Précis of *Rational Approaches to Learning and Development*

Celeste Kidd
Brain & Cognitive Sciences, University of Rochester

In this dissertation, I study decision-making mechanisms in young children and non-human primates across multiple domains—including visual attention and overt choice—in order to discover the efficacy and limitations of rational cognitive theories. Rational cognitive theories posit that intelligent organisms should choose actions that maximize utility (e.g., Anderson, 1991; Oaksford & Chater, 1994). Effective utility maximization requires an agent to generate accurate expectations about what will happen in the future—which, in turn, requires the agent to mentally model the world. Adults have a substantial amount of world experience upon which they can base such models. Very young children, however, have far less world experience, and thus—at least initially—possess no such advantage. They must sample observations from their environments in order to overcome their naïveté concerning the structure of those environments. They gradually infer complex representations and form abstract theories about the world through the sampling process. Sample by sample, they continuously build upon, update, and revise their theories.

My dissertation investigates young children's choice behavior throughout development in order to understand the decision-making mechanisms that guide knowledge acquisition. The included studies test the efficacy and limitations of rational cognitive theories in order to better understand the decision-making mechanisms that guide early learning and behavior in children and monkeys. The studies span across multiple domains—including visual attention and overt choice. The results provide empirical evidence that learners rely on utility maximization both to build complex models of the world starting from very little knowledge and, more generally, to guide their decisions and behavior. Five experiments were conducted on infants, young children, and monkeys using visual and auditory stimuli presented in sequences of events. These experiments show that children are capable of rational decisions that optimize future utility and exhibit a U-shaped relationship between stimulus complexity and attention. Similarly, monkeys' attentional patterns are also guided by stimulus complexity.

Rational decision-making in children

The dissertation opens by applying a rational cognitive approach to perhaps the most classic example of seemingly irrational behavior in young children—their poor track record in delay-of-gratification tasks, such as in the Stanford Marshmallow experiments (e.g., Mischel & Ebbesen, 1970). Though children apparently fail to maximize utility in such delay-of-gratification tasks, the cause of these apparent failures was not well understood (Chapter 2, "Rational Snacking: Young children's decision-making on the marshmallow task is moderated by beliefs about environmental reliability", Kidd, Palmeri, & Aslin, *Cognition*, 2012). For example, most 3- to 5-year-olds choose an immediately available low-value reward (e.g., one marshmallow) over one of high-value (e.g., two marshmallows) after a temporal delay (Mischel & Ebbesen, 1970).

One possible explanation of this choice is a deficiency in self-control: young children may be incapable of inhibiting their immediate-response tendencies to seek gratification (e.g., Marcovitch & Zelazo, 1999; Piaget, 1954). Previous work had implicated a child's capacity for self-control as a major causal factor in a child's later life successes (or failures). Mischel, Shoda, and Peake (1988) analyzed data from adolescents who, many years earlier, had been presented with a laboratory choice task: eat a single marshmallow immediately, or resist the temptation during a sustained delay to receive two marshmallows. With no means of distracting themselves from a treat left in view, the majority of children failed to wait for the maximum delay (15 or 20 min) before eating the marshmallow, with a mean wait time of 6 min and 5 s. Importantly, longer wait-times among children were correlated with greater self-confidence and better interpersonal skills, according to parental report. Longer wait-times also correlated with higher SAT scores (Shoda et al., 1990), less likelihood of substance abuse (Ayduk et al., 2000), and many other positive life outcomes (e.g., Mischel et al., 1989). Based on these findings, the marshmallow task was argued to be a powerful diagnostic tool for predicting personal well-being and later-life achievement—"an early indicator of an apparently long-term personal quality" (Mischel et al., 1988). The logic of the claim is that a child who possesses more self-control can resist fleeting temptations to pursue difficult goals; in contrast, children with less self-control fail to persist toward these goals and thus achieve less.

However, following the rational framework pursued in this thesis, another possibility is that children's performance may result from their expectations and beliefs (Mahrer, 1956; Mischel, 1961; Mischel & Staub, 1965), which are likely different from adults' and vary across children. Under this second theory, children engage in rational decision-making about whether waiting for the high-value reward yields an expected gain in utility. The basis for this theory centers on what it means to be rational in the context of the marshmallow task. Waiting is only the rational choice if you believe that a second marshmallow is likely to actually appear after a reasonably short delay—and that the marshmallow currently in your possession is not at risk of being taken away. This presumption may not apply equally to all children. Consider the mindset of a 4-year-old living in a crowded shelter, surrounded by older children with little adult supervision. For a child accustomed to stolen possessions and broken promises, the only guaranteed treats are the ones you have already swallowed. At the other extreme, consider the mindset of an only-child in a stable home whose parents reliably promise and deliver small motivational treats for good behavior. From this child's perspective, the rare injustice of a stolen object or broken promise may be so startlingly unfamiliar that it prompts an outburst of tears. The critical point of this vignette is that rational behavior is inferred by understanding the goals and expectations of the agent (Anderson, 1991; Anderson & Milson, 1989; Marr, 1982).

The dissertation presents data that support this second hypothesis based on strategic reasoning. We tested 3- to 5-year-old children (M = 4;6, $N = 28$) using a variant of Mischel (1974)'s marshmallow task. In our experiment, we preceded marshmallow-task testing with evidence that the experimenter running the study was either reliable or unreliable as a means of manipulating children's beliefs across conditions. Half of the children observed evidence that the researcher was reliable in advance of the marshmallow task, while half observed evidence that she was unreliable. If children employ a rational process in deciding whether or not to eat the first marshmallow, we expect children in the reliable condition to be significantly more likely to wait than those in the unreliable condition. Children who believed the experimenter was reliable waited about four times longer before eating the marshmallow than those who thought she was unreliable (12 min vs. 3 min, $p < 0.0005$), and were far more likely to wait the entire 15-minute duration ($p < 0.006$). The resulting effect of our experimental manipulation was quite robust and large. Importantly, while there were small procedural differences between our study and past studies, children—age and gender-matched to the current study—who faced similar choices without prior explicit evidence of experimenter reliability waited for around 6 minutes (e.g., 6.08 min in Shoda et al. 1990 and 5.71 min in Mischel & Ebbesen 1970). When we manipulated experimenter reliability, children waited twice that long in the reliable condition (12.03 min), and half as long in the unreliable condition (3.02 min).

These results suggest that children's wait-times are modulated by a rational decision-making process that considers environmental reliability. The results may also provide an alternative explanation for why marshmallow wait-times correlate with later life success (e.g., Mischel et al., 1989)—successful people grow up in reliable situations. Broadly, this illustrates that children build a model of the reliability of others' behavior, and use this model to inform their decisions. Because children waited in the reliable condition, our experiment provides compelling evidence that young children are indeed capable of delaying gratification in the face of temptation when provided with evidence that waiting will pay off. This suggests that strategic reasoning, rather than a deficiency in self-control determines their behavior. More broadly, the effect we observed is consistent with converging evidence that young children are sensitive to uncertainty about future rewards (Fawcett et al., 2012; Mahrer, 1956; McGuire & Kable, 2012).

To be clear, our data do not demonstrate that self-control is irrelevant in explaining the variance in children's wait-times on the original marshmallow task studies. They do, however, strongly indicate that it is premature to conclude that most of the observed variance—and the longitudinal correlation between wait-times and later life outcomes—is due to differences in individuals' self-control capacities. Rather, an unreliable worldview, in addition to self-control, may be causally related to later life outcomes, as already suggested by an existing body of evidence (e.g., Barnes & Farrell, 1992; Smyke et al., 2002). The results therefore raise questions about the most effective interventions for at-risk children: if children's behavior is rational, training on delaying gratification may not be productive.

Rational models of attention and learning

The marshmallow study relied on the fact that learners have aggregated information about the reliability of adult reward-promises prior to being tested. However, learners do not enter the world with access to most of this information—how do infants begin to make sense of the world with little or no knowledge on which to base their inferences?

Next (Chapters 3-5), I applied the same rational approach embodied by the *Rational Snacking* project to my primary line of research: infant attention. These results suggest that key attentional mechanisms filter environmental stimuli in a particularly useful way, thereby providing infants with data that are "just right" for learning (which we referred to as a "Goldilocks" effect). This work explored attentional behavior in 7- and 8-month-old infants. We showed infants visual event sequences of varying complexity, as measured by an idealized learning model, and measured when in each sequence infants decided to terminate their attention by looking away from the display. We found that infants' probability of looking away was greatest to events of either very low information content (highly predictable) or very high information content (highly surprising).

In these studies, our goal was to determine whether infants are biased to gather information from the environment in a principled way that serves as a key component of an efficient learning mechanism (e.g., Berlyne, 1960; Piaget, 1970). Specifically, we uncovered evidence that infants avoid spending time examining stimuli that are either too simple (highly predictable) or too complex (highly unexpected) according to their implicit beliefs about the probabilistic structure of events in the world. Rather, infants allocate their greatest amount of attention to events of intermediate surprisingness—events that are likely to have just enough complexity so that they are interesting, but not so much that they cannot be understood.

Many researchers have speculated about what underlying mental operations are indexed by infants' looking times or attentional patterns (Fantz, 1964; for review: Aslin, 2007). The generally accepted view is that looking times reflect some combination of (a) stimulus-driven attention, (b) memory of past stimuli, and (c) comparison between the current and the past stimuli. Previous theoretical accounts for familiarity and novelty preferences all shared a common theme: As infants attempt to encode various features of a visual stimulus, the efficiency or depth of this encoding process determines their subsequent preferences. Familiarity preferences arise when infants have not yet completed encoding the familiar stimulus into memory, or when the novel stimulus is too dissimilar from the infants' existing mental representations to be readily encoded (e.g., Dember & Earl, 1957; Hunter & Ames, 1988; Kinney & Kagan, 1976; Roder et al., 2000; Rose et al., 1982; Sokolov, 1963; Wagner & Sakovits, 1986). However, these theories lacked an objective measure of the relevant independent variable–an event's complexity or relationship to existing representations. Instead, researchers overwhelmingly relied on qualitative judgments of stimulus complexity to select materials to test infants' visual preferences. These qualitative judgments relied on inferences about infants' existing mental representations, to which researchers had no direct access.

In our studies, we overcame these problems by formalizing a notion of stimulus complexity and behaviorally testing the relationship between complexity and infants' probability of looking away at each successive point in a sequence of events. We assume that at each point in the experiment—and in everyday life—infants have used observed data to form probabilistic expectations about what events are likely and unlikely to be observed next (e.g., Téglás et al., 2011; Xu & Garcia, 2008). We modeled these expectations using an idealized observer model of our experimental stimuli based on principled Bayesian inference (a Markov Dirichlet-multinomial, MDM). We then measured complexity as the negative log probability of an event according to this idealized model. This measure quantifies each event's *information content* (Shannon, 1948). This measure has also been called *surprisal* (Tribus, 1961), since it may also be interpreted as representing the "surprise" of seeing the outcome. We used nonparametric statistical methods that could potentially reveal any relationship between surprisal and looking behavior.

The results demonstrate that infants preferentially look away at events that are either very simple (high probability) or very complex (low probability), according to the idealized model. Intuitively, high probability events convey little information–infants' attentional resources are best spent elsewhere. Low probability events may indicate that the observed stimuli are unlearnable, unstructured, or difficult to use predictively in the future. Negative log probability also quantifies the number of bits of information an ideal observer would require to encode that sequence of events in memory. Thus, infants may avoid stimuli that require encoding too much information or information that could only be extracted by prolonged attention to rare events, thereby incurring a higher processing cost than shifting attention to less complex events.

Chapter 3 demonstrates that this attentional strategy holds in multiple types of visual displays (Chapter 3, "The Goldilocks Effect: Human infants allocate attention to visual sequences that are neither too simple nor too complex", Kidd, Piantadosi, & Aslin, *PLoS ONE*, 2012). Chapter 4 demonstrates that it holds for auditory stimuli (Chapter 4, "The Goldilocks Effect in Infant Auditory Attention", Kidd, Piantadosi, & Aslin, *Child Development*, 2014). And our other work, not included in the thesis, shows that the same U-shaped effects can be found even within individual infants (see reference in Chapter 5 intro, "Rich Analysis and Rational Models: Inferring individual behavior from infant looking data," Piantadosi, Kidd, & Aslin, *Developmental Science*, 2014).

Together, these results suggest a broadly applicable principle of infant attention: infants implicitly decide to direct their attention in order to maintain intermediate rates of information absorption. Infants implicitly seek to maintain intermediate rates of information absorption and avoid wasting cognitive resources on overly simple or overly complex events. The results have important implications for two interrelated hypotheses concerning infants' attention. First, infants behave as if they are employing a principled inferential process for learning about events in the world. The particular model used in our analyses took as inputs a series of observed events or transitions between events to form probabilistic expectations about what events are most likely to occur in the future. The model was necessary to determine what complexity a set of stimulus events conveys to an ideal observer. A failure of either of these components–the probabilistic model or the linking assumption that maps level of complexity onto looking times–would have yielded null results.

Second, infants appear to allocate their attention in order to maintain an intermediate level of complexity. A powerful feature of our analyses was an ability, via a regression, to control for potential confounds such as the number of items that have not appeared yet, item repeats, and an arbitrary baseline distribution of look-away probabilities. To our knowledge, the hypothesis that infants prefer a particular level of information has not been tested while controlling for these other variables, and our analyses therefore provide several methodological advances. Rather than predicting infants' average looking time to a stimulus, our analyses predicted the precise event in a sequence when an infant would terminate (i.e., look away from) the display. Although others have observed U-shaped behavior in infants under some circumstances, our results provide the first objectively quantifiable evidence that the information-theoretic properties of a formal model provide a significant predictor of infant look-aways, over and above the effects of other variables, for a large set of arbitrary, neutral visual stimuli.

Perhaps most importantly for the field, our results also provide a formal account for why infants show novelty preferences (when two test stimuli fall on the left half of the U-shaped function, the stimulus with greater complexity elicits more attention) or familiarity preferences (when two test stimuli fall on the right half of the U-shaped function, the stimulus with lesser complexity elicits more attention). This is a longstanding methodological puzzle, and these results provide one possible solution: infant's shifting preference for novel and familiar stimuli may result from a single, underlying U-shaped curve expressing a preference to attend to intermediately predictable events.

In summary, our findings are consistent with theories that suggest infants actively seek to maintain an intermediate level of information absorption, avoiding allocating cognitive resources to either overly predictable or overly surprising events. Further investigation is required to determine how infants' preference for intermediate levels of information affects the outcome of learning, either by enhancing the rate of learning or its asymptotic level.

Intrinsic curiosity and information-seeking in monkeys

In the final experimental chapter of my dissertation, I explore the role of curiosity in guiding learners' attention. Curiosity is a function that drives intelligent creatures—including humans, apes, rats, felines, and canines—to reduce the uncertainty that is inherent in a complex world. Recent research findings in the domains of behavioral economics, memory, and motivation have provided insights into some of the neurological mechanisms that underlie curiosity-driven behavior. Under this framework, a predominant theory is that curiosity is a state of increased arousal whose termination is rewarding and facilitates memory (e.g., Berlyne, 1960; Jepma et al., 2012). In other words, curiosity is a negative reinforcer (or aversive condition) that motivates exploration and discovery. A unifying feature of this recent work is that a wide variety of researchers attempt to explain behavior through neurobiological processes—typically, the processes associated with desire, motivation, discovery, reward, and memory.

However, it remains unclear whether the brain possesses mechanisms that track informational complexity and deploy attention adaptively. Rather than asking what biological processes drive exploratory behavior, here we ask what high-level function the exploratory behavior serves from the rational perspective adopted throughout the thesis—an inquiry falling squarely within Marr's computational level of analysis (Marr, 1982). The final experimental chapter of my dissertation aims to tackle two major computational-level questions. First, what does curiosity do, exactly? Does it motivate exploration of any available novel stimulus in a random manner, or perhaps only the most informative stimuli? Addressing this question requires understanding the relationship between intrinsic curiosity and relevant features of the stimulus. This relationship would, in essence, elucidate the high-level design features that govern attentional selection and termination— what makes something inherently interesting? Second, we ask why biology should have given rise to the particular set of exploratory mechanisms that govern attentional selection, as opposed to any other? What greater purpose do exploratory mechanisms serve?

In the studies in Chapter 5, I demonstrate that juvenile rhesus monkeys, like human infants, allocate attention according to the statistical properties of stimuli in their environments. We measured monkeys' visual attention to sequential visual events that varied in their information theoretic properties (e.g., surprisal), as determined by an idealized learning model. We tested five juvenile male rhesus monkeys (Macaca mulatta) from the University of Rochester colony. Both the experiment and modeling approach were based on our earlier studies of infant visual and auditory attention (Chapters 3-4). We tested whether a rational statistical model that is similar to those used with our infant data can explain monkeys' attentional patterns. In doing this, we hope to discover the factors that influence monkeys' intrinsic curiosity and the mechanisms that govern attentional allocation and learning.

Our analysis revealed that were more likely to predictively look at more predictable object pop-up events ($\beta = -0.25$, $z = -3.89$, $p < 0.0001$). This result suggests that, like human learners, monkeys can rapidly update their expectations in accordance with the statistical properties of incoming information streams and reallocate their attention appropriately. The analysis also revealed evidence of information-seeking behavior in the monkey learners. Monkeys also exhibited increased curiosity for unknowns in the visual displays. The controlled unigram predictive-looks regression revealed that monkeys produced more predictive looks when there were more previously unseen objects ($\beta = 0.23$, $z = 5.97$, $p < 0.0001$) and on an object's first appearance ($\beta = 0.41$, $z = 6.78$, $p < 0.0001$). These results suggest that monkeys exhibited increased visual interest in boxes that had not yet revealed their contents. The transitional-model regression also revealed the indicators of information-seeking—more predictive looks for previously unseen objects ($\beta = 0.30$, $z = 7.82$, $p < 0.0001$) and on an object's first appearance ($\beta = 0.55$, $z = 9.76$, $p < 0.0001$).

The model's significant relationship with the behavioral measures is strong evidence that monkeys are able to track the statistics of the displays and that these statistics influence their attention. Our analysis has sought to discover the relationship between estimations of statistical probability and attentional behavior. This has revealed a range of behavior across measures and types of analyses. A U-shaped relationship was observed in the unigram analysis of reaction times, and other measures showed significant linear trends in predicted directions. The robustness of the unigram statistics in predicting each of the behavioral measures possibly suggests an important difference between monkeys and human learners. In general, unigram statistics were a more robust predictor of behavior than transitional statistics. This contrasts with infant attentional behavior, for which attention to sequential visual (Chapter 3) and auditory (Chapter 4) stimuli was more strongly influenced by transitional than unigram statistics. This difference may suggest that infants possess a sensitivity to transitional probabilities that monkeys lack, potentially pointing towards an important inferential precursor to language learning.

These results reflect the first behavioral evidence that monkeys' intrinsic curiosity, in the absence of any external rewards or explicit task goals, is governed by the statistical properties of stimuli in their environment. More importantly, these results are the first to demonstrate that monkeys' curiosity reflects their probabilistic beliefs and knowledge about the world, which are rapidly updated on a continuous basis as they make new observations.

Rational species-general principles of learning

A key novel feature of the research approach is the combination of behavioral methods and computational modeling. This model-driven behavioral experimentation enables me to rigorously test competing theories of decision-making and learning by quantifying otherwise unobservable cognitive processes or variables through the use of a computational model. For instance, in the "Goldilocks" work, our model is primarily used as a measure of an otherwise unobservable feature of the world, the information conveyed by a stimulus. By relating the model-based measure of information to infants' behavior, we are able to formalize and test a hypothesis about infant attention that had previously only been studied qualitatively.

This is a powerful approach because traditional infant methods typically only compare the preferences of groups of infants. My thesis builds upon these traditional methods, but attempts to formalize and test detailed predictions about behavioral patterns, allowing me to formalize and test a wide range of formal theories. As an added bonus, this approach offers the potential to generate specific predictions about the learning outcomes of individual children on the basis of their particular behavioral patterns.

In conclusion, the work in this dissertation investigated both implicit and overt measures of the choice behavior of both young children and monkeys in order to understand the decision-making mechanisms that guide the acquisition of knowledge. This work, which encompassed behavioral experimentation with young children and non-human primates across multiple domains, aimed to better understand the efficacy and limitations of rational cognitive theories. The results in this thesis presented empirical evidence that suggests that naïve learners rely on rational utility maximization both to build complex models of the world starting from very little knowledge and, more generally, to guide their decisions and behavior.

References

Anderson, J.R., & Milson, R. (1989). Human memory: an adaptive perspective. *Psychological Review, 96*, 703–719.

Anderson, J.R. (1991). *Cognitive architectures in a rational analysis*. In K. VanLehn (Ed.), Architectures for Intelligence (pp. 1–24). Hillsdale, NJ: Lawrence Erlbaum Associates.

Aslin, R.N. (2007). What's in a Look? *Developmental Science, 10,* 48–53.

Ayduk, O., Mendoza-Denton, R., Mischel, W., Downey, G., Peake, P.K., & Rodriguez, M. (2000). Regulating the interpersonal self: Strategic self-regulation for coping with rejection sensitivity. *Journal of Personality and Social Psychology, 79*, 776–792.

Barnes, G.M., & Farrell, M. P. (1992). Parental support and control as predictors of adolescent drinking, delinquency, and related problem behaviors. *Journal of Marriage and Family, 54*(4), 763–776.

Berlyne, D.E. (1960). *Conflict, Arousal, and Curiosity. (Volume 331).* New York: McGraw-Hill.

Dember, W.N., & Earl, R.W. (1957). Analysis of Exploratory, Manipulatory, and Curiosity Behaviors. *Psychological Review, 64*, 91–96.

Fantz, R.L. (1964). Visual Experience in Infants: Decreased attention to familiar patterns relative to novel ones. *Science, 146*, 668.

Fawcett, T.W., McNamara, J.M., & Houston, A.I. (2012). When is it adaptive to be patient? A general framework for evaluating delayed rewards. *Behavioural Processes, 89*, 128–136.

Hunter, M.A., & Ames, E.W. (1988). A Multifactor Model of Infant Preferences for Novel and Familiar Stimuli. *Advances in Infancy Research, 5*, 69–95.

Jepma, M., Verdonschot, R.G., van Steenbergen, H., Rombouts, S.A.R.B., & Nieuwenhuis, S. (2012). Neural Mechanisms Underlying the Induction and Relief of Perceptual Curiosity. *Frontiers in Behavioral Neuroscience, 6*, 5.

Kidd, C., Blanchard, T., Aslin, R.N., & Hayden, B. (In prep). Information-seeking behavior in human-infant and macaque learners.

Kidd, C., Piantadosi, S.T., & Aslin, R.N. (2012). The Goldilocks Effect: Human infants allocate attention to visual sequences that are neither too simple nor too complex. *PLoS ONE*, 7(5), e36399.

Kidd, C., Piantadosi, S.T., & Aslin, R.N. (2014). The Goldilocks Effect in Infant Auditory Attention. *Child Development.*

Kidd, C., Palmeri, H., & Aslin, R.N. (2013). Rational Snacking: Young children's decision-making on the marshmallow task is moderated by beliefs about environmental reliability. *Cognition.*

Kinney, D.K. & Kagan, J. (1976). Infant Attention to Auditory Discrepancy. *Child Development, 47*(1), 155–164.

Mahrer, A.H. (1956). The role of expectancy in delayed reinforcement. *Journal of Experimental Psychology, 52*, 101–106.

Marcovitch, S., & Zelazo, P.D. (1999). The A-not-B error: Results from a logistic meta-analysis. *Child Development, 70*, 1297–1313.

Marr, D. (1982). *Vision.* San Francisco: W.H. Freeman.

McGuire, J.T., & Kable, J.W. (2012). Decision makers calibrate behavioral persistence on the basis of time-interval experience. *Cognition, 124*(2), 216–226.

Mischel, W. (1961). Father-absence and delay of gratification: Cross-cultural comparisons. *Journal of Abnormal and Social Psychology, 63*, 116–124.

Mischel, W. (1974). Processes in delay of gratification. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 7, pp. 249–292). New York: Academic Press.

Mischel, W., & Ebbesen, E.B. (1970). Attention in delay of gratification. *Journal of Personality and Social Psychology, 21*, 204–218.

Mischel, W., Shoda, Y., & Peake, P.K. (1988). The nature of adolescent competencies predicted by preschool delay of gratification. *Journal of Personality and Social Psychology, 54*, 687–696.

Mischel, W., Shoda, Y., & Rodriguez. (1989). Delay of gratification in children. *Science, 244*, 933–938.

Mischel, W., & Staub, E. (1965). Effects of expectancy on working and waiting for larger rewards. *Journal of Personality and Social Psychology, 2*, 625–633.

Oaksford, M., & Chater, N. (1994). A rational analysis of the selection task as optimal data selection. *Psychological Review, 101*(4), 608-631.

Piaget, J. (1954). *The construction of reality in the child.* Oxford: Basic Books.

Piaget, J. (1970). Piaget's theory. In P. Mussen (Ed.). *Handbook of child psychology (Vol. 1)*. New York: Wiley.

Piantadosi, S.T., Kidd, C., & Aslin, R.N. (2014). Rich Analysis and Rational Models: Inferring individual behavior from infant looking data. *Developmental Science.*

Roder, B.J., Bushnell, E.W., & Sasseville, A.M. (2000). Infants' Preferences for Familiarity and Novelty During the Course of Visual Processing. *Infancy, 1,* 491–507.

Rose, S.A., Gottfried, A.W., Melloy-Carminar. P., & Bridger, W.H. (1982). Familiarity and Novelty Preferences in Infant Recognition Memory: Implications for information processing. *Developmental Psychology, 18,* 704–713.

Shannon, C.E. (1948). A Mathematical Theory of Communication. *Bell Systems Technical Journal, 27*, 379-423.

Shoda, Y., Mischel, W., & Peake, P.K. (1990). Predicting adolescent cognitive and social competence from preschool delay of gratification: Identifying diagnostic conditions. *Developmental Psychology, 26*, 978–986.

Smyke, A.T., Dumitrescu, A., & Zeanah, C.H. (2002). Attachment disturbances in young children. I: The continuum of caretaking casualty. *Journal of the American Academy of Child & Adolescent Psychiatry, 41*(8), 972–982.

Sokolov, E. (1963). P*erception and the Conditioned Reflex*. Oxford, England: Pergamon.

Téglás, E., Vul, E., Vittorio, G., Gonzalez, M., Tenenbaum, J. B., & Bonatti, L.L. (2011). Pure reasoning in 12-month-olds as probabilistic inference. *Science, 332*(6033), 1054–1059.

Tribus, M. (1961). *Thermostatics and Thermodynamics: an introduction to energy, information and states of matter, with engineering applications.* New York: D. Van Nostrand.

Wagner, S.H., & Sakovits L.J. (1986). A Process Analysis of Infant Visual and Cross-Modal Recognition Memory: Implications for an amodal code. *Advances in Infancy Research, 4,* 195–217.

Xu, F., & Garcia, V. (2008). Intuitive statistics by 8-month-old infants. *Proceedings of the National Academy of Sciences of the United States of America, 105,* 5012–5015.