PNAS
www.pnas.org

Supplementary Information for

Origins of the concepts *cause*, *cost*, and *goal* in prereaching infants

Shari Liu, Neon B. Brooks, & Elizabeth S. Spelke

Correspondence to: shariliu01@g.harvard.edu

This PDF file includes:

www.pnas.org/cgi/doi/10.1073/pnas.1904410116

Procedures

Materials and Methods

Participants. *N*=152 healthy, full-term infants (Mean age=107 days, range=91-122, 78 female) were included in our final sample. An additional 50 infants participated but were excluded from our final sample due to fussiness (28 infants), inattentiveness (5 infants), caregiver interference (1 infant), experimenter or coding error (19 infants), or technical issues (5 infants). These exclusionary criteria were set prior to the start of data collection for all experiments and were pre-registered for Experiments 4 and 5. For exclusion information broken down by experiment, see Table S2.

Experimental Procedure. Caregivers were instructed to look away from the screen and not direct their infants' attention in any way for the entirety of the study. All videos were presented using Keynote. Prior to every trial, the experimenter played an attention-getting animation until infants looked towards the screen. Then, infants watched a ~4s video of an action, which paused on the last frame. All trials began on the first frame of the video, and ended when infants either looked for 45 cumulative seconds towards the screen or looked for 2 consecutive seconds away from the screen. Infants saw between 6 and 12 habituation videos followed by 3 pairs of efficient and inefficient test videos, order counterbalanced across participants. The experiment moved from habituation to test when infants' summed looking times towards the most recent 3 habituation trials fell to below half of their summed looking times towards the first 3 habituation trials, or after 12 habituation trials, whichever came first.

Data Coding and Reliability. Looking times were measured online using XHAB (1), and then coded offline in jhab (2) or Datavyu (3) with the same thresholds as online coding to check for coding errors and inattention, and these offline values entered the final analysis. To assess the reliability of our data, 50% of test trials from participants across Experiments 1-5 (76 participants, 456 trials) were randomly selected and coded by additional researchers who were unaware of experimental condition and test trial order. The intraclass correlation coefficient (ICC) between the original data, and this newly coded data, was 0.969 (95% CI [0.946, 0.982]), 0.969 (95% CI [0.943, 0.982]), 0.968 (95% CI [0.955, 0.978]), 0.963 (95% CI [0.938, 0.977]), and (95% CI 0.936 [0.911, 0.954]), for Experiments 1 through 5, respectively.

Supplementary Results

Comparing results including and excluding influential participants.

Across all of our experiments, we checked for influential participants in every model using Cook's Distance (4), as stated in the main text. This is a method for outlier detection: The purpose of this step in our analysis is to identify individuals whose inclusion in the analysis may have undue influence by either masking the effect, which is otherwise present across the rest of the sample, or by driving the effect, which is otherwise absent across the rest of the sample. Below are the results from the main text detailing how many influential participants were detected in each analysis, and report the results including these participants. Overall, none of the main conclusions reported in the main text differ depending on the inclusion or exclusion of these participants (though see below for minor differences in the findings from Experiment 2, the analysis collapsing across Experiment 1 and Skerry et al. (5), and the analysis collapsing across Experiments 1 and 2, none of which change the conclusions reported in the main text).

In the primary analysis for Experiment 1, we detected one influential participant. Including this participant in the analysis generates the same finding as reported in the main text: Infants looked equally at the efficient vs inefficient reach of a gloved hand ([-0.169,0.209], β=0.041, B=0.02, SE=0.092, p=0.831, two-tailed). In the analysis comparing Experiment 1 to Skerry et al. (5) Experiment 3, we detected 2 influential participants. Including them results in a null difference between these two experiments ([-0.086,0.415], β=0.07, B=0.165, SE=0.128, p=0.205, two-tailed), whereas excluding them results in a marginal difference across these two experiments (see main text). This difference does not change our interpretation of Experiment 1: Infants look equally to efficient and inefficient pickup actions when the person reaching wears a glove or a mitten.

In the primary analysis for Experiment 2, we detected 2 influential participants. Including them in the sample generates a marginal effect in the same direction as that reported in the main text: Infants looked longer at the inefficient than the efficient reach of a bare hand [-0.024,0.318], β=0.297, B=0.147, SE=0.083, p=0.091, two-tailed). In the analysis comparing Experiment 2 to Skerry et al. (5) Experiment 3, we detected 1 influential participant. Including them results in the same finding as reported in the main text: Infants' looking preferences significantly differed across these two experiments ([0.048,0.536], β=0.495, B=0.292, SE=0.121, p=0.021, two-tailed). These findings do not change our conclusion in the main text: Experiments 1 and 2 overall show that infants have inconsistent, fragile expectations about the efficiency of reaches that result in displacing objects.

In the analysis collapsing across Experiments 1 and 2, we found 1 influential participant. Inclusion of that participant generates a null effect ([-0.042,0.209], β=0.139, B=0.083, SE=0.063, p=0.191, two-tailed), whereas in the main text this effect was marginal, but our conclusion is the same: Infants look equally to efficient and inefficient reaches when these actions result in objects being displaced. In the analysis comparing infants' looking preferences across Experiments 1 and 2, we found 3 influential participants, and including them results in the same result as reported in the main text: Infants' looking preferences did not differ across the two experiments, ([-0.116,0.37], β=0.211, B=0.127, SE=0.124, p=0.311, two-tailed).

In the primary analysis for Experiment 3, we detected 2 influential participants. Including them in the sample generates the same finding as reported in the main text: Infants' looking preferences for the test events differed as a function of whether they were habituated to

constrained action over a barrier (experimental group) or the same actions not over a barrier (control group) ([0.115,0.657], β=0.596, B=0.502, SE=0.114, p<.001, two-tailed). In the experimental condition, infants looked longer at the inefficient action ([0.065,0.451], β=0.398, B=0.258, SE=0.095, p=0.01, two-tailed). In the control condition, infants looked equally at the two test actions ([-0.321,0.065], β=-0.198, B=-0.128, SE=0.095, p=0.186, two-tailed).

We did not detect any influential participants in the analyses for Experiments 4 and 5.

Comparing infants' responses to mittens, gloves, and bare hands

Because Experiments 1 and 2 used the methods of Skerry et al. (5) (SCS), the primary difference between the events from Skerry et al. (5), and Experiments 1 and 2 from the main text concerned the presentation of the reaching hand, which was bare in Experiment 2, covered by tight-fitting gloves in Experiment 1, and covered by thick mittens in SCS, as in all the prior published research involving mittens training. Could infants' responses to the reaches from these experiments be explained, in part, by how easy it was to see the configuration of the hand (easy in Experiment 2, slightly harder in Experiment 1, and even harder in SCS)? To explore this question, we analyzed infants' proportion looking to the indirect, inefficient action in Experiment 1, Experiment 2, and the comparable experiment from SCS where infants had no mittens training (SCS Experiment 3), and asked whether the clarity of the person's hand in each of these experiments (2 in Experiment 2, 1 in Experiment 1, and 0 in SCS Experiment 3) predicted differences in looking preferences, controlling for correlated data within experiments. This analysis revealed that the magnitude of infants' looking preference for the inefficient over the efficient action increased with increasingly clear information about the form of the hand ([0.007,0.053], β=0.416, B=0.03, SE=0.011, p=0.011, two-tailed, excluding 4 influential participants). This finding held regardless of whether the influential participants were excluded or included (in the latter case, [0.005,0.053], β=0.359, B=0.029, SE=0.012, p=0.02, two-tailed).

Meta-analysis

To assess the effects of our experimental manipulations in Experiments 1-5 and in Skerry et al. (5), we performed an analysis over these two papers (total N=264, 12 conditions). Our analytic approach allows us to assess the independent effects of 5 manipulations: the type of or absence of motor training, the presence or absence of a barrier preventing a direct reach for the object during habituation, the nature of the goal (to change the state of an object or pick it up), the presence or absence of action on contact, and the presence or absence of mittens on the actor. The analysis also allows us to control for the participant variables of age and sex, and model the nested structure of the data (e.g. looks clustered within experiments and within papers). For ease of interpretation, we used average proportion looking to the inefficient action in this analysis, following Skerry et al (5)[1]. The findings below exclude 16 participants on the basis of Cook's

[1] Although this analysis condenses all the manipulations from these 10 experiments while taking into account data correlated within experiments and papers, only future experiments will give conclusive evidence for the independent contribution of each manipulation. For example, because no experiment in this analysis includes mittens training and state change events (only mittens training with pick up events or no mittens training with state change events), it is unclear whether the effects of these two manipulations are additive or redundant.

Distance, leaving 248 infants in the final sample. See Table S1 for results including all participants.

This analysis confirmed the findings from the individual experiments reported in the main text and in Skerry et al (5): Infants' looking preference for the inefficient action was stronger when the observed action was spatiotemporally continuous with its effect (i.e., appeared to be causal) ([0.025,0.058], β=0.467, B=0.041, SE=0.009, p<.001, two-tailed when infants received effective motor training (sticky mittens), relative to no training ([0.029,0.074], β=0.583, B=0.052, SE=0.011, p<.001, two-tailed); when the observed agent's actions were constrained by a barrier and were efficiently adapted to that barrier, relative to the same actions that were unconstrained by a barrier ([0.021,0.051], β=0.406, B=0.036, SE=0.008, p<.001, two-tailed); and when the agent pursued a state change goal, relative to a pickup goal ([0.011,0.042], β=0.302, B=0.027, SE=0.008, p=0.001, two-tailed). We also found that infants' looking preference for the inefficient reach was *smaller* when they received ineffective motor training (non-sticky mittens), relative to no training, ([-0.058,0], β=-0.33, B=-0.029, SE=0.015, p=0.051, two-tailed), and improved as the form of the hand became clearer ([0.003,0.056], β=0.333, B=0.03, SE=0.013, p=0.027, two-tailed), but neither of the latter two findings was present in the full analysis with all participants (see Table S1). These findings provide further evidence that action experience alters action interpretation, but so does causal information.

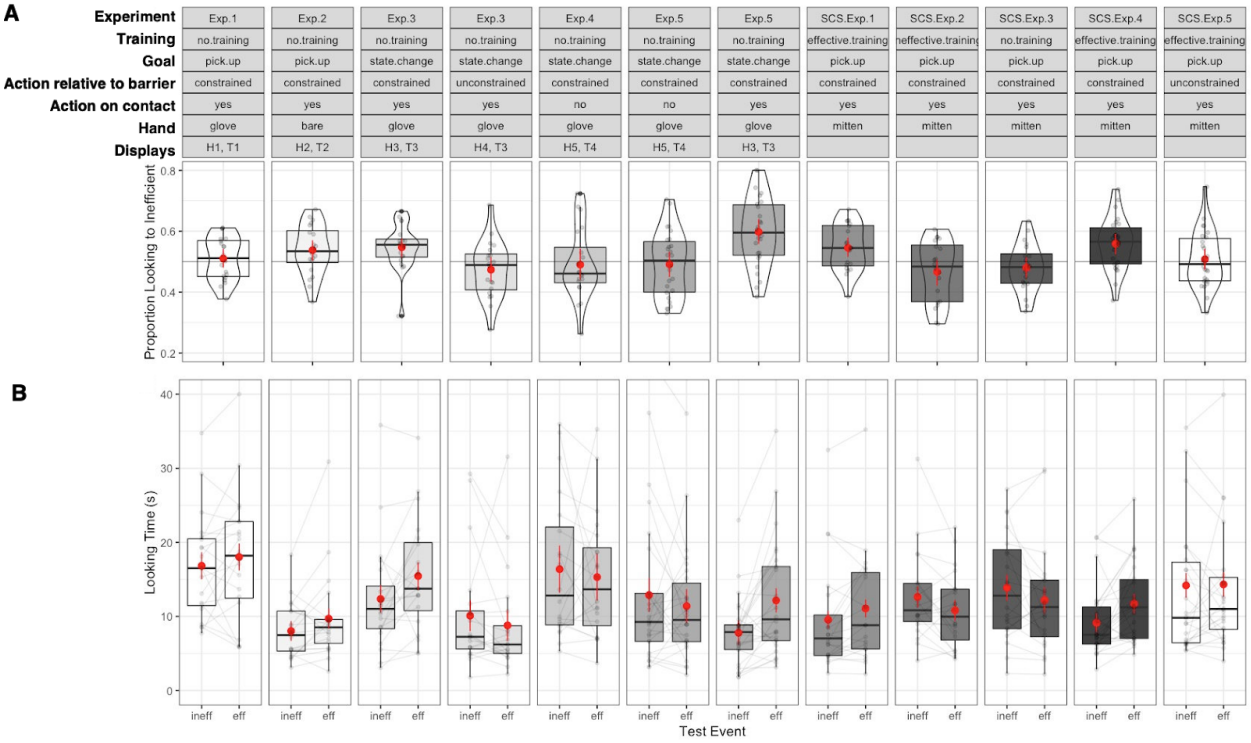

Figure S1. (A) Proportion looking towards the inefficient reach and (B) looking time in seconds towards the efficient versus inefficient reach, and at test across Experiments 1-5 (n=152) and across Experiments 1-5 in Skerry et al. (5) (SCS) (n=112). Labels above each panel list the experiment name (Exp. 1-5, SCS Exp. 1-5), type of motor training (none, ineffective non-sticky mittens, or effective sticky mittens), whether actions during habituation were constrained or unconstrained by a barrier, goal (state.change or pick.up), whether actions resulted in contact with the object, whether the actor reached with a bare, gloved, or mittened hand, and video displays listed in Figure 1. Error bars around means indicate within-subjects 95% confidence intervals (B) and bootstrapped 95% confidence intervals (A). Individual points (A) or pairs of connected points (B) indicate data from a single participant. Horizontal bars within boxes indicate medians, and boxes indicate the middle 2 quartiles of data. Violin plots (A) indicate distribution of data, area scaled proportionally to the number of observations. All data and

analyses are open source and available at https://osf.io/rcsns/.

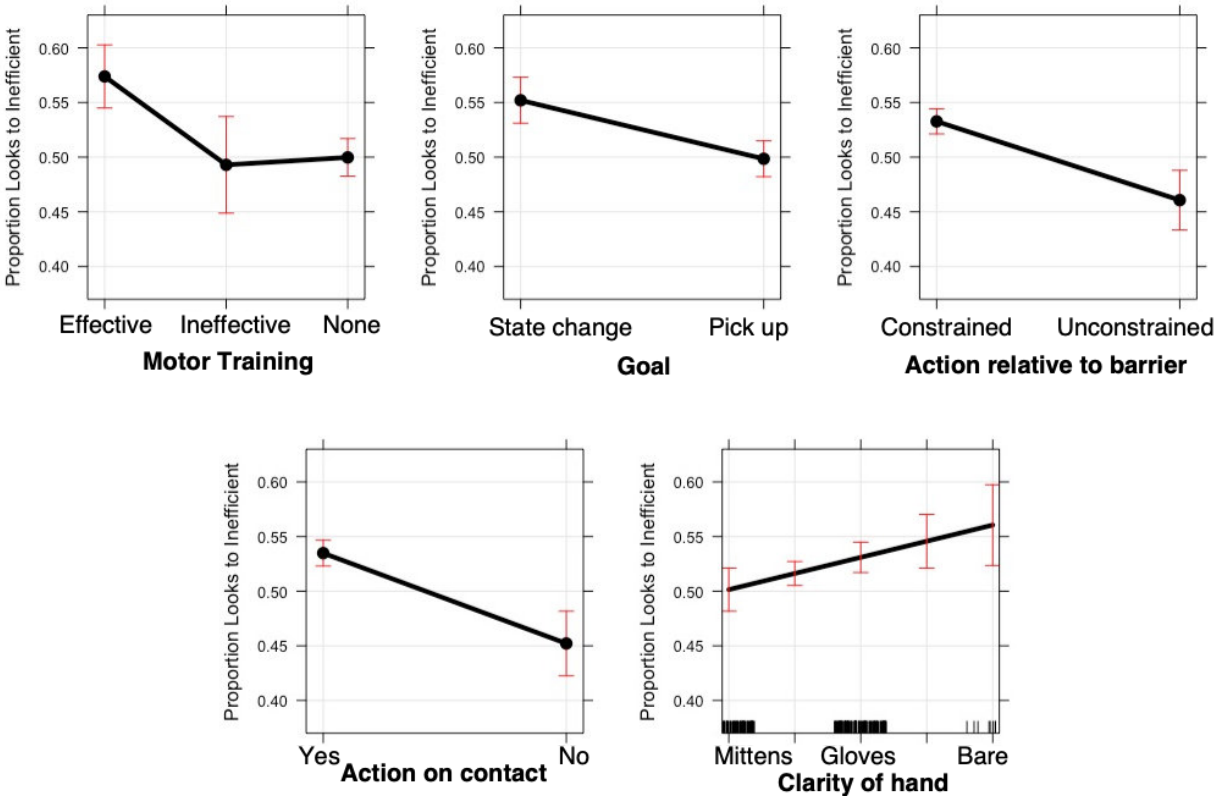

Figure S2. Effect plots for model investigating predictors of infants' looking preference for the inefficient versus efficient reach across Experiments 1-5 and Skerry et al. (5) (total N=264, 248 included in the final analysis, 16 excluded on the basis of Cook's Distance). Each point shows estimates of effects at each level of all predictors: Type of motor training (none, ineffective non-sticky mittens, or effective sticky mittens), the goal of the actor (state change vs pick up), action during habituation (constrained or unconstrained by a barrier), whether actions resulted in contact with the object (yes or no), and the clarity of the form of the hand (0=mittens, 1=gloves, 2=bare hand). Error bars indicate 95% confidence intervals. See Table S1 for full results.

Table S1. (A) Regression table for model investigating predictors of infants' looking preference for the inefficient over the inefficient reach across Experiment 1-5 and all experiments from Skerry et al. (5) (total N=264, 248 included in the final analysis, 16 excluded on the basis of Cook's Distance). (B) Regression table for the same analysis, including all participants. Dependent measure is proportion looking towards the inefficient reach, averaged across 3 test trials during test. Categorical predictors were coded using sum contrasts, and fixed effects from the model should therefore be interpreted with respect to the grand mean. Model formula: `prop.ineff.all ~ training + goal + hab + causal + clarity + (1|experiment) + (1|ageday) + (1|sex) + (1|paper).`

A	Standardized Estimate (ß)	Estimate (B)	Standard Error (SE)	df	t	p	95% CI (Lower)	95% CI (Upper)
(Intercept)	-0.716	0.457	0.012	241	37.17	<0.001	0.433	0.481
Effective training	0.583	0.052	0.011	241	4.50	0.000	0.029	0.074
Ineffective training	-0.330	-0.029	0.015	241	-1.96	0.051	-0.058	0.000
State change goal	0.302	0.027	0.008	241	3.39	0.001	0.011	0.042
Reach constrained by barrier	0.406	0.036	0.008	241	4.67	<0.001	0.021	0.051
Action on contact	0.467	0.041	0.009	241	4.86	<0.001	0.025	0.058
Clarity of hand	0.333	0.030	0.013	241	2.22	0.027	0.003	0.056

B	Standardized Estimate (ß)	Estimate (B)	Standard Error (SE)	df	t	p	95% CI (Lower)	95% CI (Upper)
(Intercept)	-0.640	0.455	0.016	10.73	28.34	<0.010	0.424	0.487
Effective training	0.578	0.058	0.015	7.55	3.79	0.006	0.028	0.088
Ineffective training	-0.355	-0.035	0.019	7.58	-1.86	0.102	-0.073	0.002
State change goal	0.301	0.030	0.011	6.75	2.79	0.028	0.009	0.051
Reach constrained by barrier	0.350	0.035	0.010	17.43	3.61	0.002	0.016	0.054

Action on contact	0.423	0.042	0.010	37.41	4.11	<0.001	0.022	0.062
Clarity of hand	0.294	0.029	0.018	7.48	1.66	0.137	-0.005	0.064

Exclusion info

Table S2. Tally of infants who participated in Experiments 1-5 but were excluded in our final sample. These exclusion criteria vary slightly across experiments (e.g. we relaxed our definition of inattentiveness from excluding all data from a participant if they missed a test trial, or if that trial was miscoded, in Experiment 3, to excluding data from just that trial in all other experiments).

Experiment	Fussy	Inattentive	Caregiver Interference	Experimenter/ Coding Error	Technical Failure	Total
Exp.1	7	0	0	2	0	7
Exp.2	6	0	0	1	2	9
Exp.3	9	5	1	12	3	30
Exp.4	0	0	0	2	0	2
Exp.5	6	0	0	2	0	8
Total	28	5	1	19	5	50

Distribution of Looking Times, Experiments 1-5

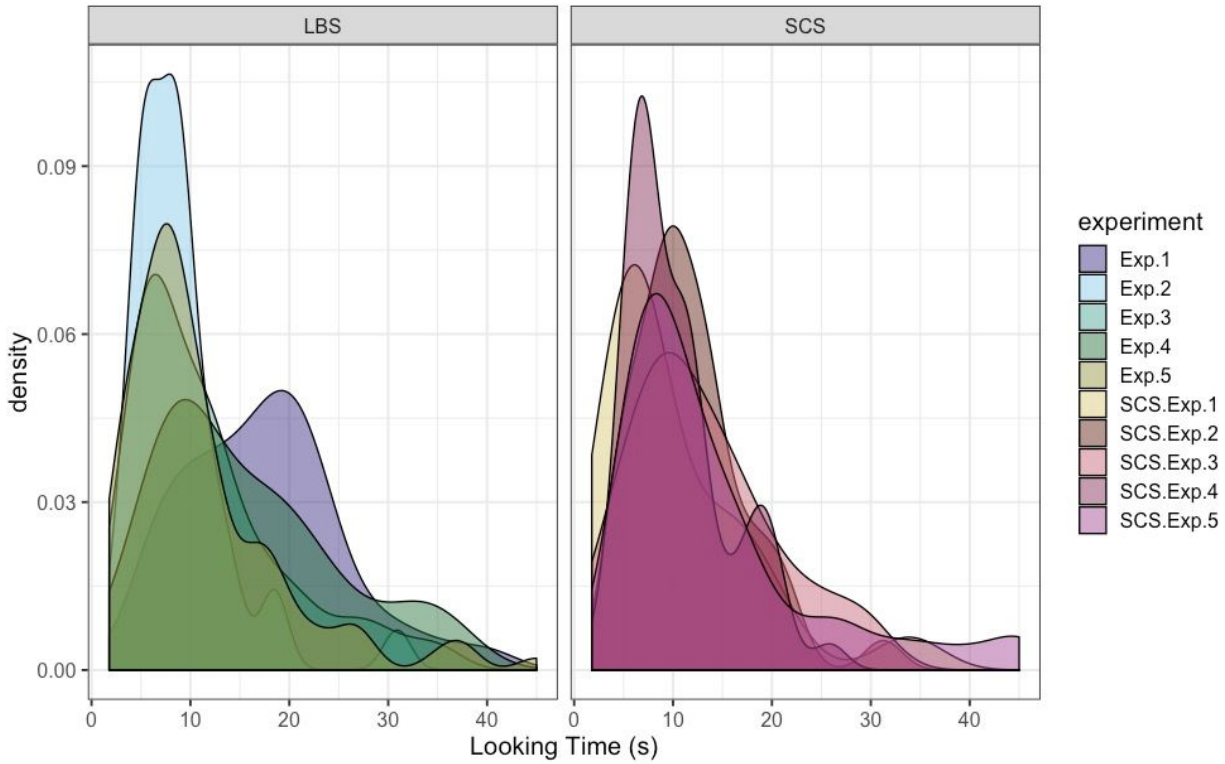

Figure S3. Density plot of looking times during test across Experiments 1-5 from the current paper (Liu, Brooks & Spelke, LBS, left panel), and Experiments 1-5 from Skerry et al. (5) (SCS, right panel) (N=264). Maximum-likelihood fitting revealed that the lognormal distribution (log likelihood=-1720.509) provides a better fit to these data than the normal distribution (log likelihood=-1842.196).

Attention to actions during habituation trials in Experiments 1-5

To ask whether infants' total attention during habituation was affected by experimental manipulations across Experiments 1-5 (action constrained vs unconstrained by a barrier, state change vs pickup goal, mittened, gloved, or bare-handed actor, and action with vs without contact with the object), and varied by gender and age, we fit a mixed effects model on these fixed effects and experiment (1-5) as a random intercept. We found that the only robust predictor of attention during habituation was age, [-3.4,-0.714], ß=-0.233, B(SE)=-2.058(0.68), p=0.003, two-tailed, such that older infants looked for a shorter time overall than younger infants.

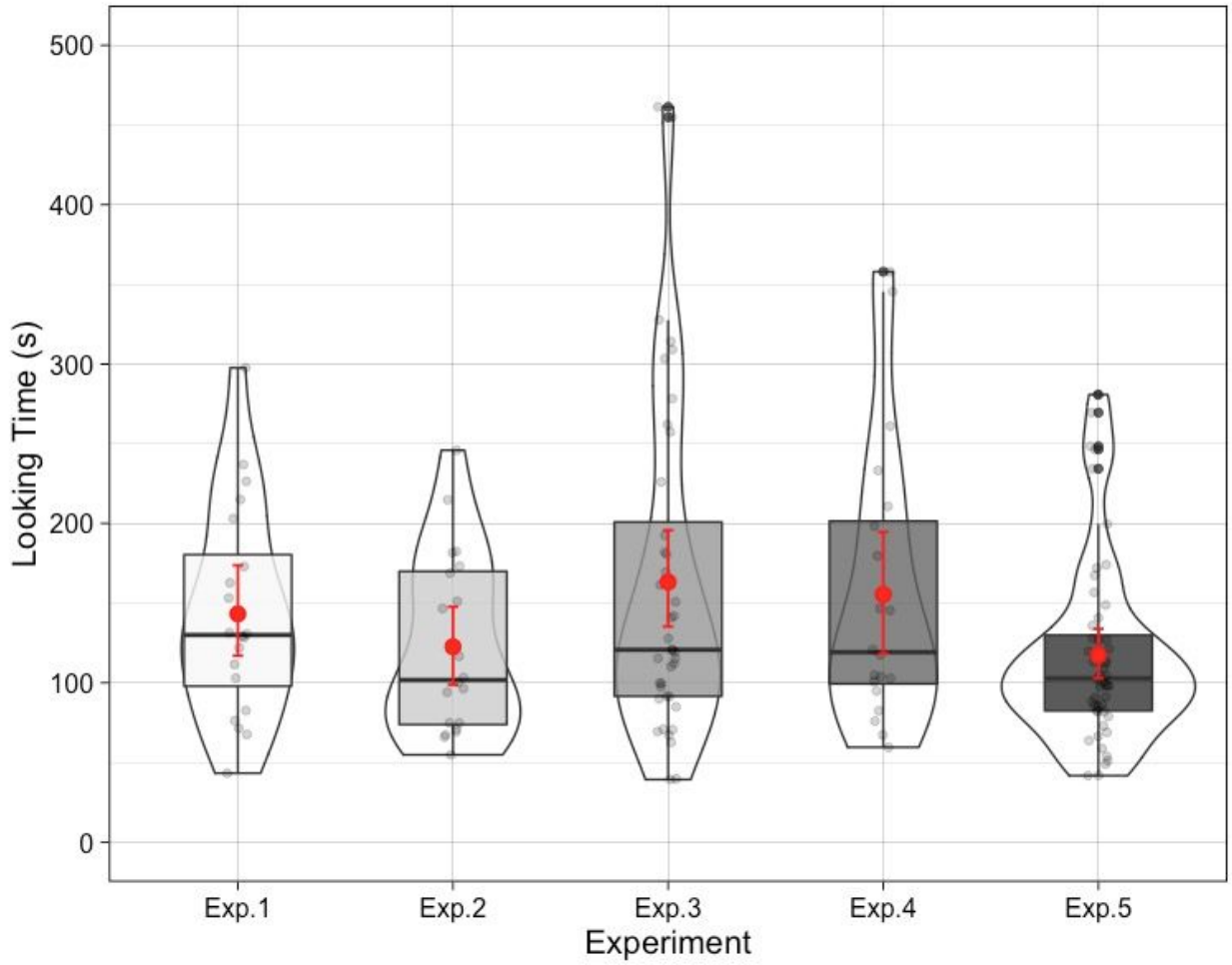

Figure S4. Total looking time in seconds during habituation across Experiments 1-5. Error bars around means indicate bootstrapped 95% confidence intervals (CIs). Individual points indicate data from a single participant. Horizontal bars within boxes indicate medians, and boxes indicate the middle 2 quartiles of data. Violin plots in indicate distribution of data, area scaled proportionally to the number of observations.

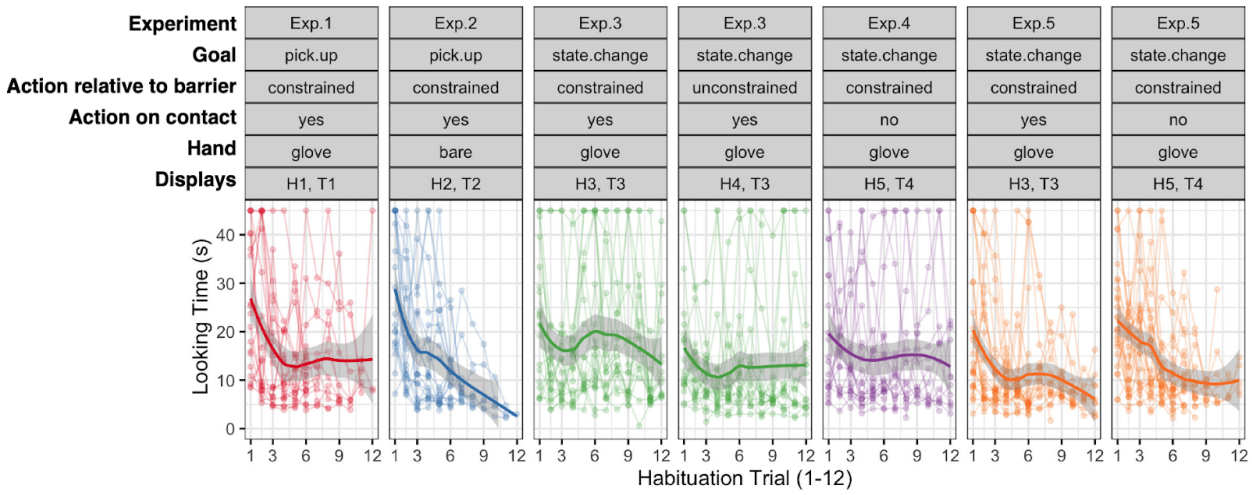

Figure S5. Looking time in seconds during each habituation trial across Experiments 1-5. Curves with 95% confidence interval ribbons indicate smoothed conditional means, generated using the loess method. Connected points indicate data from a single participant. Labels above each panel list the experiment name (Exp. 1-5), whether actions during habituation were constrained or unconstrained by a barrier, goal (state.change or pick.up), whether actions resulted in contact with the object, whether the actor reached with a mittened, gloved, or bare hand, and video displays listed in Figure 1.

Table S3. Regression table for mixed effects model analyzing the effect of age, sex, order of test events, habituation condition, goal, coverage of the hand, and causal information on total looking time habituation, controlling for other variations across Experiments 1-5. Model formula: total_hab ~ ageday + sex + first.test + hab + goal + clarity + causal + (1|experiment)

	Standardized Estimate (ß)	Estimate (B)	Standard Error (SE)	df	t	p	95% CI (Lower)	95% CI (Upper)
(Intercept)	0.171	373.160	89.49	51.41	4.170	0.000	194.14	553.006
Age in days	-0.233	-2.058	0.68	147.54	-3.026	0.003	-3.40	-0.714
Sex	0.066	5.203	6.11	148.65	0.852	0.396	-6.92	17.274
First test event	-0.006	-0.439	6.00	146.65	-0.073	0.942	-12.27	11.393
Action over a barrier	0.222	17.590	11.03	131.57	1.595	0.113	-6.44	41.220
Goal	0.007	0.589	16.02	6.18	0.037	0.972	-37.44	37.615
Clarity of hand	-0.253	-19.993	38.05	5.73	-0.525	0.619	-110.17	70.275
Action on contact	-0.055	-4.379	9.08	75.38	-0.482	0.631	-23.75	13.935

References

1. Pinto J (1995) *XHAB64* (Stanford University, Palo Alto, CA).

2. Casstevens RM (2007) *jHab: Java Habituation Software* (Chevy Chase, MD).

3. Datavyu Team (2014) *Datavyu: A Video Coding Tool* Available at: http://datavyu.org.

4. Nieuwenhuis R, te Grotenhuis M, Pelzer B (2012) Influence.ME: Tools for detecting influential data in mixed effects models. *R J* 4(2):38–47.

5. Skerry AE, Carey SE, Spelke ES (2013) First-person action experience reveals sensitivity to action efficiency in prereaching infants. *Proc Natl Acad Sci U S A* 110(46):18728–18733.