attention from the infant, if it were to demonstrate new or more fine-tuned movements for the infant to imitate. In addition, while the infants were often engaged, they were sometimes distracted by a researcher in the room or by the red ball. While such behavior is normal, robots must be able to overcome such distractions while demonstrating a new skill. These findings further motivate our exploration of surprise as a potential predictor of infant visual attention.

V. Model Validation

A. Methodology

Based on work by Itti and Baldi [15] that involved adults, we analyzed video from the head-mounted camera in the SAR leg movement study to validate the surprise model with infants. While previous work analyzed adult gaze toward prerecorded video, this work examines infant gaze toward real physical stimuli in the infant's environment. Video data from the infant's head-mounted camera were used to determine the surprise values of different areas of the infant's point of view over time. The same visual features from Itti's and Baldi's work [15]–color, intensity, movement, temporal onset/offset, and orientation–were used to determine these surprise values. Eye tracking software provided the coordinates of the infant's gaze within the video. Fig. 2 shows a video frame of the environment from the infant's head-mounted camera viewpoint and the infant's gaze location, as well as the corresponding surprise values of that frame.

Each infant's gaze locations were compared against randomly generated gaze locations to determine the extent to which infants looked toward more surprising locations. Gaze tracking data were available for 5 infants from the SAR leg movement study and formed the basis for our analysis. Infants were excluded if they shifted their gaze trackers during the study, if their eye was not visible by the camera, or if a technical issue occurred that prevented calibrated gaze tracking. A total of 162,113 frames and gaze locations were analyzed. For each video frame, we extracted the surprise value of the infant's gaze location and the surprise value of a randomly selected gaze location. Histograms of the surprise values at infant gaze locations and at random locations were compared using KL divergence.

Data were analyzed for the duration of the time that the infants wore the gaze tracker; as this part of the analysis was not dependent solely on the robot behavior but rather on the infants' general environment, data generated during the minutes before and after the interaction were analyzed as well as the interaction itself. Some infants wore the gaze tracker longer than others, and therefore contributed more gaze locations to our analysis. We account for the difference in total number of video frames for each infant when determining the average percent of gaze locations in areas with higher than average surprise value.

B. Results

The results suggest that infants were more likely to look at surprising stimuli. Fig. 3 displays histograms comparing the infants' gaze distributions to randomly generated gaze distributions. For each infant, the distribution of infant gaze locations is shifted toward more surprising areas than a randomly generated distribution. While the distance between the distributions is small, the KL divergence is at the same order of magnitude as that found in Itti's and Baldi's work [15]. We calculated 100 random distributions to compute KL divergences and used a one-tailed t-test to test the hypothesis KL divergence > 0 for each infant, $p < 0.0001$. Over 67% of infant gaze locations were in areas of the video which were more surprising than average. This number is comparable to the 72% found during Itti's and Baldi's study with adult participants observing video scenes [15]. Values for individual infants are reported in Table I.

The realistic nature of the data and the age of the population involved in this work introduce challenges to using the Bayesian surprise model as a predictor of visual attention. The video data from the SAR leg movement study are not a prerecorded set of videos, but rather footage filmed from a head-mounted camera.

As such, the camera moves significantly more than typical video footage. This may cause surprise values to be higher than those of pre-recorded videos in certain areas. In addition, the infants were sometimes fussy or distracted by people in the room. While people may generate their own low-level surprise in the video data, infants may also be looking at humans for social purposes.

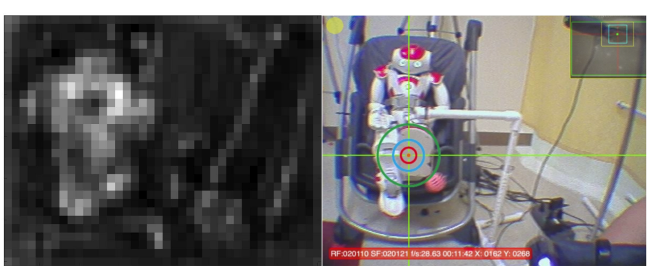

Fig. 2. Left: The surprise values of each 16x16 patch of pixels. Lighter pixels indicate higher surprise values. Right: Study environment from the infant's point of view with overlaid target to show infant gaze location during a robot kicking behavior. The circles indicate 2, 4, and 8 degrees from the estimated gaze location.

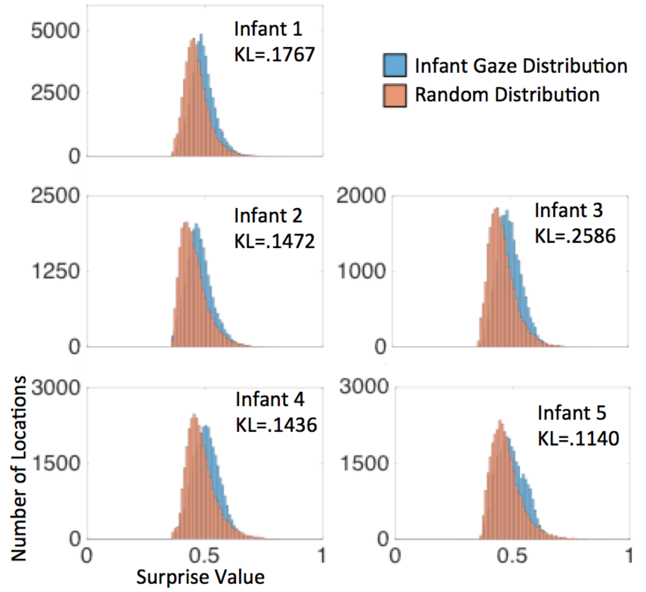

Fig. 3. Histograms and corresponding KL divergence values of infant and random gaze distributions. From each frame in an infant's gaze tracking video, we extracted the surprise value at the infant's gaze location and at a random location in the frame. This process was repeated to find the KL divergence for each infant 100 times ($p < 0.0001$ for each infant on a one-tailed t-test to test KL$>$0)

VI. Predicting Success of Robot Behavior in Acquiring Infant Visual Attention

A. Methodology

After validating the surprise model with infants, we wanted to explore more deeply how the amount of Bayesian surprise generated by the robot's behaviors could predict infant visual attention. Specifically, we were interested in determining whether the surprise model could be used to predict what percent of robot behaviors infants would look at during a specific time interval. The robot's behavior was represented as a 1-dimensional signal with a frequency of 30Hz. Values 1, 2, and 3 indicate robot kicking, robot kicking and lights, or robot kicking and sound, respectively. This numbering scheme was chosen to distinguish behaviors so that a change in behavior type may induce surprise; we also

TABLE I
PERCENT OF INFANT GAZE LOCATIONS IN REGIONS WITH HIGHER THAN AVERAGE SURPRISE VALUE

Infant	1	2	3	4	5	Weighted Average
Percent of Gaze Locations	71.24	70.60	71.97	65.08	60.42	67.97
Number of Frames	54574	25169	24753	44708	34079	

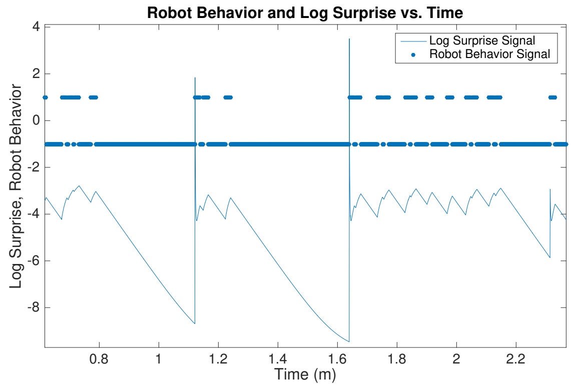

Fig. 4. The robot behavior signal and log surprise signal, ζ=0.98, for a 1.5-minute time interval. A robot behavior signal value of 1 indicates that the robot is kicking its leg, while a value of 0 indicates the robot is still.

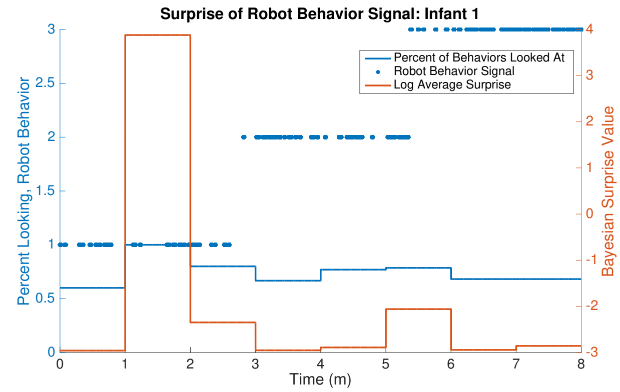

Fig. 5. Infant 1 looking behavior with robot behavior and surprise signal. The dotted line represents the robot behavior. Values 1, 2, and 3 indicate robot kicking, robot kicking and lights, or robot kicking and laughing, respectively. For Percent of Behaviors Looked At, a value of 1 on the y-axis corresponds to 100%.

evaluated the model reordering which behavior was 1, 2, or 3 to ensure the assignment did not have a significant effect. This signal was input into the surprise model. As the signal was one dimensional with a high frequency compared to number of signal value changes, we selected higher ζ values (0.98-0.99) for the forgetting factor. We also divided the behavior signal by 1000 and took the log of the surprise value to prevent the model from producing unreasonably high peaks during robot behavior onset. Fig. 4 displays the surprise signal and the robot behavior signal over a 1.5-minute window.

Video data from the head-mounted camera were annotated by two trained student annotators and one researcher to determine when the infant was looking at the robot. The method of annotation was conservative so as to minimize false positives classifying that an infant looked at the robot: an infant was only classified as looking at the robot if part of the robot was within a circle representing 2 degrees from the infant's predicted gaze location for three or more consecutive frames. 20% of the data were annotated by all three annotators. Interrater reliability was measured for this 20% using Fleiss' kappa, and a value of κ = 0.96 was achieved.

To evaluate whether surprise was predictive of the robot's success in acquiring infant attention, we compared the log of the average surprise value generated by the robot behavior signal each minute with the percent of robot behaviors an infant looked at each minute. We labeled an infant as looking at the robot if the infant looked at any part of the robot during the kick or within one second after the kick. We chose the 1 minute interval as it was small enough to make predictions over several time intervals, yet large enough that a looking behavior value would not be drastically influenced by the infant checking in with a parent. We used linear regression to generate a linear model of percent of robot behaviors looked at per minute versus log average surprise per minute.

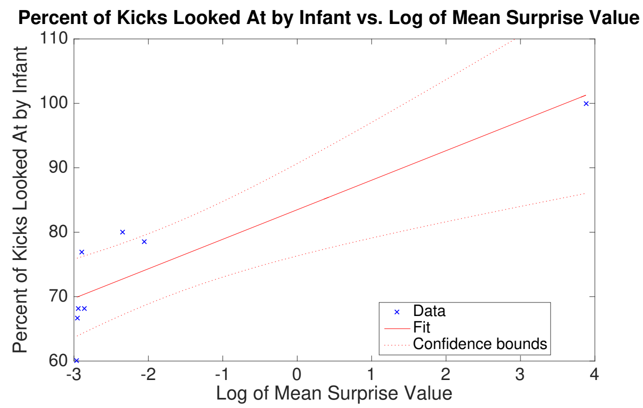

Fig. 6. The regression line and data from infant 1. The regression line is defined by equation $RBL = 83.49 + 4.58\ LAS$, showing a trend that infant 1 looked at a higher percentage of the robot behaviors during minutes with higher log average surprise value.

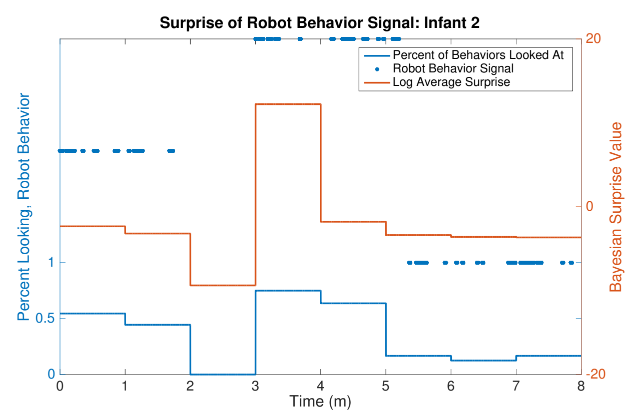

Fig. 7. Infant 2 looking behavior with robot behavior and surprise signal. The dotted line represents the robot behavior. Values 1, 2, and 3 indicate robot kicking, robot kicking and lights, or robot kicking and laughing, respectively. For Percent of Behaviors Looked At, a value of 1 on the y-axis corresponds to 100%.

B. Results

We used an ANOVA to test how well the regression equations fit the data, and whether the model was predictive for each infant. The log of the average surprise value per minute was significantly predictive of percent of robot behaviors

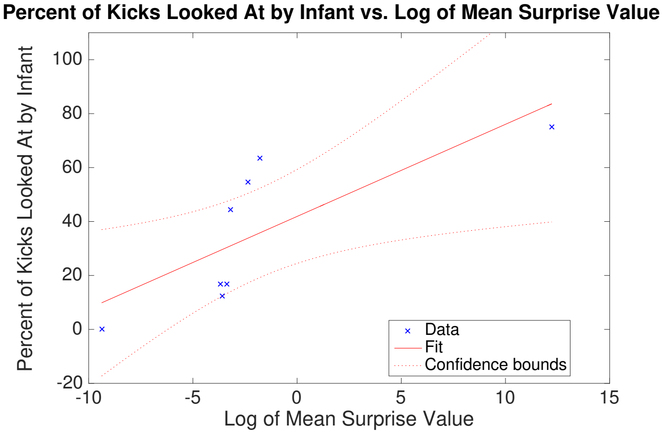

Fig. 8. The regression line and data from infant 2. The regression line is defined by equation $RBL = 41.89 + 3.42\ LAS$, showing a trend that infant 2 looked at a higher percentage of the robot behaviors during minutes with higher log average surprise value.

looked at each minute for 2 of the 5 infants, $p < 0.05$. The linear regression determined that the log of the average surprise per minute (LAS) was significantly predictive of infant 1's percent of robot behaviors looked at per minute (RBL), $F(1,6) = 20.655$, $p = 0.0039$. LAS accounted for 77.5% of the variance in RBL with the regression equation: $RBL = 83.49 + 4.58\ LAS$. Fig. 5 and Fig. 6 show infant 1's looking behavior compared with surprise.

For infant 2, LAS was predictive of RBL with $F(1,6) = 8.46$ and $p = 0.027$. LAS accounted for 58.5% of variance in the dependent variable. The regression equation was $RBL = 41.89 + 3.42\ LAS$. Infant 2's looking behavior compared with surprise is displayed in Fig. 7 and Fig. 8.

While the regression equations of infants 3, 4, and 5 suggested a positive correlation between LAS and RBL, the results for these infants were not statistically significant. We discuss possible reasons for this in the next section.

VII. Discussion and Future Work

The analysis of the video data and infant gaze locations in Section 5 demonstrates that all 5 infants tended to look at surprising areas of their environment. The Bayesian surprise model performed similarly with infant data as with adult data analyzed by Itti and Baldi [15], despite evaluating the model on moving video data from the infants' head-mounted camera instead of on pre-recorded videos used with adults. This suggests that surprise may be useful in designing and evaluating robot behaviors which attract infant attention.

The analysis of the robot behavior data and infant gaze behaviors in Section 6 shows that the surprise induced by the robot's kicking behaviors predicted the gaze behaviors of infants 1 and 2. These infants' tendencies to look more frequently during times with highly surprising behavior signals, especially as demonstrated by the rightmost data points in Fig. 6 and Fig. 8, show the promise of designing behaviors with high surprise values to attract infant visual attention.

When observing the recorded data of the other three infants, we saw that infants 4 and 5 looked less at surprising areas in general (Table I). In fact, infant 5 looked at other areas of the robot for long time intervals, instead of focusing specifically on the robot's leg during robot kicking actions. Infant 5's gaze behavior with respect to the Bayesian surprise value of the robot's behavior is shown in Fig. 9. Infant 3 appeared fussy significantly more than the other infants in the study, affecting their gaze behavior. Additionally, during infant 3's interaction, a long time interval between successive robot kicks caused a large increase in the surprise value of the Bayesian model. Since the percent of behaviors looked at by the infant is limited while the Bayesian surprise model is unbounded, the infant's looking behavior could not produce a similar spike (Fig. 10). Imposing an upper bound on the Bayesian surprise value in future work may help to mitigate against this effect. It is also possible that contextual information or external distractors may play more of a role in visual attention for some infants than for others.

This analysis serves as an initial exploration into surprise as a predictor of infant eye gaze as part of a larger effort to enable robots to acquire infant visual attention during SAR interactions. The evaluation was based on data from five infants; more work is needed to fully understand how surprise can predict infant eye gaze, and how it can be used to evaluate and inform the design of robot intervention behaviors. However, validating the Bayesian surprise model using similar analyses to Itti's and Baldi's work [15] suggests that surprise may be predictive of infant gaze location.

While our results indicate that surprise has some predictive power for infant visual attention, future work will involve a comparison with other models to provide a greater understanding of the most predictive features. Evaluating multiple models with a larger number of infants will help determine the extent to which external context and stimuli influence infant gaze behavior. This will also inform the extent to which robot behaviors need to be personalized to each infant.

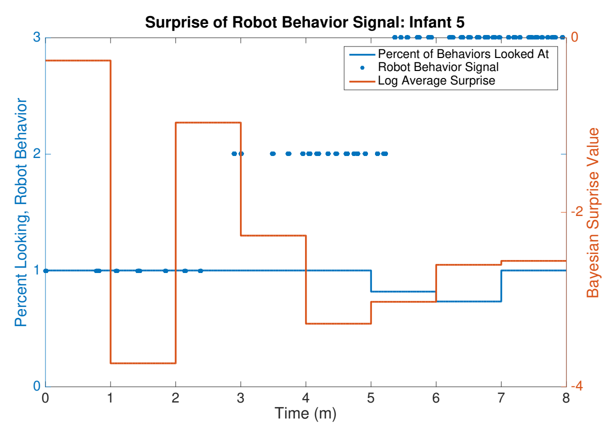

Fig. 9. Infant 5 looking behavior with robot behavior and surprise signal. The dotted line represents the robot behavior. Values 1, 2, and 3 indicate robot kicking, robot kicking and lights, or robot kicking and laughing, respectively. For Percent of Behaviors Looked At, a value of 1 on the y-axis corresponds to 100%. Infant 5 looks at the robot significantly more than the other infants, and looks at 100% of the robot behaviors during most minutes.

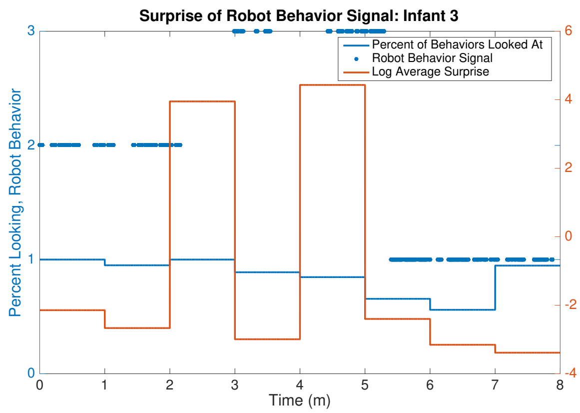

Fig. 10. Infant 3 looking behavior with robot behavior and surprise signal. The dotted line represents the robot behavior. Values 1, 2, and 3 indicate robot kicking, robot kicking and lights, or robot kicking and laughing, respectively. For Percent of Behaviors Looked At, a value of 1 on the y-axis corresponds to 100%. Significant spikes appear in the surprise signal during minutes 2 and 4 due to a long pause in robot kicking, which may reduce the predictive power of the signal.

Gender differences and distinctions between attention getting and attention holding may play a role in infant gaze behaviors; Cohen discusses that female infants are less likely than male infants to fixate longer on novel stimuli than familiar stimuli, though the number of fixations is not different [6]. Infants 4 and 5 in our study, who looked least often at surprising locations (Table I), and whose gaze behaviors were not predicted by the Bayesian surprise model, were both female. However, the size of the SAR leg movement study was too small to evaluate whether gender differences account for the variation in results between infants.

The optimal forgetting factor ζ may be different for each infant. As 6-8 month old infants are undergoing rapid cognitive development, the rate at which surprise fades may vary more in infants than in adults. The 8-minute contingency phase from the SAR leg movement study may not be enough time to learn an optimal time constant. Longer, repeated sessions may be required to more accurately predict gaze toward the robot. The large difference in regression functions between infants 1 and 2 also argues that models of attention may need to be personalized for each infant in order to best predict gaze behavior. Future work is needed to scale the size of the study and investigate the effects of individual differences on the predictive power of surprise.

VIII. Conclusion

This work studied surprise as a predictor of infant gaze during a SAR interaction involving leg movement training. The over-arching goal of the research is to inform the design of robot behaviors for future SAR interventions with infants. We found that all 5 infant participants looked most often at areas that were more surprising than average. While surprise was predictive of the percent of robot behaviors looked at per minute for two of the infants, more work is needed to understand how differences between infants and between the attention getting and attention holding phases influence the effect of surprise on visual attention to the robot. Our continuing work will explore these differences in order to further evaluate the application of surprise to attracting infant attention during SAR interactions.

Acknowledgment

We are grateful to Kiely Green and Gabriella Margarino for data annotation and to Wayne Deng, Dr. Barbara Sargent, and Dr. Nina Bradley for their valuable assistance.

References

[1] R. L. Holt and M. A. Mikati, "Care for child development: basic science rationale and effects of interventions," *Pediatric neurology*, vol. 44, no. 4, pp. 239–253, 2011.

[2] S. A. Rosenberg, C. C. Robinson, E. F. Shaw, and M. C. Ellison, "Part c early intervention for infants and toddlers: Percentage eligible versus served," *Pediatrics*, vol. 131, no. 1, pp. 38–46, 2013.

[3] R. Funke, J. C. Pulido, L. E. Eisenman, W. Deng, M. R. Rosales, N. S. Bradley, B. Sargent, B. a. Smith, and M. J. Matarić, "A data collection on the visual, physical,and behavioral reactions of an infant to a small humanoid robot," 2018.

[4] J. C. Pulido, R. Funke, J. García, B. A. Smith, and M. Matarić, "Adaptation of the difficulty level in an infant-robot movement contingency study," in *Workshop of Physical Agents*, pp. 70–83, Springer, 2018.

[5] B. Scassellati, J. Brawer, K. Tsui, S. Nasihati Gilani, M. Malzkuhn, B. Manini, A. Stone, G. Kartheiser, A. Merla, A. Shapiro, *et al.*, "Teaching language to deaf infants with a robot and a virtual human," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, p. 553, ACM, 2018.

[6] L. B. Cohen, "A two process model of infant visual attention," *Merrill-Palmer Quarterly of Behavior and Development*, vol. 19, no. 3, pp. 157–180, 1973.

[7] L. Itti and P. Baldi, "A principled approach to detecting surprising events in video," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, vol. 1, pp. 631–637, IEEE, 2005.

[8] N. T. Fitter, R. Funke, J. C. Pulido, L. E. Eisenman, W. Deng, M. R. Rosales, N. S. Bradley, B. Sargent, B. A. Smith, and M. J. Matarić, "Using a socially assistive humanoid robot to encourage infant leg motion training," *IEEE Robotics and Automation Magazine*, 2019.

[9] W. L. Johnson, J. W. Rickel, J. C. Lester, *et al.*, "Animated pedagogical agents: Face-to-face interaction in interactive learning environments," *International Journal of Artificial intelligence in education*, vol. 11, no. 1, pp. 47–78, 2000.

[10] H. Admoni, A. Dragan, S. S. Srinivasa, and B. Scassellati, "Deliberate delays during robot-to-human handovers improve compliance with gaze communication," in *2014 9th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pp. 49–56, IEEE, 2014.

[11] A. Ito, S. Hayakawa, and T. Terada, "Why robots need body for mind communication-an attempt of eye-contact between human and robot," in *RO-MAN 2004. 13th IEEE International Workshop on Robot and Human Interactive Communication (IEEE Catalog No. 04TH8759)*, pp. 473–478, IEEE, 2004.

[12] C. Yu, M. Scheutz, and P. Schermerhorn, "Investigating multimodal real-time patterns of joint attention in an hri word learning task," in *Proceedings of the 5th ACM/IEEE international conference on Human-robot interaction*, pp. 309–316, IEEE Press, 2010.

[13] A. Bruce, I. Nourbakhsh, and R. Simmons, "The role of expressiveness and attention in human-robot interaction," in *Proceedings 2002 IEEE International Conference on Robotics and Automation (Cat. No. 02CH37292)*, vol. 4, pp. 4138–4142, IEEE, 2002.

[14] C. Kidd, S. T. Piantadosi, and R. N. Aslin, "The goldilocks effect: Human infants allocate attention to visual sequences that are neither too simple nor too complex," *PloS one*, vol. 7, no. 5, p. e36399, 2012.

[15] L. Itti and P. F. Baldi, "Bayesian surprise attracts human attention," in *Advances in neural information processing systems*, pp. 547–554, 2006.

[16] C. Jiang, C. J. Lane, E. Perkins, D. Schiesel, and B. A. Smith, "Determining if wearable sensors affect infant leg movement frequency," *Developmental neurorehabilitation*, vol. 21, no. 2, pp. 133–136, 2018.