Ten-month-old infants infer the value of goals from the costs of actions

Authors: Shari Liu*[1,3], Tomer D. Ullman[1,2,3], Joshua B. Tenenbaum[2,3], & Elizabeth S. Spelke[1,3]

Affiliations:
[1]Department of Psychology, Harvard University
[2]Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology
[3]Center for Brains, Minds, and Machines (CBMM)

*Correspondence to: shariliu01@g.harvard.edu

Abstract: Infants understand that people pursue goals, but how do they learn which goals people prefer? Here, we test whether infants solve this problem by inverting a mental model of action planning, trading off the costs of acting against the rewards actions bring. After seeing an agent attain two goals equally often at varying costs, infants expected the agent to prefer the goal it attained through costlier actions. These expectations held across three experiments conveying cost through different physical path features (jump height and width; incline angle), suggesting that an abstract variable, such as 'force', 'work' or 'effort', supported infants' inferences. We model infants' expectations as Bayesian inferences over utility-theoretic calculations, providing a bridge to recent quantitative accounts of action understanding in older children and adults.

One Sentence Summary: Infants use the amount of work or effort an agent is willing to expend, to determine the value of the agent's goal.

Main Text:

When we observe people's actions, we see more than bodies moving in space. A hand reaching for an apple is not just one object decreasing its distance from another; it can indicate hunger (in the person who is reaching), helpfulness (if the person is reaching on behalf of someone else) or compromise (if the person reaching would prefer a banana, but not enough to go buy one). This fast and automatic ability to interpret the behavior of others as intentional, goal-directed, and constrained by the physical environment is often termed 'intuitive psychology' (*1–4*). Here, we use behavioral experiments and computational models to probe the developmental origins and nature of this ability.

Over the past two decades, research has revealed that the building blocks of our intuitive psychology are present as early as the first year of life. Despite infants' limited experience, their interpretations of other people's actions are guided by assumptions about agents' physical properties (*5*), intentions and goals (*6*), mental states (*7–10*), causal powers (*11*), and dispositions to act efficiently (*7*, *12*, *13*). This wealth of findings does not reveal, however, whether infants' capacities depend on a host of distinct local abilities (*14–16*), or on a single coherent system supporting inference, prediction, and learning (*3*, *17–19*).

Here, we tackle this question in a case study, based on a computationally precise proposal for a coherent, abstract, and productive system for action understanding (Fig. 1). Previous studies suggest that infants are sensitive to the costs of agents' actions (*3*, *7*, *12*, *13*), and can infer agents' preferences (*6*, *9*). Decision theorists for hundreds of years have recognized these as the two central factors guiding the decisions of rational agents (*20–22*). Here we ask whether infants

can integrate these dimensions to infer agents' goals: Do infants use the cost that an agent expends to attain a goal state in order to infer the value of that goal state for the agent?
[Fig 1 here]

Such an inference has been proposed to rest on three nested assumptions that together constitute a "naïve utility calculus" (*23*), analogous to classical economic thinking. First, agents act to maximize their utility U, under constraints (*2*, *4*, *24*, *25*). Second, this utility separates into rewards and costs, two distinct components that can be individual targets of inference (*26*). That is, if R(S) is the reward of a goal state S, and C(A) is the cost of an action, then an agent acts to maximize the following:

$$U(A,S) = R(S) - C(A). \qquad (1)$$

Third, the cost of an action is not arbitrary but depends in general on properties of both the agent and the situation they are in, which determine how much effort the agent might need to exert to carry out that action.

These assumptions can be formalized as generative models that successfully predict the quantitative and qualitative behavior of adults and older children (*4*, *23*, *26*). In these models, observers who reason that other agents are maximizing their expected utility according to Eq. 1 can use what they know about rewards and costs to predict the agents' future actions. Inverting this process, observers can use the agents' overt actions to infer their hidden rewards and costs, according to:

$$P(R,C \mid A) \propto P(A \mid C,R) \cdot P(R,C), \qquad (2)$$

where $P(R,C|A)$ is the posterior distribution over the rewards and costs of an agent. By Bayes' theorem, this distribution is proportional to the product of $P(A|C,R)$ — the likelihood of the agent choosing action A given rewards R and costs C, given by a rational planning procedure (*4*, *23*) —and $P(R,C)$, a prior distribution over costs and rewards.

Do infants apply the logic of cost-reward reasoning? Past research suggests that infants are sensitive to the relative value of different goal objects for an agent who chooses to approach one object over another (*6*, *27*) as well as the relative efficiency of the actions taken by an agent who approaches a goal object (*12*, *13*, *28*). Past studies do not reveal, however, whether infants have a unified intuitive psychology in the form of a generative model, or separate representations for variables like cost and reward that become unified only later in development, as children gain experience exerting themselves to achieve goals or communicating with others about their desires and actions. It is also an open question whether infants consider cost and reward in terms of abstract variables, such as work, effort, desire or value, or whether their understanding is restricted to perceptual features of actions, such as the distance or duration an agent travels, or the number of times it selects a particular goal. In physical action contexts, effort often covaries with perceptible properties such as the length or duration of a path traveled, but it depends ultimately on the amount of force that the agent must exert over time and distance (i.e., the amount of work the agent must do). Likewise, value often covaries with the number of times agent selects a goal, but ultimately depends on how strongly the agent desires a goal relative to the cost of achieving it or its value relative to other options.

We designed and conducted three experiments to test whether infants learn about the reward agents place on goals from cost, working backwards from the assumption that agents maximize utility and inferring relative rewards from observed actions under varying costs. We then use the data from these experiments, together with the findings from past experiments (*6*, *7*, *13*), to test a variety of computational models of infants' performance, including models with integrated versus isolated, and abstract versus cue-based, representations of costs and rewards (see model description in Supplementary Materials). Our empirical and computational findings support the view that a productive system grounded in costs-reward tradeoffs guides action understanding toward the end of the first year of life.

[Fig 2 here]

We tested N =80 ten-month-old infants in three experiments with pre-specified designs, procedures, sample sizes, and analysis plans (*29*). In all experiments, infants first saw an agent move to and refuse to move to each of two target goals under conditions of varying cost. Then infants watched test events in which the agent chose either the higher or the lower value target when both were present at equal cost. If infants infer the reward of the targets to the agent from the effort undertaken to reach them, and then they should be more surprised when the agent chooses the lower value target, looking longer at the test trials displaying that action (*30*).

In Experiment 1 (N=24), we leveraged events widely used in studies of early action understanding, wherein animated characters jump efficiently over barriers of variable heights to arrive at goal objects (*3*, *7*, *13*, *31*), and indicate their preferences by selecting one goal over another (*6*, *9*). During familiarization, infants watched six trials consisting of four different events involving a central agent and one of two target individuals on a level surface (Fig. 2A; Movie S1). In each event, the target jumped and made a noise, and the agent responded by turning to face and beginning to approach the target, whereupon a barrier fell onto the stage between directly in the agent's path. On two of these events (one for each target), the agent looked to the top of the barrier, made a positive "Mmmm!" sound, backed up and then jumped over the barrier, landing next to the target. On the other two events, the agent looked to the top of the barrier, made a neutral "Hmmm…" sound, backed away, and returned to its initial position. The critical distinction between these events concerned the height of the barrier, and therefore the length, height, and speed of the jump that the agent undertook so as to clear it (all jumps were equated for duration). For one target, the agent jumped over a low barrier and declined to jump a medium barrier; for the other target, the agent jumped the medium barrier and declined a tall barrier. After this familiarization, the agent appeared between the two equidistant targets on a level surface. Infants viewed two pairs of looped test events (Fig. 2D-2E, Movies S4-5), order counterbalanced, in which the agent looked at each of the targets and then repeatedly approached either the higher or the lower value target. Our pre-specified dependent measure was average log-transformed looking time (*32*) across test trials; we predicted differential looking at the test events but did not pre-specify the direction of this difference.

Infants looked longer at test trials in which the agent chose the target for whom it had jumped a lower barrier (M=28.41s, SD=14.85), relative to the target for whom it had jumped a higher barrier (M=21.79s, SD=12.29) (Fig. 3), 95% CI [0.062, 0.591], B=0.327, SE=0.130, β=0.502, $t(24)$=2.523, p=.019, two-tailed, mixed effects model with random intercept for participant (*30*). These findings suggest that infants inferred the rewards that the central agent placed over the targets from the cost the agent was willing to expend to reach these targets, and they therefore expected the agent to choose that target at test. Nevertheless, Experiment 1 does not show whether infants used the physical effort undertaken by the agent, or variables that merely correlate with effort (e.g. distance, speed), in their predictions.

[Fig 3 here]

To control for distance and speed of travel, Experiment 2 (N=24) used ramps of three different inclines to convey cost (Fig. 2B; Movie S2). On each familiarization trial, a target appeared on the top of one ramp, and the agent looked up the ramp and either climbed to the target or returned to its starting position. The agent climbed the shallow ramp and declined to climb the medium ramp for one target, and climbed the medium ramp and declined the steep ramp for the other target. The methods were otherwise the same as in Experiment 1. Consistent with our pre-specified directional prediction, infants again looked longer at the test events in which the agent approached the lower value target (*M*=30.94s, *SD*=13.31) than the higher value target (*M*=27.05s, *SD*=17.55) (Fig. 3), 95% CI [0.028, 0.472], *B*=0.250, *SE*=0.109, *ß*=0.408, *t*(24)=2.294, *p*=.015, one-tailed (*30*). This finding further suggests that infants understand agents' actions in accord with abstract, general and interconnected concepts of cost and reward, but narrower explanations remain. In Experiments 1 and 2, the agent was confronted with an obstacle to its forward motion (a barrier or ramp), and the size of the obstacle covaried with the cost of the agent's action, requiring the agent to move further upward to attain the higher value target. Because infants become sensitive to the effects of gravity on objects' on inclined planes well before 10 months of age (*33*), they may learn that agents will move to greater heights or overcome higher obstacles for more rewarding targets, without invoking a more abstract representation of physical effort. Experiment 3 was undertaken to explore these interpretations.

In Experiment 3 (*N*=32), the agent was separated from each of the two targets during familiarization not by an obstacle but by a horizontal gap in the supporting surface (Fig. 2C, Movie S3). Infants first saw a ball roll off the edge of a narrow, medium, and wide gap, and shatter (Movie S6). During familiarization, these three trenches, requiring jumps of variable lengths and speeds but of equal durations and heights, were interposed between the agent and target; the agent moved to the edge of a trench, looked at the far side, and then jumped over a narrow trench for one target (and refused the medium trench), and a medium trench for the other target (and refused the widest trench). The methods were otherwise unchanged (Fig. 2E, Movie S5). The methods and analyses for Experiment 3 were preregistered at https://osf.io/k7yjt/ (*29*) and tested the same directional prediction as Experiment 2. Infants again looked longer at the lower value choice (*M*=23.05s, *SD*=13.58) relative to the higher value choice (*M*=17.47s, *SD*=10.69) (Fig. 3), 95% CI [0.020, 0.501], *B*=0.260, *SE*=0.119, *ß*=0.403, *t*(32)=2.185, *p*=.018, one-tailed (*30*).

Regardless of whether an agent cleared higher barriers (Exp. 1), climbed steeper ramps (Exp. 2) or jumped wider gaps (Exp. 3) for one target over the other, infants expected the agent to choose that target at test. Across all experiments, infants looked longer at the lower value action (*M*=26.99s, *SD*=14.13) than the higher value action (*M*=21.64s, *SD*=13.94), 95% CI [0.139, 0.415], *B*=0.277, *SE*=0.070, *ß*=0.424, *t*(80)=3.975, *p*<.001, one-tailed, mixed effects model with random intercepts for participant and experiment, supporting our general hypothesis that infants infer the values of agents' goals from the costs of their actions. Although past research had shown that infants represent the goal of an agent's action from observations of an agent's choices between two objects (*6*) and expect agents to give different emotional responses when agents complete versus fail to complete their goals (*31*), the present experiments provide evidence that infants develop ordinal representations of reward even when the number of choices and expressed emotions are equated across the actions and only the costs of the actions vary. Moreover, they show that infants do not simply attribute higher reward to goals that agents pursue for a longer duration or attain with greater frequency, because these variables were equated as well. The findings provide evidence for longstanding suggestions that infants represent physical cost as a continuous variable that agents seek to minimize (*3*, *13*): Infants make appropriate cost assessments even when the specific physical features that distinguished lower- from higher-cost actions, including the relative length, curvature, duration or speed of a motion trajectories, systematically varied. Together, Experiments 1-3 suggest that infants represent cost and reward as interconnected, abstract variables that they apply to a wide range of events.

The discovery that infants infer the rewards of goals from the costs of achieving them provides empirical support for the thesis that an abstract and productive system guides infants' analysis of agents and their actions (*3*, *17*, *19*). Specifically, we suggest that the cognitive machinery supporting infants' intuitive psychology includes a mental model both of how agents plan actions in the forward direction, in accord with maximizing their utilities (Eq. 1) (*23*), and a procedure for inverting this model, in accord with the computational framework of inverse planning (Eq. 2) (*4*). Applying this general framework to our specific experiments, we posit that infants have developed a model of action planning prior to the experiment: they assume that agents value some goal objects more than others, and to engage in costlier actions to achieve goals with higher reward. When the infants see the agent take costlier actions to arrive at one target than at another, they invert this model to infer the relative reward of the two targets to that agent. Then when they see the agent flanked by the two targets in a situation where costs are equal, they apply their knowledge of the targets' relative value to the agent to run their planning model for that agent forward, predicting the target that it will approach. We have implemented this hypothesis in a computational model that accounts not only for the findings of the present experiments but also for a range of past studies of early action understanding (*6*, *7*, *13*). Furthermore, we compared this model to an array of simpler models that focus only on relative costs or rewards in isolation, or on particular cues to effort or value. We find that the only the full model with abstract variables for costs and rewards can account for all of the findings (Fig. S3; see Supplementary Material for details).

The present studies raise key questions for future research. First, the cognitive architecture underlying infants' assessment of cost remains to be explored. Our experiments suggest that infants are responding to an abstract notion of cost, rather than specific physical path features such as vertical motion (controlled for in Exp. 3), horizontal motion (controlled for in Exp. 1), or raw path length (controlled for in Exp. 2). We do not know, however, whether infants represent the abstract costs of actions by drawing on a concept of experienced effort or exertion within the domain of naïve psychology, or by leveraging an intuitive concept of force or work done (i.e. the integral of force applied over a path) from the domain of naïve physics (*34*, *35*), or perhaps both. Next, our experiments investigated only one class of goal states and target-directed actions, leaving open the breadth and generality of infants' intuitive psychology. In particular, cost can be defined in terms of work or effort to produce physical forces, but there are other kinds of costs: Agents could consider variables like the mental effort of planning (*36*, *37*) and the risks of choosing certain actions, neither of which involves applications of force. It is an open question whether these other variables trade off against reward in infants' intuitive psychology the way that physical work or effort does. Lastly, our studies do not speak to the origins of these abilities. Although 10-month-old infants cannot perform the actions from our experiments or communicate with others about them, their productive system for reasoning about costs and rewards may arise through their experiences observing the actions of other agents or performing actions within their repertoire, such as lifting their arms or balancing their bodies against the force of gravity. Alternatively, this system of intuitive psychology may guide infants' action understanding from the beginning. Testing these possibilities would address fundamental questions concerning the nature, origins, and interrelations between our intuitive psychology and intuitive physics.

However these questions are answered, the present study suggests that our propensity to understand the minds and actions of others in terms of abstract, general and interrelated concepts begins early. Before human infants learn to walk, leap and climb, they leverage mental models of agents and actions—forward models of how agents plan, and inverse models for working backwards from agents' actions to the causes inside their minds.

References and Notes:

1. F. Heider, M. Simmel, An experimental study of social behavior. *Am. J. Psychol.* **57**, 243–259 (1944).
2. D. C. Dennett, *The Intentional Stance* (The MIT Press, London, 1987).
3. G. Gergely, G. Csibra, Teleological reasoning in infancy: The naïve theory of rational action. *Trends Cogn. Sci.* **7**, 287–292 (2003).
4. C. L. Baker, R. Saxe, J. B. Tenenbaum, Action understanding as inverse planning. *Cognition*. **113**, 329–349 (2009).
5. R. Saxe, T. Tzelnic, S. Carey, Five-month-old infants know humans are solid, like inanimate objects. *Cognition*. **101**, B1–B8 (2006).
6. A. L. Woodward, Infants selectively encode the goal object of an actor's reach. *Cognition*. **69**, 1–34 (1998).
7. G. Gergely, Z. Nádasdy, G. Csibra, S. Bíró, Taking the intentional stance at 12 months of age. *Cognition*. **56**, 165–193 (1995).
8. K. H. Onishi, R. Baillargeon, Do 15-month-old infants understand false beliefs? *Science*. **308**, 255–8 (2005).
9. Y. Luo, R. Baillargeon, Do 12.5-month-old infants consider what objects others can see when interpreting their actions? *Cognition*. **105**, 489–512 (2007).
10. Y. Luo, S. C. Johnson, Recognizing the role of perception in action at 6 months. *Dev. Sci.* **12**, 142–149 (2009).
11. P. Muentener, S. Carey, Infants' causal representations of state change events. *Cogn. Psychol.* **61** (2010), pp. 63–86.
12. A. E. Skerry, S. E. Carey, E. S. Spelke, First-person action experience reveals sensitivity to action efficiency in prereaching infants. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 18728–33 (2013).
13. S. Liu, E. S. Spelke, Six-month-old infants expect agents to minimize the cost of their actions. *Cognition*. **160**, 35–42 (2017).
14. A. L. Woodward, Infants' grasp of others' intentions. *Curr. Dir. Psychol. Sci.* **18**, 53–57 (2009).
15. M. Paulus, Action mirroring and action understanding: an ideomotor and attentional account. *Psychol. Res.* **76**, 760–7 (2012).
16. C. M. Heyes, C. D. Frith, The cultural evolution of mind reading. *Science*. **344**, 1243091 (2014).
17. E. S. Spelke, K. D. Kinzler, Core knowledge. *Dev. Sci.* **10**, 89–96 (2007).
18. J. B. Tenenbaum, C. Kemp, T. L. Griffiths, N. D. Goodman, How to grow a mind: statistics, structure, and abstraction. *Science*. **331**, 1279–1285 (2011).
19. R. Baillargeon, R. M. Scott, L. Bian, Psychological Reasoning in Infancy. *Annu. Rev. Psychol.* **67**, 159–186 (2016).
20. D. Bernoullli, Exposition of a new theory on the measurement of risk. *Econometrica*. **22**, 23–36 (1954).

21. J. S. Mill, *Utilitarianism* (Oxford University, 1863).
22. J. Bentham, *An Introduction to the Principles of Morals and Legislation* (Clarendon Press, 1879).
23. J. Jara-Ettinger, H. Gweon, L. E. Schulz, J. B. Tenenbaum, The Naïve Utility Calculus: Computational Principles Underlying Commonsense Psychology. *Trends Cogn. Sci.* **20**, 589–604 (2016).
24. C. G. Lucas *et al.*, The child as econometrician: A rational model of preference understanding in children. *PLoS One*. **9** (2014), doi:10.1371/journal.pone.0092160.
25. A. Jern, C. Kemp, A decision network account of reasoning about other people's choices. *Cognition*. **142**, 12–38 (2015).
26. J. Jara-Ettinger, H. Gweon, J. B. Tenenbaum, L. E. Schulz, Children's understanding of the costs and rewards underlying rational action. *Cognition*. **140**, 14–23 (2015).
27. Y. Luo, R. Baillargeon, Can a self-propelled box have a goal? - Psychological reasoning in 5-month-old infants. *Psychol. Sci.* **16**, 601–608 (2005).
28. G. Csibra, G. Gergely, S. Bíró, O. Koós, M. Brockbank, Goal attribution without agency cues: The perception of "pure reason" in infancy. *Cognition*. **72**, 237–267 (1999).
29. Detailed methods and analyses from the experiments reported in this paper are available in the Supplemental Materials.
30. Prior to conducting Exp. 1, we were unsure whether infants would express their expectations by longer or shorter looking at the unexpected event, and so we pre-specified a two-tailed alpha. After Exp. 1, however, we had strong reason to believe that infants would look longer at the test actions that they found less probable, and so we pre-specified (for Exp. 2) and pre-registered (for Exp. 3) one-tailed tests and report these numbers in the text. However, results of all three experiments were also statistically significant by two-tailed tests, with p=.031 and p=.036 for Exp. 2 and 3, respectively.
31. A. E. Skerry, E. S. Spelke, Preverbal infants identify emotional reactions that are incongruent with goal outcomes. *Cognition*. **130**, 204–216 (2014).
32. G. Csibra, M. Hernik, O. Mascaro, D. Tatone, M. Lengyel, Statistical treatment of looking-time data. *Dev. Psychol.* **52**, 521–536 (2016).
33. I. K. Kim, E. S. Spelke, Infants' sensitivity to effects of gravity on visible object motion. *J. Exp. Psychol. Hum. Percept. Perform.* **18**, 385–393 (1992).
34. E. Téglás *et al.*, Pure reasoning in 12-month-old infants as probabilistic inference. *Science (80-.).* **332**, 1054–9 (2011).
35. T. D. Ullman, E. Spelke, P. Battaglia, J. B. Tenenbaum, Mind games: Game engines as an architecture for intuitive physics. *Trends Cogn. Sci.* **xx**, 1–17 (2017).
36. A. Shenhav *et al.*, Toward a Rational and Mechanistic Account of Mental Effort. *Annu. Rev. Neurosci.* **40**, 99–124 (2017).
37. W. Kool, S. J. Gershman, F. A. Cushman, Cost-Benefit Arbitration Between Multiple Reinforcement-Learning Systems. *Psychol. Sci.* **28**, 1321–1333 (2017).
38. B. Foundation, Blender (2016), (available at blender.org/download).
39. J. Pinto, XHAB64 (1995).
40. R. M. Casstevens, jHab: Java Habituation Software (2007).
41. R. D. C. Team, R: A language and environment for statistical computing (2015), (available at https://www.r-project.org/).
42. D. Bates, M. Mächler, B. Bolker, S. Walker, Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* **67** (2015), doi:10.18637/jss.v067.i01.

43. R. Nieuwenhuis, M. te Grotenhuis, B. Pelzer, Influence.ME: Tools for detecting influential data in mixed effects models. *R J.* **4**, 38–47 (2012).
44. H. Wickham, *ggplot2: Elegant graphics for data analysis* (Springer-Verlag, 2009).
45. G. Gergely, H. Bekkering, I. Király, Rational imitation in preverbal infants. *Nature*. **415**, 755 (2002).
46. A. N. Meltzoff, Understanding the intentions of others : Re- enactment of intended acts by 18-month-old children. *Dev. Psychol.* **31**, 838–850 (1995).
47. Y. Mou, J. M. Province, Y. Luo, Can infants make transitive inferences? *Cogn. Psychol.* **68**, 98–112 (2014).
48. T. Kushnir, F. Xu, H. M. Wellman, Young children use statistical sampling to infer the preferences of other people. *Psychol. Sci. a J. Am. Psychol. Soc. / APS*. **21**, 1134–1140 (2010).
49. G. Csibra, S. Bíró, O. Koós, G. Gergely, One-year-old infants use teleological representations of actions productively. *Cogn. Sci.* **27**, 111–133 (2003).
50. A. Jern, C. Kemp, in *Proceedings of the 36th {A}nnual {C}onference of the {C}ognitive {S}cience {S}ociety*, P. Bello, M. Guarini, M. McShane, B. Scassellati, Eds. (2014).
51. J. . Hamlin, T. Ullman, J. Tenenbaum, N. Goodman, C. Baker, The mentalistic basis of core social cognition: experiments in preverbal infants and a computational model. *Dev. Sci.* **16**, 209–226 (2013).
52. S. Russell, P. Norvig, *Artificial Intelligence: A Modern Approach, Third edition* (2014).
53. R. Sutton, A. Barto, Reinforcement Learning: An Introduction. *MIT Press. Cambridge, Massachusetts* (1998).
54. N. D. Goodman, V. K. Mansinghka, D. M. Roy, K. Bonawitz, J. B. Tenenbaum, Church: a language for generative models. *Uncertain. Artif. Intell.* (2008).

Acknowledgements: All authors wrote the paper, S.L. designed the experiments and collected and analyzed the data under the direction of E.S., and T.U. built the computational models under the direction of J.T. This material is based upon work supported by the Center for Brains, Minds and Machines (CBMM), funded by National Science Foundation STC award CCF-1231216, and by a National Science Foundation Graduate Research Fellowship under grant DGE-1144152. We thank the families who volunteered to participate, and A. Aguirre, R. Guzman, C. Kerwin, N. Kalra, and H. Tarr for research assistance. All data, analyses, and materials from the experiments and models reported in this paper are archived on the Open Science Framework at https://osf.io/mykph/.

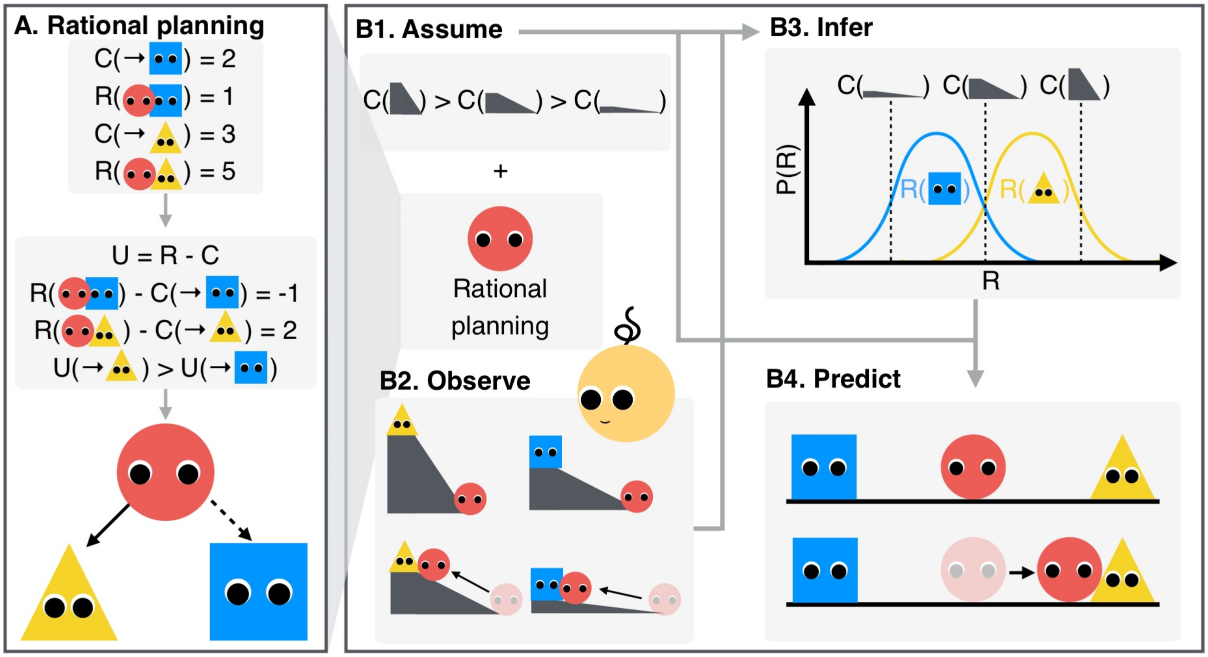

Fig. 1. A schematic of our computational model. The forward direction (A) defines the agent as a rational planner that calculates the utilities of different actions from their respective costs and rewards, and then selects an action stochastically in proportion to its utility. In this case, the overall utility for approaching Triangle is higher than for approaching Square, so the central agent (Circle) will likely choose Triangle over Square. An observer (B1) assuming this model and some priors over the costs of different actions, can (B2) observe a series of actions and then (B3) infer a posterior distribution over the hidden values of an agent's costs and rewards given its actions. These posteriors can then be used to (B4) predict the actions of the agent in a new situation, in which the same goal states can be reached by different actions.

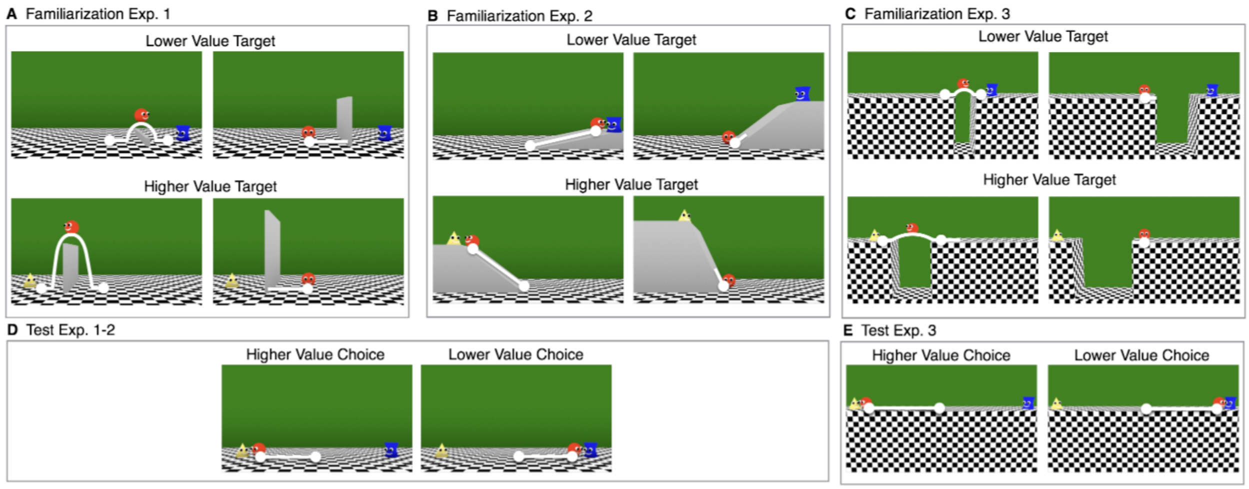

Fig. 2. Structure of Experiments 1-3. During familiarization (A-C), the central agent (Circle) accepted a low and refused a medium cost for the lower value target (in this case, Square), and accepted a medium and refused a high cost for the higher value target (Triangle). Other than the sizes of the barriers, ramps, and trenches, and the consequent trajectories of motion, the pairs of events displaying approach or refusal of approach to the two targets were identical. At test (D-E), the agent looked at each of the two targets and chose either the lower or higher value target. White circles indicate start- and end-points of action, and white lines indicate trajectories.

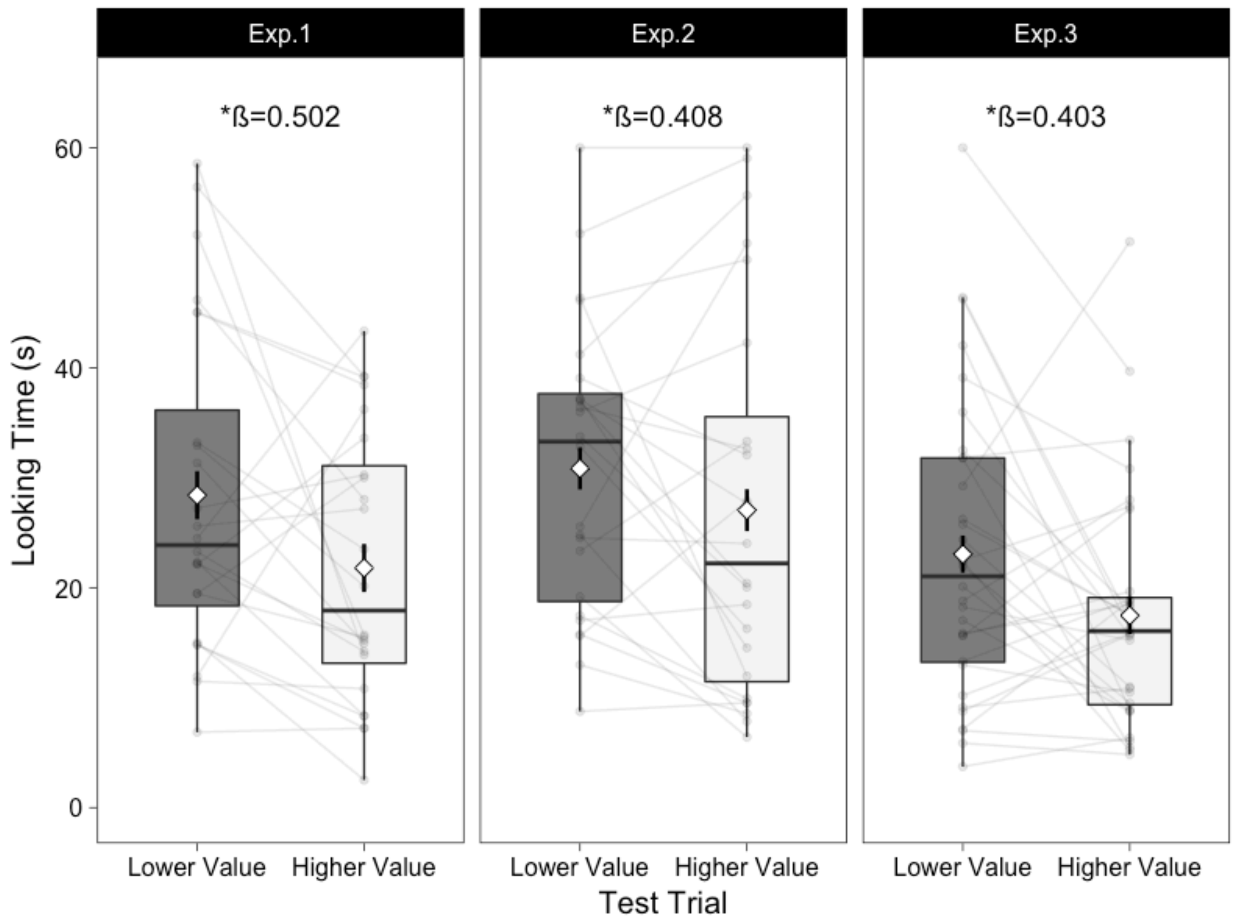

Fig. 3. Boxplots of average looking time towards the higher and lower value choice during test in Experiments 1-3. White diamonds indicate means, with error bars indicating within-subjects standard errors. Horizontal lines indicate medians, boxes indicate middle quartiles, and whiskers indicate points within 1.5 times the interquartile range from the upper and lower edges of the middle quartiles. Light grey points connected across boxes indicate looking times from individual participants. Beta coefficients indicate effect sizes in standard deviations, and asterisks indicate significance relative to pre-specified (Experiments 1-2) and pre-registered (Experiment 3) alphas (*<.05). See text and SM for statistical analyses.

Science AAAS

Supplementary Materials for

Ten-month-old Infants Infer the Value of Goals from the Costs of Actions

Shari Liu, Tomer D. Ullman, Joshua B. Tenenbaum, & Elizabeth S. Spelke

correspondence to: shariliu01@g.harvard.edu

This PDF file includes:

Other Supplementary Materials for this manuscript includes the following:

Movies S1-S6

Materials and Methods:

Experiment 1

Methods

Participants. Our final sample included 24 healthy, full-term infants (15 female, M_{age}=9.95m, range=9.43-10.53). Eight more infants were tested, excluded, and replaced (1 for fussiness that prevented study completion, 1 for technical failure, 5 for inattentiveness during test events, and 1 for experimental error). Sample size and exclusion criteria were fixed prior to the start of data collection, and decisions concerning exclusions were made by researchers unaware of the order of the test events viewed by the infant (and therefore blind to the data that the infant provided). All participants were recruited from the greater Boston area and tested at the Laboratory for Developmental Studies at Harvard University with parental informed consent. Families received a small thank-you gift (e.g. a t-shirt or toy) for participating. All study protocols were approved by the Committee on the Use of Human Subjects at Harvard University.

Materials and Design. All animated events were created in Blender (*38*), synchronized with a custom audio track in iMovie, and presented using Keynote on a 101.6cm by 132.1cm LCD projector screen. Two speakers flanking the screen played all stimuli-related sounds. Infants' looking time data were coded online using Xhab64 (*39*) software and offline using jHab (*40*).

The experiment consisted of 3 pairs of familiarization trials, 1 pre-test trial, and 2 pairs of test trials. All familiarization (Movie S1) and test trials (Movie S2) began with an attention-getting animation and sound (3.0s), followed by looped sequences of events. *Familiarization* sequences consisted of 4 videos (8s or 8.9s; see below) and *test* sequences consisted of a single video (5.2s), each played on a loop with black screens (0.5s) interspersed between the events. Events featured three agents with eyes and a smiling mouth: a central red spherical agent, and two target agents (a cone and cylinder). All familiarization events featured the agent paired with one of the two targets at a time; all test events featured both targets together. The timing of all actions was held constant within familiarization and test blocks. The location of the targets was constant across participants, and the identity of the higher value target was counterbalanced across participants.

During *familiarization* (Movie S1), the agent responded to the call of one of the targets by accepting or refusing to jump over a small (1 units tall), medium (6 units), or large (10 units) barrier that fell with a thud between the agent and target. In events where the agent accepted the cost (8.9s), it looked up at the barrier, made a positive "Mmmm!" sound, and leapt over it to reach the target. The agent always acted efficiently (Gergely et al., 1995) adapting the height of its jump to the height of the barrier; all jumps were accompanied with a popping sound. In events in which the agent refused the cost (8.0s), it looked up at the barrier, made a mildly negative "Hmmmm…" sound, and backed away, returning to the center of the screen. Each block of familiarization events consisted of 4 videos, wherein the agent accepted a small cost and refused a medium cost for one target, and accepted a medium cost and refused a large cost for the other target. (For convenience, we describe events where the agent does not take a costly action as "refusing" or "declining" to take an action, which is how these events appear to adults. But for the purposes of our hypothesis about infants' action understanding, and for our computational model, what matters in these cases is only that the agent is presented with a costly action, considers it and does not take it.) Each familiarization trial consisted of looped blocks either in the above order (small, medium, medium, large) or the opposite order. Across the 6 familiarization trials, both orders were presented 3 times in an ABABAB pattern. The identity of the higher value target and the first familiarization block were counterbalanced across participants.

A single non-looped *pre-test* event following familiarization featured a still image of the two targets without the agent. During *test* (Movie S4), the agent reappeared, rotated left then right while saying "Hmmm…", and then approached one of the targets: either the higher- or lower value agent. The same sound accompanied each approach. Across 2 pairs of test trials presented in alternation, the agent approached the Triangle target twice and the Square target twice. The first test trial (higher- or lower value approach) was counterbalanced across participants.

Procedure. Infants were seated on their caregivers' laps approximately 1.5m away from the screen. Caregivers were instructed to keep their eyes closed and to refrain from interacting with their infants throughout the experiment, and were monitored for compliance.

After calibrating infants to the screen using a toy, the researcher began the experiment. The researcher had access to a video feed of the infants' faces, a computer screen indicating the current trial, and a third screen indicating when to conclude a trial. The researcher ran the experiment and coded looking time online while unaware of the order of events (and therefore unaware of the infant's differential reactions to the displays), but could determine the start of each trial as well as the timing of actions (e.g. when the central agent approached one of the two targets) based on auditory cues.

Across both the familiarization and test phases of the experiment, the researcher began coding a trial immediately following the attention getter, and concluded the trial once the infant had attended to the screen for 60s cumulatively or looked away for 2s consecutively. During pre-test, the researcher waited until the infant looked towards each target agent at least once, and then began the test trials. These criteria were fixed prior to the start of data collection.

Coding and analysis. Videos of all test sessions were coded offline by observers who were unaware of the order of events that infants viewed, using the same thresholds as online coding, and reviewed for predetermined exclusion criteria (fussiness that prevented study completion, online coding error, experimenter error, technical failure, and parental interference). Further, if infants were determined to have missed a critical part of the test trial (i.e., looked away for the entirety of the approach at test), then that test pair was marked and excluded from subsequent analyses. If infants missed a critical portion of both test pairs, then they were dropped from the sample and replaced. To assess the reliability of the offline-coded data, 100% of the test trials were recoded independently by an additional researcher who was unaware of test pair order. The two coders agreed on trial cutoffs for 95% of the test trials, and the intraclass correlation (ICC) between the two raters was 0.994, 95% CI [0.991, 0.996]. Thus, the primary offline coding data were used in our analyses.

The primary dependent measure was log-transformed looking time (*32*) but plots and descriptive statistics feature raw values for ease of interpretation.

All models were fit in R (*41*). Linear mixed models were fit using the lme4 package (*42*). Detection of influential observations was conducted using the influence.ME package (*43*). Plots were produced using the ggplot2 package (*44*). To explicitly take into account repeated measures, all mixed models included participant identity as a random intercept. Three classes of models were fit: (1) null models, featuring participant identity as the only predictor, (2) hypothesis-driven models, which included additional manipulated factor(s), and (3) exploratory models, which included additional non-hypothesis driven factors. We leveraged likelihood ratio tests (LRTs) and the Aikake Information Crtierion (AIC) to evaluate model fit and parsimony..

All degrees of freedom from mixed effects models were calculated using the Satterthwaite approximation method. Bracketed values indicate 95% confidence intervals.

We predicted that if infants can infer value from effort and use this information to predict the agents' subsequent actions, then they will differentiate between the more probable outcome (when the agent chooses the higher value goal) and the less probable one (when the agent chooses the lower value goal) by showing a looking preference in either direction.

Results

Hypothesis-driven Results. A model with the single predictor of test trial (higher- or lower value) revealed that infants looked longer at the lower value action (M=28.41s, SD=14.85) than the higher value action (M=21.79s, SD=12.29), B=0.327, SE=0.130, β=0.502, $t(24)$=2.523, p=.019, [0.062, 0.591]. This model outperformed a null model by a LRT, $X^2(1)$=5.648, p=.017. A leverage analysis using Cook's Distance revealed 1 influential observation in this model. Removal of this case produced an inferentially equivalent result, B=0.263, SE=0.119, β=0.454, $t(23)$=2.221, p=.037, [0.021, 0.506].

Exploratory Results. A first exploratory analysis tested for an effect of test pair order by including an interactive effect between test trial presentation order and trial type. Infants discriminated between the test events to a similar degree regardless of whether they were assigned to watch the lower- or higher value event first, B=0.212, SE=0.255, β=0.325, $t(24)$=0.829, p=0.415, [-0.310, 0.733]. Removing one influential case produced an inferentially equivalent result, B=0.354, SE=0.226, β=0.610, $t(23)$=1.569, p=0.130, [-0.107, 0.816]. A second model was fit with the additive effects of test pair order and test trial type, summed looking time during familiarization, sex, the identity of the higher value character, and the order of the first block of familiarization. Infants' looking preferences were not predicted by any of the exploratory factors, with all CIs containing 0, ps>0.1. Removal of one influential case produced inferentially equivalent results. The best model out of the above 4 was the hypothesis-driven model (AIC=92.500).

Experiment 2

Methods

Participants. Our final sample included 24 healthy, full-term infants (15 female, M_{age}=9.88m, range=9.47-10.43). Ten more infants were tested, excluded, and replaced (1 for fussiness that prevented study completion, 1 for technical failure, 2 for online coding errors, 2 for parental interference, and 4 for inattentiveness during test events).

Materials, design, and procedure. All materials, design, and procedure were identical to those from Exp. 1 except that during familiarization (Movie S2), each target appeared at the top of a ramp and the agent either refused or accepted to climb the ramp to reach it. To manipulate action cost while controlling for path length, the angle of the ramps varied (11.51°, 39.26°, and 64.09°) such that the agent began 10 Blender units away from the top of the ramp. When the agent accepted the cost (6.8s), the agent moved up the ramp once, slid back down, and then moved all the way up to the target. When the agent refused the cost (5.5s), the agent moved up the ramp once, slid back down, and then turned away from the target, back towards the center for the screen.

Coding and analysis. All coding and analysis procedures were identical to those from Exp. 1. To assess the reliability of the offline-coded data, 100% of the test events were recoded independently by an additional researcher who was unaware of test pair order. The two coders agreed on trial cutoffs for 98% of the test trials, and the intraclass correlation (ICC) between the two raters was 0.978, 95% CI [0.967, 0.985]. Thus, the primary offline coding data were used in our analyses. Because of our strong directional prediction, all reported p-values in hypothesis-driven results in Exp. 2 are one-tailed. All other reported *p*-values are two-tailed.

Results

Hypothesis-driven Results. As in Exp. 1, infants looked longer at the lower value action (M=30.84, SD=13.79) than the higher value action (M=27.05s, SD=17.55), *B*=0.250, *SE*=0.109, β=0.408, *t*(24)=2.294, *p*=.015, [0.028, 0.472]. This model outperformed a null model by LRT, $X^2(1)=4.760$, *p*=.029. No influential cases were detected.

Exploratory Results. An exploratory model testing explicitly for presentation order revealed that infants differentiated between the test events differently depending on whether they were assigned to watch the lower value versus the higher value approach first, *B*=0.521, *SE*=0.190, β=0.851, *t*(24)=2.741, *p*=.011, [0.133, 0.909]. No influential cases were detected. which revealed that whereas infants who saw the lower value test event first looked longer at the lower value (M=28.70s, SD=11.12) versus higher value (M=20.01s, SD=14.83) test trials, *B*=0.511, *SE*=0.134, *t*(24)=3.797, *p*=.001, [0.233, 0.788], infants who saw the higher value choice did not differentiate between the lower- (M=32.97s, SD=15.39) and higher value test events (M=34.08, SD=17.78), *B*=-0.011, *SE*=0.134, *t*(24)=0.079, *p*=.938, [-0.288, 0.267]. An additional model testing for effects of summed attention during test, sex, the identity of the higher value target, and the first familiarization loop revealed that no further effects other than one of first familiarization loop, where infants assigned to watch a sequence of low to high cost first looked longer overall at test, *B*=0.440, *SE*=0.185, β=0.851, *t*(24)=2.374, *p*=.023, [0.062, 0.819]. Removal of 3 influential observations from this model yielded inferentially equivalent results. The best model of the four reported was the simpler exploratory model with the single interactive effect (AIC=78.219).

Experiment 3

Methods

Participants. Our final sample included 32 healthy, full-term infants (15 female, M_{age}=10.03m, range=9.57-10.50). Six more infants were tested, excluded, and replaced (1 for fussiness that prevented study completion, 3 for online coding errors, 1 for parental interference, and 1 for inattentiveness during test events). Sample size was determined from a simulation power analysis over data from Exp 1-2. The design, procedure, and analyses of this experiment were pre-registered via the Open Science Framework (https://osf.io/k7yjt/).

Materials, design, and procedure. All materials, design, and procedure were identical to those from Exp. 1-2 except as follows. During familiarization (Movie S3), each target appeared at the far end of a trench (15 units deep), and the agent either accepted this cost by jumping across it or refused to jump. To manipulate action cost while controlling for movement against gravity, the width of the trench varied (5, 10, and 15 units) such that the agent began 13, 18, or 23 units away from the target. When the agent accepted the cost (8.4s for the small cost, 9.3s for the medium), the agent looked down at the bottom of the trench, looked at its target, and backed up and leapt across the trench. When the agent refused the cost (7.2s), it looked down at the bottom of the trench, looked at its target, and backed up the turned away, back towards the center for the screen. The agent backed away a longer distance and accelerated more when accepting the medium versus the small cost. During test (Movie S5), infants saw the same choice events as in Exp. 1-2, but the events took place on a platform at the same height used for familiarization. Prior to familiariazation, infants watched an additional video (Movie S6) , in which a ball rolled across the platform and off the small, medium, and wide trench (5.8s, 5.7s, and 5.0s), identical in width and depth to those during familiarization and situated at the center of the screen. During each segment of the video, the ball traveled in a parabolic trajectory after it rolled off the edge and shattered against the far wall of the trench. The location of impact depended on the width of the trench, such that the ball shattered at a lower point on the far wall for wider trenches. The side of the screen that the ball emerged from (left vs. right) was consistent across all videos within participants and was counterbalanced across participants.

Coding and analysis. All coding and analysis procedures were identical to those from Exp. 1-2. To assess the reliability of the offline-coded data, 100% of the test events were recoded independently by an additional researcher who was unaware of test pair order. The two coders agreed on trial cutoffs for 96% of the test trials, and the intraclass correlation (ICC) between the two raters was 0.995, 95% CI [0.993, 0.996]. Thus, the primary offline coding data were used in our analyses. All reported p-values in hypothesis-driven results in Exp. 3 are one-tailed. All other reported *p*-values are two-tailed.

Results

Hypothesis-driven Results. As in Exp. 1-2, infants looked longer at the lower value action (M=23.05s, SD=13.58) than the higher value action (M=17.47, SD=10.69), *B*=0.260, *SE*=0.119, β=0.403, *t*(32)=2.185, *p*=.018, [0.020, 0.501]. This model outperformed a null model by LRT, $X^2(1)=4.452$, *p*=.035. Removal of 1 influential case produced an inferentially equivalent result, *B*=0.286, *SE*=0.120, β=0.466, *t*(31)=2.379, *p*=.011, [0.043, 0.529].

Exploratory Results. An exploratory model testing explicitly for presentation order revealed that like in Exp. 2, infants differentiated between the test events differently depending on whether they were assigned to watch the lower value versus the higher value approach first, *B*=0.770, *SE*=0.195, β=1.194, *t*(32)=3.941, *p*<.001, [0.357, 1.165]. We detected 1 influential observation in this model and removed it from subsequent pairwise comparisons, which revealed that whereas infants who saw the lower value test event first looked longer at the lower value (M=27.95s, SD=14.11) versus higher value (M=15.05s, SD=9.06) test trials, *B*=0.645, *SE*=0.117, *t*(31)=5.519, *p*<.001, [0.407, 0.883], infants who saw the higher value choice showed a weak preference for the higher-value (M=20.62, SD=11.93) relative to the lower-value (M=16.56s, SD=9.84) test events, *B*=-0.236, *SE*=0.121, *t*(31)=-1.959, *p*=0.059, [-0.483, 0.010]. An additional model testing for effects of summed attention during test, sex, the identity of the higher value target, and the first familiarization loop revealed that no further effects other than one of total attention in seconds during familiarization, where infants who were more attentive during familiarization looked marginally longer overall at test, *B*=0.002, *SE*=0.001, β=0.261, *t*(32)=1.948, *p*=.060, [0.000, 0.005]. Removal of two influential cases from this model yielded an inferentially equivalent result, and revealed an additional finding that infants randomly assigned to conditions where the higher value target was on the right looked longer overall at test, *B*=0.368, *SE*=0.149, β=0.610 *t*(30)=2.464, *p*=.020, [0.066, 0.670] The best model of the four reported was the simpler exploratory model with the single interactive effect (AIC=114.24).

Results Across Exp 1-3. As reported in the main text, across all experiments, infants looked longer at the lower value action (M=26.99s, SD=14.13) than the higher value action (M=21.64s, SD=13.94), *B*=0.277, *SE*=0.070, β=0.424, *t*(80)=3.975, *p*<.001, one-tailed, [0.139, 0.415], mixed effects model with random intercepts for participant and experiment. Removal of one influential case produced an inferentially equivalent result, *B*=0.258, *SE*=0.068, β=0.406, *t*(79)=3.799, *p*<.001, one-tailed, [0.123, 0.393]. To test explicitly for differences in responses to test events across experiments, an additional model with an interactive effect between experiment and test event was fit and revealed no differences in looking preference across experiments, and removal of one influential observation produced an inferentially equivalent result.

Two further models tested explicitly for the effect of presentation order on attention to the lower- vs. higher value test events, and additional predictors of summed attention during test, sex, the identity of the higher value target, and the first familiarization loop. These models revealed a robust effect of presentation order, B=0.528, SE=0.126, β=0.808, $t(80)$=4.179, p<.001, [0.277, 0.778], and for an effect of attention during familiarization, B=0.002, SE=0.001, β=0.209, $t(79.65)$=2.221, p=.030, [0.000, 0.006]. mixed effects model with random intercepts for participant and experiment. Removal of influential cases produced inferentially equivalent results. A comparison of model fit and parsimony revealed that the model including the single interactive effect provided the best description of the data (AIC=280.90).

Fifty out of the 79 infants who demonstrated a preference in either direction (63.29%; 1 infant in Experiment 2 looked for 60 seconds on all test trials) looked longer on average at the lower value test events, p=.033, 95% CI [0.269, 0.490], exact binomial test. The proportion of infants who looked in the predicted direction did not differ across Experiments 1 (17/24, 70.83%), Experiment 2 (14/23, 60.87%), and Experiment 16 (19/32, 59.38%), $X^2(2)$=1.022, p=.600. See Figure S3. Further non-parametric analyses on raw looking times in seconds supported the finding that infants looked longer at the lower value test events across all experiments, [2.165, 8.185], V=2241, p=.001, Wilcoxon signed rank test, and [0.720, 8.235], bootstrapped median difference in looking times across test events with 10,000 samples.

To compare attention during familiarization across the experiments, we fit a linear model including summed looking time in seconds as the dependent variable and experiment (1, 2, or 3) as a predictor. We found that infants looked longer during familiarization in Exp. 1 (M=261.96s, SD=61.29) than in Exp. 2 (M=216.50, SD=63.49), B=45.47, SE=19.22, β=0.653, $t(77)$=2.366, p=.021, [7.196, 83.74], and in Exp. 3 (M=209.70, SD=72.33), B=52.26, SE=17.98, β=0.751, $t(77)$=2.907, p=.005, [16.460, 88.056]. See Figure S2.

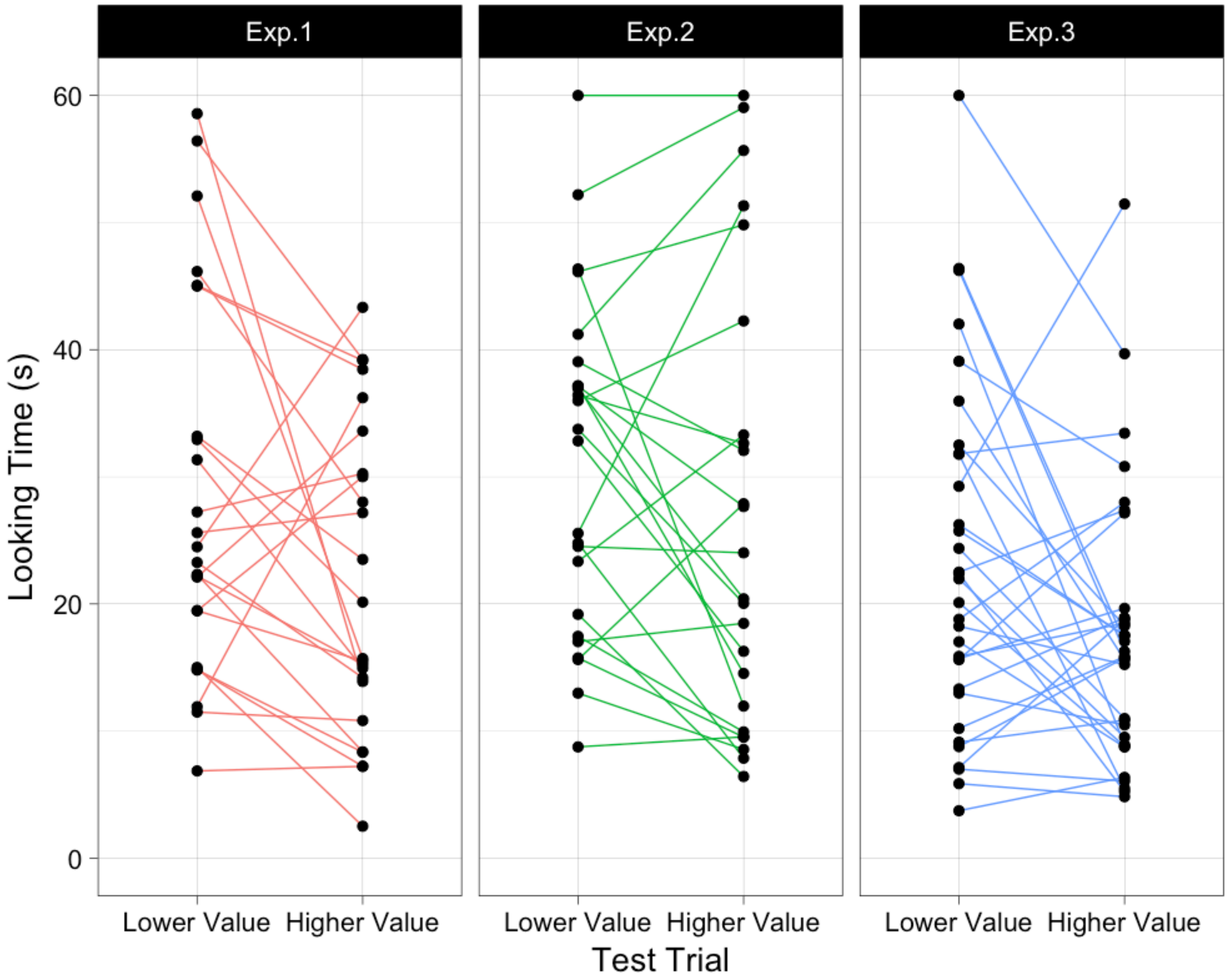

Fig S1. Looking times averaged across two test pairs towards the lower value and higher value test event from all participants in Experiments 1-3.

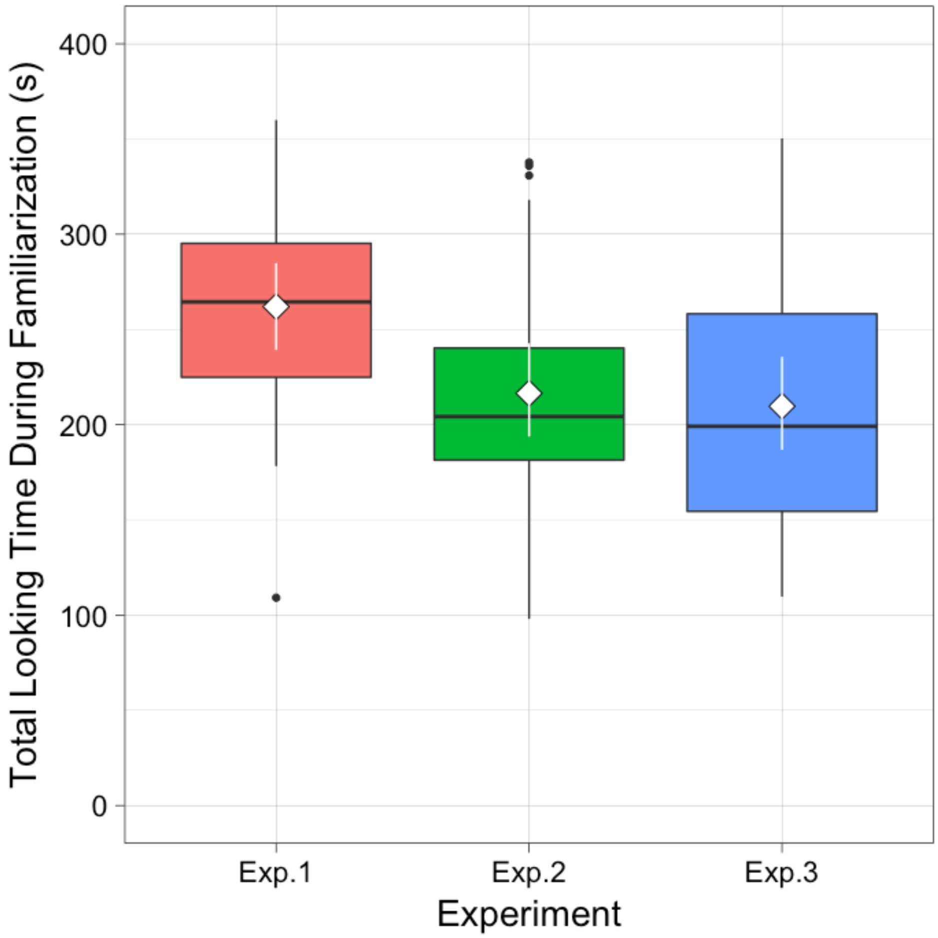

Fig S2. Total looking time in seconds during familiarization across Exp 1-3. Boxes indicate middle quartiles, vertical lines indicate points within 1.5 times the interquartile range from the 25th and 75th percentiles, and horizontal lines indicate medians. Means and 95% confidence intervals are plotted in white.

Movie S1. Sample familiarization event for Experiment 1.
Movie S2. Sample familiarization event for Experiment 2.
Movie S3. Sample familiarization event for Experiment 3.
Movie S4. Sample test event for Experiments 1-2.
Movie S5. Sample test event for Experiment 3.
Movie S6. Pre-familiarization event for Experiment 3.

Computational Modeling Details

We constructed several alternative models for predicting the actions of a central red agent, given previous stimuli showing its behavior. We evaluated the models both as accounts of our present three experiments testing infants' ability to integrate costs and rewards in action understanding, and as accounts of five previous experiments that tested infants' abilities to make inferences about rewards (*6*) or costs (*7*, *13*) individually. As our main proposed model, we constructed a probabilistic program for inferring rewards over possible goals through a Bayesian computation, based on the Naïve Utility Calculus (NUC) and inverse planning. The cost-function of this model was either based on distance (NUC-distance), or physical effort through the application of force (NUC-force). We next constructed three alternative models that were lesions of various aspects of this model. Only one of our two NUC models (NUC Force; Fig. S3B) accounts for all of these findings, including Experiment 2 in which infants saw the agent cover the same total distance at the same speed but on trajectories requiring different amounts of physical effort due to the force of gravity. In contrast, two models that do not integrate costs and rewards (Costs Only, Fig. S3E and Rewards Only, Fig. S3F), as well as two integrative but simpler models using only perceptual cues for value (Counts, Fig. S3D) or distance-based representations of cost (NUC Distance, Fig. S3C) instead of abstract effort-reward tradeoffs, account for just a subset of this body of empirical findings.

The eight experiments shown in Fig. S3 only begin to test the explanatory scope of our computational framework. This approach naturally extends to explain other findings (*23*), including infants' rational imitation of other agents' goal-directed actions (*45*, *46*), understanding of the transitivity of agents' preferences (*47*), inferences of agents' preferences from their non-random selections of objects (*48*), and inferences about the existence and location of an occluded object from the cost of an agent's action (*49*).

In the rest of this section we describe the full NUC model, then the alternative proposals. We then compare the predictions of the different models to the empirical results of eight different experiments with young children that all relate to reasoning about cost, reward and efficient behavior.

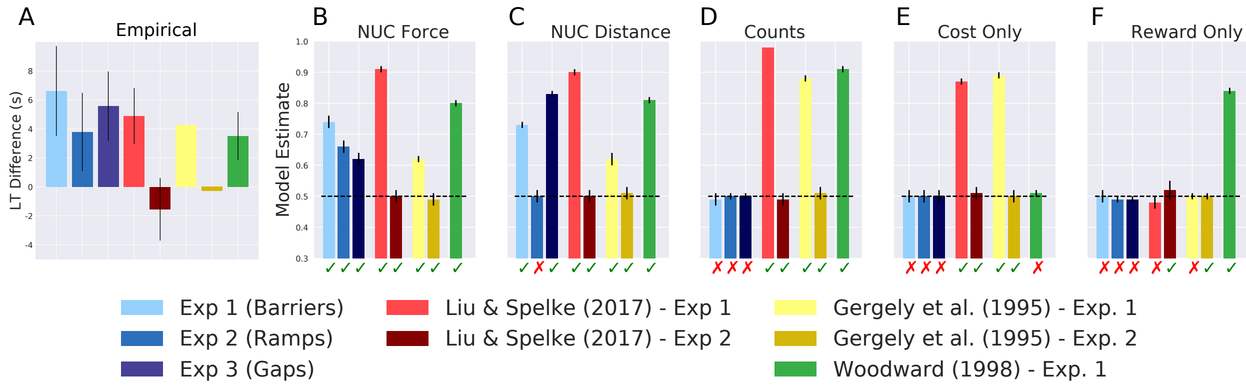

Fig S3. Comparison of model predictions (B-F) to the empirical results (A) of Experiments 1-3 in this paper, as well as additional experiments on infants' sensitivity to costs and rewards (*6*, *7*, *13*). Model estimates were generated from 5,000 MH samples with a 10-step interval and burn-in of 1,000. Error bars indicate 95% CIs (model estimates) and standard errors (where available from empirical findings). Check marks and crosses respectively indicate consistency or inconsistency with empirical findings.

Decision-making framework

Following the logic of "Bayesian Theory of Mind" (*4*, *23*, *50*, *51*), the primary model assumes agents have a planning procedure for generating actions that maximize their utility or expected utility, and that agents then use observed actions to invert this procedure to reason about the utilities and constraints of the planning agents.

Following a standard decision-making framework for rational planning (e.g. *48*) we assume that a rational planning agent:

(i) Divides the world into possible states *S*, such that for each state $s \in S$ there exists a set of possible actions A_s.

(ii) Has access to a transition function *T* such that:

$$T(S, A_s) \rightarrow P(S'), \quad (1)$$

where *P(S')* is the probability distribution over the states of the worlds S' that can result from taking action A_s in state S.

(iii) Has a utility function *U* such that:

$$U(A, S) = Reward(S) - Cost(A), \quad (2)$$

where *Reward* and *Cost* are functions that map from states and actions to real numbers. This separation of the utility into independent components follows recent work by Jara-Ettinger et al. (2016), who showed that both components can be separate targets of inference, and used as explanatory variables by young children across different scenarios.

(iv) Is guided by a decision-function $D(S, A_s) \rightarrow P(A)$ that selects a given action in a state, in order to maximize the expected utility U. For our simple scenarios, we assume a soft-max decision function (*53*):

$$P(\mathrm{A}_s = \mathrm{a} | U, S = s) = \frac{e^{\beta\, U(a, T(s,a))}}{\sum_j e^{\beta\, U(a, T(s,a))}}, \quad (3)$$

where β is a noise-parameter (inverse-temperature) adjusting the agent's determinism. As $\beta \rightarrow 0$, the agent will behave in a more random fashion. As $\beta \rightarrow \infty$, the agent will tend to use greedy action selection. Intermediate values of β (around 1) approximate probability matching behavior. Thus, we place an equal prior probability on β taking one of two values, β = 3 (representing a near-optimal rational agent) and β = 0 (representing a random, non-rational agent). The exact value of the parameter for a rational agent is not important for the qualitative pattern of results discussed in the following sections.

The Specific Decision-making Environment

States: In our studies infants saw the agent making a single choice between staying put and moving towards one of two possible goal agents. The full set of possible states in Experiments 1-3 are $s \in \{Start, Target_A, Target_B\}$, where *Target$_i$* can be the Square or Triangle goal agent. Not all states are available in every situation, depending on the stimuli (for example, if the stimuli shows the decision making agent near a ramp with the Square goal agent at the top, the sub-set of states for this environment does not include reaching the Triangle goal agent).

Actions: In Experiment 1 the actions in any situation were a subset of $\{Nothing, Left, Right,\ Jump\ tall,\ Jump\ medium,\ Jump\ short\}$.

In Experiment 2 the actions were a subset of $\{Nothing, Left, Right, Climb\ steep, Climb\ moderate, Climb\ shallow\}$.

In Experiment 3 the actions were a subset of $\{Nothing, Left, Right, Jump\ wide, Jump\ medium, Jump\ narrow\}$.

The set of possible actions and states was similarly altered for the other 5 experiments we considered, that were not part of the current study.

Cost functions: We do not know the specific cost that infants believe the agent incurs for jumping over the barriers or climbing the inclines, but we assume that the cost for jumping/climbing is greater than staying put, and that cost generally scales with distance. There are at least two potential models for how to take this distance into account when calculating cost. The first model (NUC-distance) ignores physical forces and effort, and uses only total distance traveled between states as the input to the cost function:

$$Distance_{cost}(S_1 \rightarrow S_2) \propto \int_C a\, ds, \quad (4)$$

where C is the trajectory from S_1 to S_2, and a is some constant factor. We considered each Blender unit of distance to be one unit of cost for the model.

The second model (Force) considers the physical work required to get from one state to the next. Specifically, we consider the work done in the presence of a conservative force field (gravity):

$$Work_{cost}(S_1 \rightarrow S_2) = \int_C F \cdot ds = \int_{t1}^{t2} F \cdot dt. \quad (5)$$

For Experiments 1&2, we considered the force of over-coming friction as well as gravity, meaning $F = f \cdot g \cdot \cos(\alpha) + g \cdot \sin(\alpha)$, where α is the angle opposite the vertical part of the incline, and f is the friction coefficient. The mass (m) was set arbitrarily to 1 and so does not appear. Without the friction component the non-accelerated movement of an agent on a non-inclined plane is effortless. We set the friction coefficient arbitrarily to $f = 0.5$, similar to wood sliding on wood or metal sliding on wood, though we note that the exact choice of f in the range 0 to 1 does not affect the qualitative results. For experiment 3 we considered in addition the impulse forces that accelerate and decelerate the agent. This instantaneous change in velocity can be related to the work done as a change in kinetic energy, meaning $W = \Delta E_k = \frac{1}{2}(V_1^2 - V_2^2)$, where V_1 and V_2 are the velocities before and after the application of the impulse. Both models would predict similar cost differences for Experiments 1 and 3, but they would diverge with regards to Experiment 2. $Distance_{cost}$ of the steep and shallow inclines is the same, while $Work_{cost}$ is sensitive to the incline of the ramp.

Rewards: We do not make a-priori assumptions about the reward associated with reaching the goal agents, except that they can potentially exceed the range of the costs:

$$\text{Reward(Target A)} \sim \text{Uniform}(0, \text{Reward}_{max}),$$
$$\text{Reward(Target B)} \sim \text{Uniform}(0, \text{Reward}_{max}).$$

Inference

Using the previous assumptions, we can calculate the posterior distribution over the rewards and costs conditioned on an observed action A. Applying Bayesian reasoning, this is a combination of the likelihood of the observed action (given by the planning procedure), and the prior distributions over costs and rewards:

$$P(Reward, Cost \mid A, S) \propto P(A \mid S, Cost, Reward)\, P(Reward, Cost) \quad (6)$$

We constructed a probabilistic program that samples rewards and costs from this posterior, conditioned on observed actions matching the actions of the agent . The program was written in Church (*54*), a probabilistic programming language based on Scheme.

This program can be used to infer the value of the reward for the different experiments (cases 1-8 as detailed above). Below we consider the specific case of Experiment 2 from the current paper (Ramps), using the NUC-force model. The sampling procedure in Church uses the Metropolis-Hastings algorithm, and Fig. S4 shows the resulting approximation to the posterior in Eq. 6 for Experiment 2 and the NUC-force model, with 5,000 samples at 10-step intervals with a burn-in of 1,000 samples. As can be seen in Fig. S4, the model shifts probability away from a uniform distribution for both goals, such that the reward for the higher-effort target is higher in expectation than the lower-effort one. For both the targets, most of the probability mass is in between the accepted and rejected costs, as expected.

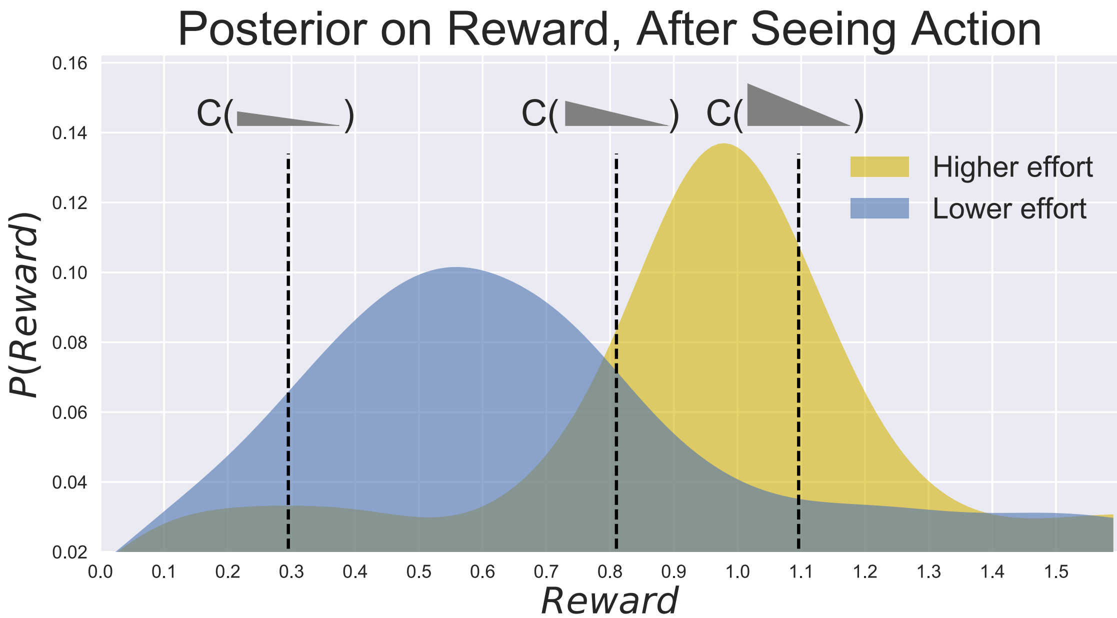

Fig S4. The model's inferred posterior probability distribution over possible reward values, after seeing the actions taken in Experiment 2 when using a Force-based cost function. Note that for the force-based model the cost of the different ramps is a combination of the effort of working against gravity and overcoming friction.

Action Prediction

The 'forward planning' part of the program implements Eq. 1-6 and the decision-making framework of the first section. This forward planning can use the posterior distribution inferred from observing the agent's action to predict its next actions, P(Action | Previous stimuli). This probability can be calculated for cases in which there is only one reward and multiple possible actions (as in Gergely et al. 1995, Experiments 1&2) or multiple rewards (the other experiments considered). In all model comparisons we used 5,000 samples at 10-step intervals with a burn-in of 1,000 samples for the inference of the reward distributions, and for the sampling of action predictions. We discounted the 'Do Nothing' action predictions, as the infants never saw these in the test stimuli, and were faced rather with a direct comparison between two actions.

To give an example of a particular way in which action prediction proceeds, we consider again Experiment 2 from the current paper under the NUC-force model. For the test stimuli, the situation includes no barriers, and both targets are obtainable. The possible actions are: going to target A, going to target B, or doing nothing. We assume that:

Reward(Target A) ~ P(Reward(Target A) | Previous Stimuli), (7)

Reward(Target B) ~ P(Reward(Target B) | Previous Stimuli), (8)

where *P(Reward(Target) |Previous Stimuli)* is calculated using the program approximating Eq. 6.

Fig. S5 shows the resulting prediction for the probability distribution over the agent's action. As can be seen in Fig. S5, the model predicts the agent will move towards the Target that it expended higher effort to reach earlier. Consequently, when the agent moves towards the lower effort Target, this goes against the prediction.

The amount of surprise the model predicts can be directly related to the inverse of the probability of a given event or action (*34*).

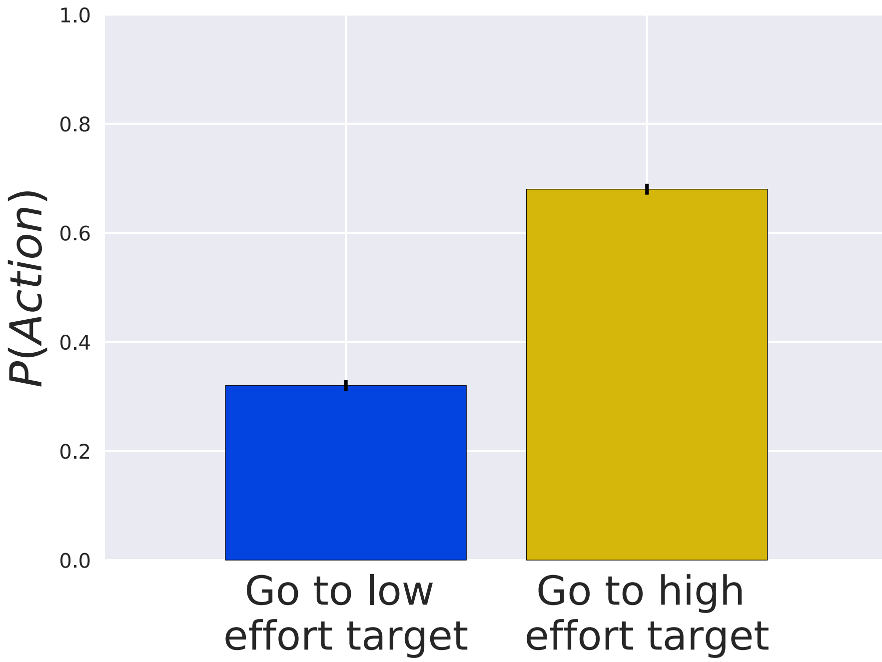

Fig S5. Probability of predicted next action by the NUC-force model for Experiment 2 (Ramps), for the test case in which both targets are equally accessible. Black lines show bootstrapped 95% confidence intervals.

Alternative Models

The NUC-force and NUC-distance models outlined so far assume that infants are able to use Bayesian reasoning to integrate inferences about cost, reward, and agent rationality to predict the next action of an agent. The NUC-distance model assumes that cost is proportional to perceptually observable distance, while the NUC-force model assumes that cost is proportional to physical effort.

But an alternative option is that infants are performing a much more low-level analysis of the scene, or reasoning about costs and rewards in a non-integrated way. Consider for example the findings from Experiment 1 in Woodward (1998), showing that 9-month old infants expected an agent to reach for a goal object A over B, after A and B had switched positions compared to habituation. A full NUC mental model could be deployed to infer that the reward for A is higher than the reward for B, and then correctly predict the action. But it is also possible that infants are reasoning along the lines of "A goal previously reached for will be reached for again, regardless of its position".

We thus consider 3 alternatives to the NUC model, all of which involve different forms of lower level or non-integrated reasoning, as follows:

1) **Count-based:** This is a perceptual-based account that relies on directly observable cues: reaching an item, and the distance to the item. This model is similar to the NUC in that agents are assumed to be rational planners that act to achieve goals under constraints. However, rather than using full Bayesian inverse planning to reason about rewards, it uses an easy-to-calculate proxy for the reward of a target: The model tallies the number of times a goal object has been approached or reached, and considers that tally in proportion to the total number of times any goals have been reached. Furthermore, the model considers the cost as proportional to the distance. Thus, for this model:

$$P(Reward_i |\, Stimuli) = \frac{\#(Reward_{\,i}\ approached)}{\sum_j \#(Reward_{\,j}\ approached)}\ . \quad (9)$$

The cost is calculated as in Eq. 4, and the planning proceeds as in Eq. 1-3.

2) **Cost Only:** This model assumes all goal states are equally rewarding, but is sensitive to the cost in the form of distance as in Eq. 4. Such a model is able to reason about agents acting efficiently to reach goals, in the sense of minimizing distance as in Gergely et al. (1995), for example.

3) **Reward Only:** This model assumes rewards can vary and performs the correct Bayesian updating on the probability distribution over reward as in Eq. 6, but without the cost factor. Such a model is sufficient for correctly predicting behavior when reasoning about the choices of agent that shows a simple preference for one target over another, as in Woodward (1998), for example.

Comparison to Empirical Data from Eight Experiments

In this section we compare the outputted prediction of the NUC models as well as the alternative models, to eight different experiments examining infants' expectations about cost, reward and efficient behavior. The experiments considered were:

1) Experiment 1 in the current paper (Barriers).
2) Experiment 2 in the current paper (Ramps).
3) Experiment 3 in the current paper (Gaps).
4) Experiment 1 in Liu and Spelke (2017), in which six-month-old infants first view an agent jumping over barriers of varying sizes to get to a goal. In the test phase, the agent is blocked by a previously unseen small barrier and either makes a small jump (expected) or a large jump (unexpected) over the barrier to reach the goal.
5) Experiment 2 in Liu and Spelke (2017), which is identical to (3), except that the barriers are placed behind the goal such that they are not blocking the path of the agent. Infants' looking time was at chance when shown a small or large jump, presumably inferring that the agent is irrational based on previous behavior.
6) Experiment 1 in Gergely et al. (1995), in which 12-month-olds first view an agent jumping over a barrier to reach a goal. In the test phase, the barrier is removed and the agent either moves straight to the goal (expected) or performs the previously seen jump on the way to the goal (unexpected).
7) Experiment 2 in Gergely et al. (1995), which is identical to (5) except the barrier is placed behind the goal such that it is not blocking the path of the agent. Infants' looking time was at chance when shown the agent moving straight or making a jump on the way to the goal.
8) Experiment 1 in Woodward (1998), in which 9-month-old infants first saw an agent reaching for and grasping one of two possible goals. In the test phase, the positions of the goals were switched. The agent either reached to the same goal using a new trajectory (expected) or follow the same trajectory to reach a new goal (unexpected).

We use the same model set-up and parameters for all experiments, varying only the actions and stimuli observed, and the different cost and reward functions used depending on the model.

Following (*34*), we relate the amount of surprise in infants' looking time to 1-P(outcome) as predicted by the model. Specifically, we relate this measure to the difference in looking time (measured in seconds) between the unexpected event and the expected event. The results of the models and the comparison to the eight experiments are shown in Fig. S3.

Code

All the code implementing the computational models and the analysis of their output is available on Open Science Framework (https://osf.io/crx4d/).