Experiential reward learning outweighs instruction prior to adulthood

Johannes H. Decker,
Sackler Institute for Developmental Psychobiology, Weill Cornell Medical College, 1300 York Avenue, Box 140, New York, NY 10065, USA

Frederico S. Lourenco,
Sackler Institute for Developmental Psychobiology, Weill Cornell Medical College, 1300 York Avenue, Box 140, New York, NY 10065, USA

Bradley B. Doll, and
Center for Neural Science, New York University, New York, NY, USA

Catherine A. Hartley
Sackler Institute for Developmental Psychobiology, Weill Cornell Medical College, 1300 York Avenue, Box 140, New York, NY 10065, USA

Abstract

Throughout our lives, we face the important task of distinguishing rewarding actions from those that are best avoided. Importantly, there are multiple means by which we acquire this information. Through trial and error, we use experiential feedback to evaluate our actions. We also learn which actions are advantageous through explicit instruction from others. Here, we examined whether the influence of these two forms of learning on choice changes across development by placing instruction and experience in competition in a probabilistic-learning task. Whereas inaccurate instruction markedly biased adults' estimations of a stimulus's value, children and adolescents were better able to objectively estimate stimulus values through experience. Instructional control of learning is thought to recruit prefrontal–striatal brain circuitry, which continues to mature into adulthood. Our behavioral data suggest that this protracted neurocognitive maturation may cause the motivated actions of children and adolescents to be less influenced by explicit instruction than are those of adults. This absence of a confirmation bias in children and adolescents represents a paradoxical developmental advantage of youth over adults in the unbiased evaluation of actions through positive and negative experience.

Keywords

Development; Reinforcement learning; Decision-making; Cognitive control

C. A. Hartley: cah2031@med.cornell.edu.

Learning to obtain rewards and avoid punishment is critical for the survival of all organisms. An approach to this challenge that is employed across species is trial and error-based learning. By aggregating positive and negative feedback stemming from our previous actions, we are able to estimate how beneficial a given action might be in the future. Although such feedback-driven learning is effective, the need to learn about the consequences of our actions through direct experience can be inefficient at best, and dangerous when the potential outcomes are extremely negative.

Recruiting a sophisticated capacity for symbolic communication, humans regularly circumvent these shortcomings of experiential learning by conveying the value of an action through rules, advice, or other forms of explicit instruction. By selecting actions based on instruction, a learner is able to benefit from the prior experience and knowledge of others. The utility of transmitting information through instruction is particularly evident in the context of development. Children and adolescents receive a great deal of instructed information from parents, teachers, and public policy campaigns that seek to educate and protect them, as well as from their peers. An assumption inherent in providing such guidance is that instruction can direct children and adolescents' behavior as effectively as, or better than, their own experiential learning. To date, few studies have directly examined whether the efficacy of learning from instruction versus experience changes across development. However, our understanding of the cognitive processes and neural circuits implicated in such learning, and their maturational trajectories, suggests that there may be qualitative changes in the recruitment of instructed versus experiential learning across development.

Previous research has demonstrated that providing adults with instruction or advice induces a behavioral "confirmation bias," in which recommended actions are valued more highly than those learned solely through experience, even when those recommendations turn out to be inaccurate (Biele, Rieskamp, & Gonzalez, 2009; Biele, Rieskamp, Krugel, & Heekeren, 2011; Doll, Hutchison, & Frank, 2011; Doll, Jacobs, Sanfey, & Frank, 2009). This instructional biasing of experiential learning is thought to stem from the influence of the prefrontal cortex, implicated in rule-guided behavior (Bunge & Zelazo, 2006; Miller & Cohen, 2001), on feedback-based evaluative processes in the striatum (McClure, Berns, & Montague, 2003; O'Doherty, Dayan, Friston, Critchley, & Dolan, 2003; Pagnoni, Zink, Montague, & Berns, 2002). This process has been modeled computationally as an instruction-consistent distortion of error-driven reinforcement-learning signals (Doll et al., 2009). Developmentally, the striatal signals implicated in feedback-based reward learning appear to be relatively mature in children and adolescents (Cohen et al., 2010; Galvan et al., 2006; van den Bos, Cohen, Kahnt, & Crone, 2012). In contrast, connectivity between the prefrontal cortex and the striatum exhibits marked structural changes from childhood through adulthood (Imperati et al., 2011; Liston et al., 2006). Consistent with the proposal that these connectivity changes reflect fine-tuning of the information exchange between these regions (Somerville & Casey, 2010), cognitive functions that depend on the integrity of frontostriatal pathways typically show continued maturation into adulthood (Liston et al., 2006; Rubia et al., 2006; Somerville & Casey, 2010). This neural model suggests that the biasing influence of explicit instruction on value-based choices might be diminished in

children and adolescents, predisposing them to exhibit greater reliance on experiential learning.

In the present behavioral study, we tested this hypothesis by having children, adolescents, and adults complete a probabilistic reward-learning task consisting of a learning phase immediately followed by a test phase. In the learning phase, three pairs of stimuli were presented, and participants could learn experientially, through trial and error, which stimulus within each pair was most likely to yield reward. Importantly, participants were given an inaccurate instruction that a lower-valued stimulus within one pair was likely to be rewarding. Participants could discover that this information was inaccurate through the subsequent positive or negative feedback following each choice. During the test phase, participants were presented with all possible pairings of the six stimuli from the learning phase, and they attempted to select the higher-valued option, receiving no feedback. By comparing their performance for instructed and uninstructed stimuli of equal value, we could determine the extent to which the false instruction biased their experiential learning of the true stimulus value, providing a quantitative measure of the influence of instruction on experiential learning. Previous studies in adults have demonstrated that inaccurate instruction strongly biases experiential value learning (Doll et al., 2011, 2009; Staudinger & Buchel, 2013). We hypothesized that children and adolescents would be less susceptible to this bias, instead relying predominantly upon their own experience to guide their choices.

Materials and methods

Participants

Participants were recruited through community-based events (e.g., street fairs) and flyers posted within institutions in the New York City metropolitan area. All participants (or parents of minors) were screened by phone prior to participation to ensure that the participant had no history of diagnosed neurological or psychiatric disorders, was not taking medication, and was typically developing cognitively and behaviorally (based on self- or parental report). We also ensured that all participants were not colorblind. All participants provided written consent to participate and were paid for their participation. They were debriefed following the experiment about the misleading nature of the instructions.

A total of 87 (51 female, 36 male) paid volunteers completed the study and were included in the analyses: 30 children (18 female, 12 male; 6–12 years of age, $M = 9.5$, $SD = 1.8$), 31 adolescents (15 female, 16 male; 13–17 years of age, $M = 14.8$, $SD = 1.5$), and 26 adults (17 female, nine male; 18–34 years of age, $M = 23.0$, $SD = 4.3$). Previous studies (Biele et al., 2009; Doll et al., 2009) had reported large instruction-bias effect sizes in adults ($d = 0.9$ and $d = 1.0–1.3$, respectively). Because we considered the possibility that children or adolescents might show a smaller effect, we targeted a sample size of 25 participants per group, which would enable us to detect a significant effect of at least 0.6 in each age group with 80 % power (alpha of .05, two-tailed). Additional participants were recruited to ensure adequate power in the event of subject attrition, particularly in the child and adolescent groups.

Behavioral paradigm

Participants completed an instructed probabilistic selection task (Doll et al., 2009) that was adapted for use across development, which consisted of a learning phase followed immediately by a test phase. Participants were told that their task was to feed a hungry mouse by helping him find the cheese hidden behind one of two mouseholes. During learning, participants saw one of three stimulus pairs on each trial, referred to here as AB, CD, and EF, which consisted of uniquely colored mouseholes (Fig. 1). These stimuli were chosen to make them easily distinguishable and to be as engaging as possible for our younger participants.

Participants were given positive or negative feedback (a happy mouse with cheese, or a sad mouse) after each choice during the learning phase, indicating whether they had made a "correct" or an "incorrect" choice. Although participants did not receive monetary rewards, previous studies have suggested that purely cognitive feedback in learning tasks recruits underlying neurocircuitry similar to that in reward-based reinforcement learning (Daniel & Pollmann, 2010; Rodriguez, Aron, & Poldrack, 2006; van den Bos et al., 2012). Though both stimuli in each pair were occasionally correct or incorrect, each pair had an optimal choice. The stimuli were probabilistically reinforced; for the AB pair, choosing "A" resulted in positive feedback on 80 % of the trials, whereas "B" led to positive feedback on 20 % of the trials. The other two pairs (CD and EF) had reward contingencies of 70 % (C/E) and 30 % (D/F), but participants were given inaccurate instruction about stimulus F (verbatim: "We'll get you started with a hint—this mousehole has a good chance of containing cheese"). This instruction was provided in textual format on the screen, accompanied by an image of the recommended mousehole. Thus, the instruction did not have a clear social source and was not directly associated with the experimenter, or with any specific individual. Before starting, participants completed a brief quiz on the task instructions, during which they were prompted to recall the recommended stimulus and were again visually reminded of this instruction. The participants saw each stimulus pair 60 times, pseudorandomized in ten-trial blocks, with side of presentation counterbalanced for each participant. Participants had 2.5 s to choose a stimulus and received feedback for 1 s.

Before the test phase, participants were told that they would now be tested on what they had just learned. Participants were presented with all 15 possible stimulus pairings (3 original, 12 novel), but were not given feedback after making a choice. For each pair, they were asked to "choose the mousehole that feels more correct based on what you've learned; if you're not sure which one to pick, go with your gut feeling." Participants saw each pair six times, randomly intermixed in the 90 test trials. A blank screen was presented between trials (150 ms duration), and participants had no time limit when making a choice.

Data analysis

Learning-phase choice behavior data were analyzed using a generalized linear mixed-effects model using the lme4 package for the R statistical language (Bates, Maechler, Bolker, & Walker, 2014). Optimal choice (i.e., choosing the higher-probability option) was modeled with independent predictors of age group (factors: Children, Adolescents, Adults), pair (factors: AB, CD, EF), trial (1–180, *z*-normalized), and all two- and three-way interactions.

We used a maximal random-effects structure (Barr, Levy, Scheepers, & Tily, 2013), including a per-participant adjustment to the intercept ("random intercepts"), as well as per-participant adjustments to the pair, trial, and pair-by-trial interaction terms ("random slopes"). In addition, we included all possible random correlation terms among the random effects. The p values and 95 % confidence intervals of the log-odds were determined through bootstrapping with 400 simulations using the bootMer function as implemented in the lme4 package, and p values for the analyses of variance were determined using likelihood ratio tests as implemented in the mixed function of the afex package. The data are presented visually in Fig. 2 below, using mean percent choices for each pair by age group, in ten-trial blocks.

We assumed that any biasing influence of the instructions on experiential learning would be revealed by a tendency to make instruction-consistent choices in the test phase, an effect previously observed when participants were instructed either that a suboptimal stimulus was good or an optimal stimulus was bad (Doll et al., 2009). We examined participants' test-phase choices for pairs that included the equally valued but differentially instructed 30 % stimuli (D and F) to determine the extent to which the learned stimulus values were biased by instruction. We first assessed whether participants chose in accordance with the instruction for the equally valued pair of 30 % stimuli (DF pair, 30:30 instructed; see Fig. 3A below). We then compared performance for a set of pairs from the test phase, in order to generate an instruction-bias score: AD (80:30), AF (80:30 instructed), DB (30:20), and FB (30 instructed:20); see Fig. 3B. These comparisons were chosen because any difference in performance—measured as the proportion of choices of the optimal, higher-probability option—between the two 80:30 pairs or 30:20 pairs was likely to be due to the instruction, since the stimuli were otherwise identical. The bias score was the mean of two difference scores: the difference between AD and AF performance, and the difference between FB and DB performance, each of which could vary between −1 and +1. Positive numbers would indicate an instruction-consistent bias, negative numbers an instruction-inconsistent bias (i.e., participants chose against the instruction), and values close to zero would indicate no instruction bias. The choice behavior data for these pairs were analyzed similarly to the training-phase analysis, except that the independent variables were age group (Children, Adolescents, Adults), pair (factors: Easy 80:30, Hard 30:20), and instruction (factors: Instructed, Uninstructed). Post-hoc testing was performed using t tests of the instruction-bias score mentioned above. To establish that all age groups exhibited above-chance experiential learning, we performed an additional analysis testing performance on all uninstructed pairs (A, B, C, and D combinations), with age group as the independent variable.

Response time data from each phase of the task were analyzed separately using a linear mixed-effects model. Models were constructed as before, with response time as the dependent variable and choice and its interactions added as additional, independent variables. All p values were determined using conditional F tests with Kenward–Roger correction of the degrees of freedom, as implemented in the ANOVA function (with Type III F tests) from the car package.

We used reinforcement-learning models to attempt to characterize how participants integrated the positive and negative feedback received during the learning phase. We used

participants' test-phase choices as indications of their learned stimulus values (Doll et al., 2011, 2009; Frank, Moustafa, Haughey, Curran, & Hutchison, 2007; Frank & O'Reilly, 2006; Frank, Seeberger, & O'Reilly, 2004) and fit two reinforcement-learning models to these test-phase choices by maximum a posteriori estimation (Daw, 2011; den Ouden et al., 2013) using the MATLAB Optimization Toolbox (The MathWorks, Inc., Natick, MA). For each model, we estimated the parameters that best captured how learning-phase feedback could be integrated to yield the choices observed in the test phase. The first model was a standard reinforcement-learning model, in which prediction errors (δ) were used to update the values (Q) associated with each stimulus. Feedback that was better than expected would yield a positive prediction error, and feedback worse than expected would yield a negative prediction error (δ). The learning rate parameter (a) determined the extent to which these prediction errors were incorporated into the updated stimulus value. This learning algorithm has been widely used to model an experiential trial- and error-based learning process (Bayer & Glimcher, 2005; Pessiglione, Seymour, Flandin, Dolan, & Frith, 2006; Watkins & Dayan, 1992). Specifically, we updated stimulus values (Q) on each trial according to the following model:

$$Q_s(t+1) = Q_s(t) + \alpha\delta(t) \quad (\text{for all stimuli A, B, C, D, E, and F}),$$

where $\delta(t) = r(t) - Q_s(t)$ is the difference between the outcome at time t (1 = *reward*, 0 = *no reward*) and the current expected stimulus value.

The second reinforcement-learning model included an additional bias parameter (a_I) that altered the integration of feedback following choices of the instructed stimulus, enabling instruction-consistent feedback to be amplified (multiplying positive prediction errors, δ_+, which were set to zero on negative prediction error trials, by the bias parameter), and instruction-inconsistent feedback to be diminished (dividing negative prediction errors, δ_-, which were set to zero on positive prediction error trials, by the bias parameter) (Doll et al., 2011, 2009). For the instruction bias reinforcement-learning model, stimulus values were updated as follows:

$$Q_s(t+1) =$$
$$Q_s(t) + \alpha\delta(t) \quad (\text{for uninstructed stimuli A, B, C, D, and E})$$
$$Q_s(t+1) =$$
$$Q_s(t) + \alpha_I\alpha\delta_+(t) + \alpha\delta_-(t)/\alpha_I \text{ (for instructed stimulus F)}.$$

For both models, the final stimulus values were then fit to participants' test-phase choices, with each trial modeled using the softmax choice rule:

$$P_{s1}(t) = \{1 + exp[-\beta(Q_1 - Q_2)]\}^{-1},$$

where the inverse temperature parameter (β) describes how deterministic an individual's choices are, given the difference in Q values. Parameter estimates were compared at the group level using nonparametric tests. Model fits were compared to one another using the Akaike information criterion (Akaike, 1974).

We took the Beta(1.1, 1.1) distribution as a prior for the learning rate parameter (α), and the Gamma(1.2, 5) distribution as a prior for both the bias (α_I) and inverse temperature (β) parameters. These priors were chosen to be uninformative over the previously observed ranges of parameter estimates in similar tasks and to ensure smooth parameter boundaries (Daw, 2011; Daw, Gershman, Seymour, Dayan, & Dolan, 2011).

Results

Learning phase

During the learning phase (Fig. 2), we found a significant difference in performance by age group ($\chi^2 = 6.96$, $df = 2$, $p = .031$): Children performed significantly worse than adults [log-odds difference = −0.547, 95 % confidence interval (CI) (−1.018, −0.161), $p = .015$], and showed a trend toward worse performance than adolescents [log-odds difference = −0.357, CI (−0.721, 0.082), $p = .095$], but adolescents did not differ from adults [log-odds difference = −0.190, CI (−0.627, 0.198), $p = .40$]. Performance also differed significantly depending on the stimulus pair ($\chi^2 = 29.7$, $df = 2$, $p < .0001$). Performance was significantly better for the easy uninstructed pair (AB 80/20) than for the falsely instructed pair (EF 70/30) [log-odds difference = 1.046, CI (0.721, 1.387), $p < .005$], and marginally better than for the hard uninstructed pair (CD 70/30) [log-odds difference = 0.183, CI (−0.001, 0.338), $p = .055$]. Performance for the easy uninstructed pair (CD 70/30) was also significantly better than that for the falsely instructed pair (EF 70/30) [log-odds difference = 0.863, CI (0.573, 1.182), $p < .005$]. We also observed a significant linear improvement in performance across the learning phase ($\chi^2 = 36.1$, $df = 1$, $p < .0001$) [log-odds estimate = 0.297, CI (0.202, 0.378)].

The linear improvement in performance across the learning phase differed by age groups ($\chi^2 = 7.45$, $df = 2$, $p = .024$). Children showed slower improvement in performance than adults [log-odds difference = −0.283, CI (−0.502, −0.066), $p < .005$], and marginally slower improvement than adolescents [log-odds difference = −0.183, CI (−0.363, 0.037), $p = .080$], and no difference was apparent between adolescents and adults [log-odds difference = −0.100, CI (−0.309, 0.109), $p = .344$]. The linear improvement in performance also differed by stimulus pair ($\chi^2 = 2.53$, $df = 2$, $p = .009$). There was a slower improvement in performance for both the easy uninstructed pair (AB) [log-odds difference = −0.265, CI (−0.397, −0.135), $p = .010$] and the hard uninstructed pair (CD) [log-odds difference = −0.169, CI (−0.300, −0.034), $p = .030$] than for the falsely instructed pair (EF), suggesting that performance quickly reached asymptote in the uninstructed pairs. However, there was no difference between the two uninstructed pairs [log-odds difference = −0.096, CI (−0.208, 0.016), $p = .110$]. We found a marginally significant age-group-by-stimulus-pair interaction effect ($\chi^2 = 9.33$, $df = 4$, $p = .053$): Whereas adults performed better than children and adolescents on uninstructed pair choices (AB/CD), their performance decreased for the instructed pair (EF), particularly when compared to adolescents (Fig. 2). Finally, we found no evidence of a pair-by-age-by-trial interaction ($\chi^2 = 2.89$, $df = 4$, $p = .576$).

Test phase

Replicating previous results (Doll et al., 2009), during the test phase, adults showed a bias toward the instructed stimulus (F) when it was part of the equally valued DF stimulus pair

[30:30 instructed; $t = 5.98$, $df = 25$, $p < .0001$, mean = 81.4 %, CI (70.6, 92.2)]. However, children only showed a marginal effect [$t = 1.98$, $df = 29$, $p = .0573$, mean = 60.8 %, CI (49.6, 71.9)], and adolescents showed no effect of instructions [$t = 0.66$, $df = 30$, $p = .515$, mean = 53.8 %, CI (42.1, 65.4)] (Fig. 3A). Children [$t = -2.64$, $df = 54.0$, $p = .011$, percent difference = −20.7, CI (−35.8, −5.5)] and adolescents [$t = -3.56$, $df = 55.0$, $p = .0008$, percent difference = −27.6, CI (−43.2, −12.1)] chose the instructed stimulus significantly less than did adults, and no difference in preference emerged between children and adolescents [$t = 0.92$, $df = 59.0$, $p = .36$, percent difference = 7.0 %, CI (−8.8, 22.8)]. This provided an initial indication of a persistent instruction bias in adults that was absent in children and adolescents.

The computed instruction bias score (Fig. 3B) assesses choice preferences across a broader set of pairs that are equally valued but differentially instructed. Again, we found that adults were more biased than children [$t = 2.24$, $df = 51.8$, $p = .029$; instruction bias difference = 0.179, CI (0.018, 0.340)] and adolescents [$t = 2.29$, $df = 51.0$, $p = .026$; instruction bias difference = 0.182, CI (0.026, 0.341)], and that children were not differently biased from adolescents [$t = 0.002$, $df = 58.7$, $p = .998$; instruction bias difference = 0.003, CI (−.284, .295); Fig. 3C]. This pattern of age group differences was also present for each individual subcomponent of the bias score (80/30 and 30/20). The effect of age group remained significant when gender was included as a predictor of instruction bias, with only adults showing a significant bias ($p = .0005$). There was also a significant gender interaction for the adult group ($p = .0133$), but not for children ($p = .40$) or adolescents ($p = .25$). This effect was due to adult females having a larger instruction bias than adult males did [$t = 2.57$, $df = 24$, $p = .017$; instruction bias difference = 0.298, CI (0.059, 0.537)]. We also found that the instruction bias increased linearly with age ($r = .246$, $p = .021$; Fig. 3D).

We examined the choices for the pairs that measured the instruction bias (Fig. 3B)—pairs AD 80/30, AF 80/30 instructed, DB 30/20, and FB 30/20 instructed—using a mixed-effects regression with pair difficulty, instruction, age group, and all interactions as independent variables. As expected, we found a significant pair-by-instruction interaction effect (i.e., an instruction bias) on performance ($\chi^2 = 23.19$, $df = 1$, $p < .0001$) [log-odds estimate = −0.510, CI (−0.729, −0.302)], indicating that the false instruction impaired performance for the otherwise easy pair (80/30) and improved performance for the otherwise hard pair (30/20). Additionally, this instruction bias effect differed across age groups ($\chi^2 = 11.59$, $df = 2$, $p = .0031$), mirroring the effects seen in the instruction bias score. Children and adolescents were equally unaffected by instructions [log-odds difference = 0.024, CI (−0.428, 0.471), $p = .945$], whereas adults were significantly more affected by instructions than were children [log-odds difference = −0.793, CI (−1.311, −0.278), $p < .002$] and adolescents [log-odds difference = −0.769, CI (−1.325, −0.267), $p = .008$]. We found no other significant effects of pair or instruction (both $ps > .6$).

We observed an overall difference in age group optimal choices for these pairs ($\chi^2 = 16.58$, $df = 2$, $p = .0003$), reflecting age-related differences in overall probabilistic learning, independent of instructions. Children [log-odds difference = −1.310, CI (−1.963, −0.665), $p < .002$] and adolescents [log-odds difference = −0.868, CI (−1.486, −0.275), $p = .006$] chose

less optimally than did adults, but not differently from one another [log-odds difference = −0.443, CI (−1.024, 0.115), p = .108].

Performance on all uninstructed pairs (combinations of A, B, C, and D stimuli) showed that all age groups performed better than chance [children: log-odds estimate = 0.436, CI (0.128, 0.744), p = .006; adolescents: log-odds estimate = 0.674, CI (0.367, 0.981), p < .0001; adults: log-odds estimate = 1.28, CI (0.932, 1.621), p < .0001]. Children [log-odds difference = −0.840, CI (−1.302, −0.378), p = .0004] and adolescents [log-odds difference = −0.603, CI (−1.063, −0.143), p = .010] performed less well than adults and were not significantly different from each other [log-odds difference = −0.238, CI (−0.672, 0.197), p = .284], demonstrating age differences in experiential learning similar to those evident in the learning-phase uninstructed choices.

During both the learning and test phases, response times were unrelated to choice, instruction, age group, or any of their interactions. Response times decreased significantly over the learning phase (response time effect = −51.4 ms, SEM = 7.75 ms, χ^2 = 35.58, df = 1, p < .0001), with no difference by age groups. During the test phase, response times were significantly longer for the hard (30:20) than for the easy (80:20) pairs, regardless of instruction (response time difference = 244.8 ms, SEM = 38.5 ms) [$F(1, 93.58)$ = 9.58, p = .0026]. No other significant effects on test-phase response times were apparent.

Reinforcement learning (RL)

To explore the process by which instructions might influence the integration of feedback during the learning phase, we fit participants' test-phase choices using a standard and modified instruction bias RL model. The standard RL model describes a feedback-driven learning process that has been proposed to underlie experiential reward learning (Bayer & Glimcher, 2005; Pessiglione et al., 2006; Watkins & Dayan, 1992). The instruction bias RL model adds a bias parameter that amplifies the influence of instruction-consistent outcomes and diminishes instruction-inconsistent feedback, yielding an instruction bias (Doll et al., 2009). The standard RL model is equivalent to an instance of the instruction bias RL model in which the bias parameter is set to 1 (i.e., no bias). Model comparison based on the median AIC values for each age group (see Table 1) indicated that the choices of children and adolescents were fit better by the standard RL model, whereas adults were fit better by the modified model that included an instruction bias parameter. This result suggests that children and adolescents recruited an undistorted feedback-based integration process during the learning phase, but adults biased the integration of feedback for the instructed stimulus, altering the weighting of positive and negative outcomes in an instruction-consistent manner.

In both the standard and instruction bias RL models, age group differences in the estimated learning rate parameters (α) suggested that children were more influenced by recent outcomes than were adolescents or adults (Kruskal–Wallis standard: H = 12.87, p = .002; bias: H = 8.82, p = .013; see Table 1 for group comparisons). The softmax inverse temperature parameter (β), which reflects how deterministically a participant used learned stimulus values to make test-phase choices, also showed increases across age groups (standard: H = 16.26, p = .0003; bias: H = 18.71, p < .0001; see Table 1 for group

comparisons). These age differences in choice consistency paralleled the differences in test-phase performance observed for the experientially learned stimuli in the regression results. In the instruction bias RL model, bias parameter estimates (α_I) exhibited the same pattern of age group differences observed for both the instruction bias score and the regression analysis ($H = 14.25$, $p = .0009$). Adults' bias parameter estimates were significantly higher than both children's ($W = 156.5$, $p = .0002$) and adolescents' ($W = 237$, $p = .021$), with no difference appearing between children and adolescents ($W = 430.5$, $p = .62$).

Collectively, these modeling results suggest qualitative differences as a function of age in the manner in which instructions influenced experiential feedback-based learning. Whereas instructions biased the integration of feedback during learning for adults, both children and adolescents were less influenced by instructions, integrating feedback in a relatively unbiased manner.

Discussion

In this study, we examined whether the influence of instructions on experiential learning changes across development. Despite age differences in performance, children, adolescents, and adults were all able to recruit experiential feedback to learn to preferentially select the higher-valued stimulus of each uninstructed pair. In all age groups, choices for the instructed pair were initially biased toward the inaccurately recommended stimulus and gradually (rapidly in adolescents) shifted toward the higher-valued alternative stimulus as participants received continued negative feedback. However, performance during the test phase suggested marked qualitative differences across development in how instructions influenced the processing of this experiential feedback. Consistent with previous findings (Doll et al., 2009; Staudinger & Buchel, 2013), we found that adults showed a strong instruction-consistent bias, suggesting that inaccurate instruction distorted their feedback-based value learning. In contrast, both children and adolescents showed a minimal influence of instruction on test-phase performance, suggesting that they integrated positive and negative feedback more objectively during the learning phase in order to estimate the value of the instructed stimulus. These data suggest that when explicit instruction or advice conflicts with experiential feedback about the value of an action, children and adolescents weight their own experience more heavily.

Our analyses of instruction bias focused on decisions made during the test phase, in which participants' novel choices revealed the value estimated for each stimulus through the integration of learning-phase feedback. In contrast, choices during the learning phase can reflect potentially heterogeneous evaluation strategies adopted by participants (e.g., hypothesis testing across multiple trials; Doll et al., 2011; Frank et al., 2007), which may obscure current stimulus value estimates. Past studies employing variants of this task suggested that test-phase choices might provide the most reliable indication of learning and are selectively sensitive to various pharmacological, genetic, psychological, and neurological factors thought to alter the incremental experiential-learning process (Doll et al., 2011, 2009; Frank et al., 2007; Frank & O'Reilly, 2006; Frank et al., 2004). In our study, age group differences in the influence of instruction were not evident in choices made during the learning phase. However, we saw robust evidence of an instruction bias in the

test-phase choices of adults, but not of children and adolescents. Our reinforcement-learning analyses establish a link between feedback received during the learning phase and test-phase decisions, by formalizing the potential underlying processes for learning stimulus values. Crucially, reinforcement-learning parameters for the initial learning phase were fit to the test-phase choices that reflected the final learned stimulus values. These analyses suggested that whereas experienced outcomes during the learning phase were objectively weighted in children's and adolescents' value estimates, adults biased the weighting of outcomes for the instructed stimulus to be more consistent with the explicit instruction that they had received.

Experiential learning is thought to depend critically on dopaminergic prediction errors, through which the striatum can learn the value of an action (McClure et al., 2003; O'Doherty et al., 2003; Pagnoni et al., 2002; Schultz, Dayan, & Montague, 1997). Explicit instruction is proposed to bias this striatal learning process through the top-down influence of the pre-frontal cortex (Biele et al., 2009, 2011; Doll et al., 2011, 2009, 2014; Li, Delgado, & Phelps, 2011; Staudinger & Buchel, 2013), which enables task-relevant rules and instructions to influence goal-directed behavior (Miller & Cohen, 2001). A theoretical model supported by our reinforcement-learning analyses (Doll et al., 2009) posits that the prefrontal cortex amplifies the effect of instruction-consistent outcomes and diminishes the influence of instruction-inconsistent outcomes on the striatal learned values. This produces a behavioral "confirmation bias," through which recommended actions are more highly valued than those learned solely through experience, even when the recommendation is inaccurate. Previous studies examining the instructional control of experiential value learning in adults have largely supported this model, demonstrating both the hypothesized alteration of striatal feedback-driven error signals (Biele et al., 2011) and a correlation between instruction-guided choice outcomes and pre-frontal cortex activation (Li et al., 2011). Collectively, this evidence suggests that functional interaction between the pre-frontal cortex and the striatum may have mediated the instructional biasing of learning that we observed in our adult participants.

By extension, the relative absence of instructional influence on experiential learning in children and adolescents may stem from the reduced functional efficacy of prefrontal–striatal pathways prior to adulthood. Functional imaging studies have revealed intact striatal prediction error signals from childhood onward (Galvan et al., 2006; van den Bos et al., 2012), consistent with evidence of feedback-based experiential reward learning across development (Cohen et al., 2010; Peters, Braams, Raijmakers, Koolschijn, & Crone, 2014; van den Bos et al., 2012; van den Bos, Güro lu, van den Bulk, Rombouts, & Crone, 2009). In contrast, both structural and functional connectivity between the prefrontal cortex and the striatum exhibit marked changes from childhood through adulthood (Imperati et al., 2011; Liston et al., 2006). Cognitive functions that depend on the integrity of this neural pathway typically show continued maturation into adulthood (Liston et al., 2006; Rubia et al., 2006; Somerville & Casey, 2010), suggesting that developmental changes in frontostriatal connectivity may facilitate information exchange between these regions. On the basis of the neuroscientific model of instructional control of learning in adulthood, we hypothesize that the prolonged maturation of prefrontal–striatal connectivity underlay the resistance of children and adolescence to the biasing effects of inaccurate instruction in our task. Our present study focused solely on behavior. However, we expect that a functional imaging

study of our task might show that adults exhibit an instruction-consistent bias in striatal prediction error signals for choices of the instructed stimulus during the learning phase, with positive signals being amplified and negative signals diminished relative to those for the equally valued uninstructed stimulus. We expect that this biased signaling in adults would be accompanied by greater prefrontal–striatal connectivity following instructed than following uninstructed choice outcomes. In contrast, we expect that children and adolescents would show no such differences in prediction error signals for instructed versus uninstructed stimuli. These specific hypotheses about the potential neural substrates of our behavioral results could be tested directly in a subsequent developmental neuroimaging study.

The maturational increase in instructional influence on learning that we observed in this study concurs with a broader literature suggesting a gradual developmental emergence of cognitive control (Bunge & Zelazo, 2006; Diamond, 2006; Munakata, Snyder, & Chatham, 2012). A primary challenge of cognitive development is to acquire knowledge across a variety of stimulus domains about the nature of the environment, which is accomplished in large part through inductive statistical learning. Such experientially acquired knowledge may be more flexibly applied and easily generalized than explicit rule-based learning, a principle that has long been recognized in pedagogical theory (Hayes, 1993; Kolb, 1984). Implicit learning processes typically recruit evolutionarily conserved subcortical regions including the basal ganglia (Bischoff-Grethe, Goedert, Willingham, & Grafton, 2004; Rauch et al., 1997). Such learning is evident early in development (Amso & Davidow, 2012; Kirkham, Slemmer, & Johnson, 2002; Saffran, Aslin, & Newport, 1996) and may continue to improve into adulthood (Thomas et al., 2004). Although reduced prefrontal control is often portrayed as a developmental handicap, it may confer distinct advantages by enabling implicit learning to occur unhindered (Thompson-Schill, Ramscar, & Chrysikou, 2009). Providing instruction, whether false or veridical, has been shown to interfere with multiple forms of implicit experiential learning, reducing task performance relative to when no instruction is given (Reber, 1989). Increased sensitivity to underlying patterns in the reinforcement of actions may facilitate children's and adolescents' acquisition of language, social conventions, and other complex behaviors.

An important consideration not addressed in this study is whether the social source of instruction might modulate its influence. In the present study, the instruction provided to participants was simply presented on the screen, lacking any specific social origin. In contrast, real-world advice often comes from peers (friends, classmates, and colleagues) or authority figures (parents, teachers, and bosses), which may yield different effects on behavior than a printed message does. The source of advice may be a particularly important factor during adolescence—a period of increasing independence and heightened sensitivity to peers (Chein, Albert, O'Brien, Uckert, & Steinberg, 2011; Galvan, 2014; Gardner & Steinberg, 2005; Jones et al., 2014; Steinberg & Monahan, 2007). Moreover, the influence of instruction has been shown to depend on the perceived expertise of the advisor (Meshi, Biele, Korn, & Heekeren, 2012), and peers and authority figures may be viewed as experts in different behavioral domains at different developmental stages. Thus, advice from different social sources may vary in salience across both age groups and decision contexts. Future work might explore whether manipulating the social source of instruction would alter the developmental differences in instruction bias reported here.

Both parents and policymakers commonly rely on rules and instruction to deter children and adolescents from actions that carry potentially harmful consequences. In particular, increased independence during adolescence often presents opportunities to experiment with behaviors (e.g., sex, drug experimentation, or reckless driving) that frequently yield positive social or hedonic outcomes, but that can have rare yet serious negative effects. Positive experienced outcomes may come to predominate in adolescents' risk estimates (Reyna & Farley, 2006). In our study, participants of all ages initially adhered to the instruction, consistent with other evidence that the actions and decisions of adolescents can be influenced by advice (Engelmann, Moore, Capra, & Berns, 2012). However, when the feedback they received provided evidence contradictory to their prior instruction, both children and adolescents, but not adults, showed greater reliance upon their own experience. Public policy campaigns attempting to deter adolescents from risky behavior through explicit guidance or information have had limited efficacy (Ennett, Tobler, Ringwalt, & Flewelling, 1994; Trenholm et al., 2007). The present results suggest a cognitive mechanism underlying such resistance to instruction. This finding highlights the importance of research aimed at identifying effective ways for both parents and public health campaigns to advise adolescents as they navigate real-world risky behavioral domains (Reyna & Farley, 2006).

In summary, by placing instruction and experience in competition, we have shown here that the relative weighting of these two sources of information shifts over the course of development. Consistent with the protracted maturation of the circuitry implicated in instructional control of learning, children and adolescents showed less influence of instruction on choice than did adults. Whereas instruction alters the processing of experiential feedback in adults, our results suggest that children and adolescents remain attuned to the true reward contingencies in their environment, enabling experience to prevail in directing their actions. Many aspects of cognition (e.g., working memory; Crone, Wendelken, Donohue, van Leijenhorst, & Bunge, 2006), attentional control (Rueda, Posner, & Rothbart, 2005), and executive function (Diamond, 2006) improve as individuals mature from childhood through adulthood, typically conferring advantages for adults in learning and decision-making. Similarly, the effective recruitment of instruction to guide one's actions may generally be advantageous, allowing an individual to benefit from the knowledge and prior experience of others. However, our results suggest that this ability may also come at the cost of introducing pronounced bias in the processing of experiential feedback. The absence of confirmation biases in the children and adolescents observed in this study represents a paradoxical developmental advantage of youth over adults in the unbiased evaluation of actions through positive and negative experience.

Acknowledgments

References

Akaike H. A new look at the statistical model identification. IEEE Transactions on Automatic Control, AC-19. 1974:716–723. doi:10.1109/TAC.1974.1100705.

Amso D, Davidow J. The development of implicit learning from infancy to adulthood: Item frequencies, relations, and cognitive flexibility. Developmental Psychobiology. 2012; 54:664–673. [PubMed: 22714674]

Barr DJ, Levy R, Scheepers C, Tily HJ. Random effects structure for confirmatory hypothesis testing: Keep it maximal. Journal of Memory and Language. 2013; 68:255–278. doi:10.1016/j.jml. 2012.11.001.

Bates, D.; Maechler, M.; Bolker, B.; Walker, S. lme4: Linear mixed-effects models using Eigen and S4 [Software]. 2014. Retrieved from http://CRAN.R-project.org/package=lme4

Bayer HM, Glimcher PW. Midbrain dopamine neurons encode a quantitative reward prediction error signal. Neuron. 2005; 47:129–141. [PubMed: 15996553]

Biele G, Rieskamp J, Krugel LK, Heekeren HR. The neural basis of following advice. PLoS Biology. 2011; 9:e1001089. doi:10.1371/journal.pbio.1001089. [PubMed: 21713027]

Biele G, Rieskamp J, Gonzalez R. Computational models for the combination of advice and individual learning. Cognitive Science. 2009; 33:206–242. doi:10.1111/j.1551-6709.2009.01010.x. [PubMed: 21585468]

Bischoff-Grethe A, Goedert KM, Willingham DT, Grafton ST. Neural substrates of response-based sequence learning using fMRI. Journal of Cognitive Neuroscience. 2004; 16:127–138. [PubMed: 15006042]

Bunge SA, Zelazo PD. A brain-based account of the development of rule use in childhood. Current Directions in Psychological Science. 2006; 15:118–121.

Chein J, Albert D, O'Brien L, Uckert K, Steinberg L. Peers increase adolescent risk taking by enhancing activity in the brain's reward circuitry. Developmental Science. 2011; 14:F1–F10. [PubMed: 21499511]

Cohen JR, Asarnow RF, Sabb FW, Bilder RM, Bookheimer SY, Knowlton BJ, Poldrack RA. A unique adolescent response to reward prediction errors. Nature Neuroscience. 2010; 13:669–671.

Crone EA, Wendelken C, Donohue S, van Leijenhorst L, Bunge SA. Neurocognitive development of the ability to manipulate information in working memory. Proceedings of the National Academy of Sciences. 2006; 103:9315–9320.

Daniel R, Pollmann S. Comparing the neural basis of monetary reward and cognitive feedback during information-integration category learning. Journal of Neuroscience. 2010; 30:47–55. [PubMed: 20053886]

Daw, ND. Trial-by-trial data analysis using computational models. In: Delgado, MR.; Phelps, EA.; Robbins, TW., editors. Decision making, affect, and learning: Attention and performance XXIII. Oxford University Press; New York, NY: 2011. p. 3-38.

Daw ND, Gershman SJ, Seymour B, Dayan P, Dolan RJ. Model-based influences on humans' choices and striatal prediction errors. Neuron. 2011; 69:1204–1215. [PubMed: 21435563]

den Ouden HE, Daw ND, Fernandez G, Elshout JA, Rijpkema M, Hoogman M, Cools R. Dissociable effects of dopa-mine and serotonin on reversal learning. Neuron. 2013; 80:1090–1100. [PubMed: 24267657]

Diamond, A. The early development of executive functions. In: Bialystok, E.; Craik, F., editors. Lifespan cognition: Mechanisms of change. Oxford University Press; New York, NY: 2006. p. 70-95.

Doll BB, Hutchison KE, Frank MJ. Dopaminergic genes predict individual differences in susceptibility to confirmation bias. Journal of Neuroscience. 2011; 31:6188–6198. [PubMed: 21508242]

Doll BB, Jacobs WJ, Sanfey AG, Frank MJ. Instructional control of reinforcement learning: A behavioral and neurocomputational investigation. Brain Research. 2009; 1299:74–94. [PubMed: 19595993]

Doll BB, Waltz JA, Cockburn J, Brown JK, Frank MJ, Gold JM. Reduced susceptibility to confirmation bias in schizophrenia. Cognitive, Affective, & Behavioral Neuroscience. 2014; 14:715–728. doi:10.3758/s13415-014-0250-6.

Engelmann JB, Moore S, Capra CM, Berns GS. Differential neurobiological effects of expert advice on risky choice in adolescents and adults. Social Cognitive and Affective Neuroscience. 2012; 7:557–567. [PubMed: 22563008]

Ennett ST, Tobler NS, Ringwalt CL, Flewelling RL. How effective is drug abuse resistance education? A meta-analysis of Project DARE outcome evaluations. American Journal of Public Health. 1994; 84:1394–1401. [PubMed: 8092361]

Frank MJ, Moustafa AA, Haughey HM, Curran T, Hutchison KE. Genetic triple dissociation reveals multiple roles for dopamine in reinforcement learning. Proceedings of the National Academy of Sciences. 2007; 104:16311–16316.

Frank MJ, O'Reilly RC. A mechanistic account of striatal dopamine function in human cognition: Psychopharmacological studies with cabergoline and haloperidol. Behavioral Neuroscience. 2006; 120:497–517. doi:10.1037/0735-7044.120.3.497. [PubMed: 16768602]

Frank MJ, Seeberger LC, O'Reilly RC. By carrot or by stick: Cognitive reinforcement learning in parkinsonism. Science. 2004; 306:1940–1943. doi:10.1126/science.1102941. [PubMed: 15528409]

Galvan A. Insights about adolescent behavior, plasticity, and policy from neuroscience research. Neuron. 2014; 83:262–265. [PubMed: 25033176]

Galvan A, Hare TA, Parra CE, Penn J, Voss H, Glover G, Casey BJ. Earlier development of the accumbens relative to orbitofrontal cortex might underlie risk-taking behavior in adolescents. Journal of Neuroscience. 2006; 26:6885–6892. [PubMed: 16793895]

Gardner M, Steinberg L. Peer influence on risk taking, risk preference, and risky decision making in adolescence and adulthood: An experimental study. Developmental Psychology. 2005; 41:625–635. doi:10.1037/0012-1649.41.4.625. [PubMed: 16060809]

Hayes RL. A facilitative role for counselors in restructuring schools. Journal of Humanistic Education and Development. 1993; 8:156–162.

Imperati D, Colcombe S, Kelly C, Di Martino A, Zhou J, Castellanos FX, Milham MP. Differential development of human brain white matter tracts. PLoS ONE. 2011; 6:e23437. doi:10.1371/journal.pone.0023437. [PubMed: 21909351]

Jones RM, Somerville LH, Li J, Ruberry EJ, Powers A, Mehta N, Casey BJ. Adolescent-specific patterns of behavior and neural activity during social reinforcement learning. Cognitive, Affective, & Behavioral Neuroscience. 2014; 14:683–697. doi:10.3758/s13415-014-0257-z.

Kirkham NZ, Slemmer JA, Johnson SP. Visual statistical learning in infancy: Evidence for a domain general learning mechanism. Cognition. 2002; 83:B35–B42. [PubMed: 11869728]

Kolb, DA. Experiential learning: Experience as the source of learning and development. Prentice-Hall; Englewood Cliffs, NJ: 1984.

Li JA, Delgado MR, Phelps EA. How instructed knowledge modulates the neural systems of reward learning. Proceedings of the National Academy of Sciences. 2011; 108:55–60.

Liston C, Watts R, Tottenham N, Davidson MC, Niogi S, Ulug AM, Casey B. Frontostriatal microstructure modulates efficient recruitment of cognitive control. Cerebral Cortex. 2006; 16:553–560. [PubMed: 16033925]

McClure SM, Berns GS, Montague PR. Temporal prediction errors in a passive learning task activate human striatum. Neuron. 2003; 38:339–346. [PubMed: 12718866]

Meshi D, Biele G, Korn CW, Heekeren HR. How expert advice influences decision making. PLoS ONE. 2012; 7:e49748. doi:10.1371/journal.pone.0049748. [PubMed: 23185425]

Miller EK, Cohen JD. An integrative theory of prefrontal cortex function. Annual Review of Neuroscience. 2001; 24:167–202. doi:10.1146/annurev.neuro.24.1.167.

Munakata Y, Snyder HR, Chatham CH. Developing cognitive control: Three key transitions. Current Directions in Psychological Science. 2012; 21:71–77. [PubMed: 22711982]

O'Doherty JP, Dayan P, Friston K, Critchley H, Dolan RJ. Temporal difference models and reward-related learning in the human brain. Neuron. 2003; 38:329–337. [PubMed: 12718865]

Pagnoni G, Zink CF, Montague PR, Berns GS. Activity in human ventral striatum locked to errors of reward prediction. Nature Neuroscience. 2002; 5:97–98.

Pessiglione M, Seymour B, Flandin G, Dolan RJ, Frith CD. Dopamine-dependent prediction errors underpin reward-seeking behaviour in humans. Nature. 2006; 442:1042–1045. [PubMed: 16929307]

Peters S, Braams BR, Raijmakers ME, Koolschijn PC, Crone EA. The neural coding of feedback learning across child and adolescent development. Journal of Cognitive Neuroscience. 2014; 26:1705–1720. doi:10.1162/jocn_a_00594. [PubMed: 24564463]

Rauch SL, Whalen PJ, Savage CR, Curran T, Kendrick A, Brown HD, Rosen BR. Striatal recruitment during an implicit sequence learning task as measured by functional magnetic resonance imaging. Human Brain Mapping. 1997; 5:124–132. [PubMed: 10096417]

Reber AS. Implicit learning and tacit knowledge. Journal of Experimental Psychology: General. 1989; 118:219–235. doi:10.1037/0096-3445.118.3.219.

Reyna VF, Farley F. Risk and rationality in adolescent decision making implications for theory, practice, and public policy. Psychological Science in the Public Interest. 2006; 7:1–44.

Rodriguez P, Aron A, Poldrack R. Ventral–striatal/nucleus-accumbens sensitivity to prediction errors during classification learning. Human Brain Mapping. 2006; 27:306–313. [PubMed: 16092133]

Rubia K, Smith AB, Woolley J, Nosarti C, Heyman I, Taylor E, Brammer M. Progressive increase of frontostriatal brain activation from childhood to adulthood during event-related tasks of cognitive control. Human Brain Mapping. 2006; 27:973–993. doi:10.1002/hbm.20237. [PubMed: 16683265]

Rueda RM, Posner MI, Rothbart MK. The development of executive attention: Contributions to the emergence of self-regulation. Developmental Neuropsychology. 2005; 28:573–594. [PubMed: 16144428]

Saffran JR, Aslin RN, Newport EL. Statistical learning by 8-month-old infants. Science. 1996; 274:1926–1928. doi:10.1126/science.274.5294.1926. [PubMed: 8943209]

Schultz W, Dayan P, Montague PR. A neural substrate of prediction and reward. Science. 1997; 275:1593–1599. [PubMed: 9054347]

Somerville LH, Casey BJ. Developmental neurobiology of cognitive control and motivational systems. Current Opinion in Neurobiology. 2010; 20:236–241. doi:10.1016/j.conb.2010.01.006. [PubMed: 20167473]

Staudinger MR, Buchel C. How initial confirmatory experience potentiates the detrimental influence of bad advice. NeuroImage. 2013; 76:125–133. [PubMed: 23507392]

Steinberg L, Monahan KC. Age differences in resistance to peer influence. Developmental Psychology. 2007; 43:1531–1543. doi:10.1037/0012-1649.43.6.1531. [PubMed: 18020830]

Thomas KM, Hunt RH, Vizueta N, Sommer T, Durston S, Yang Y, Worden MS. Evidence of developmental differences in implicit sequence learning: An fMRI study of children and adults. Journal of Cognitive Neuroscience. 2004; 16:1339–1351. [PubMed: 15509382]

Thompson-Schill SL, Ramscar M, Chrysikou EG. Cognition without control when a little frontal lobe goes a long way. Current Directions in Psychological Science. 2009; 18:259–263. [PubMed: 20401341]

Trenholm, C.; Devaney, B.; Fortson, K.; Quay, K.; Wheeler, J.; Clark, M. Impacts of four title V, section 510 abstinence education programs: Final report. Mathematica Policy Research, Inc; Princeton, NJ: 2007.

van den Bos W, Cohen MX, Kahnt T, Crone EA. Striatum–medial prefrontal cortex connectivity predicts developmental changes in reinforcement learning. Cerebral Cortex. 2012; 22:1247–1255. doi:10.1093/cercor/bhr198. [PubMed: 21817091]

van den Bos W, Güro lu B, van den Bulk BG, Rombouts SA, Crone EA. Better than expected or as bad as you thought? The neurocognitive development of probabilistic feedback processing. Frontiers in Human Neuroscience. 2009; 3(52):1–11. doi:10.3389/neuro.09.052.2009. [PubMed: 19255629]

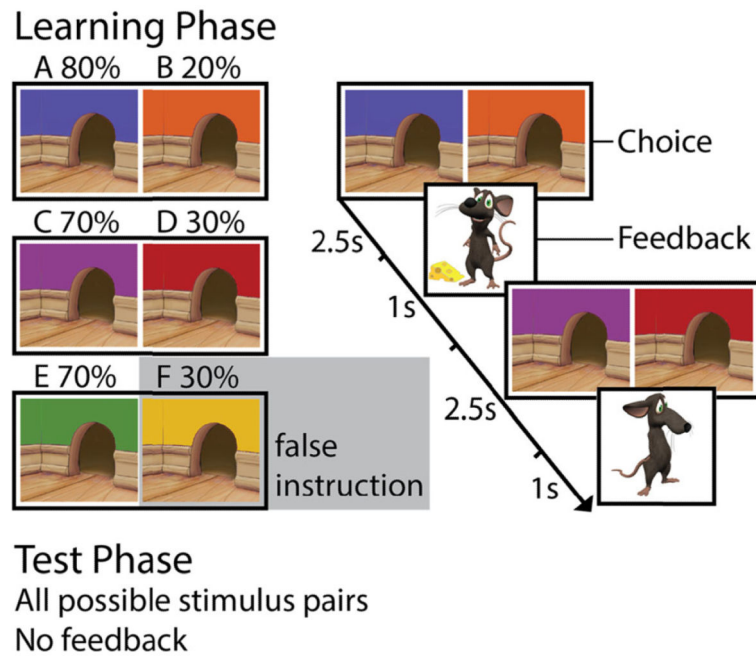Watkins CJ, Dayan P. Q-learning. Machine Learning. 1992; 8:279–292.

Fig. 1.
Probabilistic-learning paradigm. The learning phase consisted of 180 choices between six probabilistically reinforced stimuli presented in three pairs. Participants were falsely instructed that one stimulus had a high likelihood of being rewarded, when in actuality it did not. Positive or negative feedback was given following each trial. The test phase, consisting of all 15 possible stimulus pairs with no feedback, enabled assessment of the extent to which the learned stimulus values were biased by the instruction
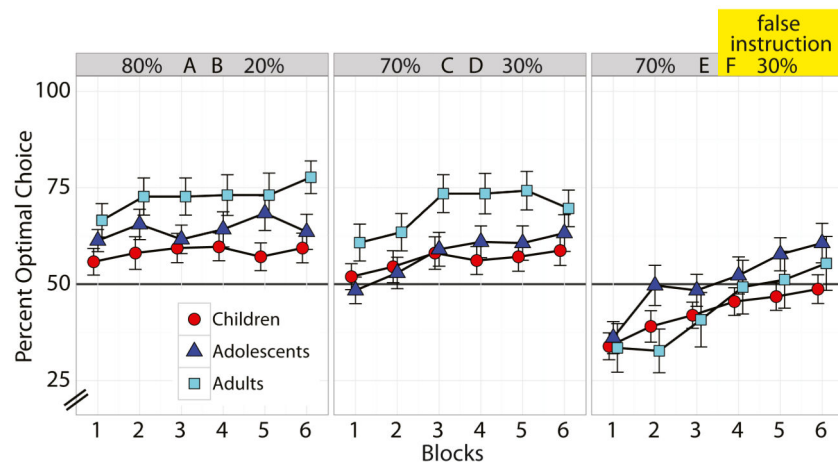
Fig. 2.

Learning-phase performance. Adults (*cyan squares*) performed better than children (*red circles*) and marginally better than adolescents (*blue triangles*) for the two uninstructed pairs (AB, CD). All groups initially adhered to the false instruction (F) and gradually learned through experience to select the higher-valued alternative (E), with adolescents showing the fastest improvement. Error bars represent *SEM*s
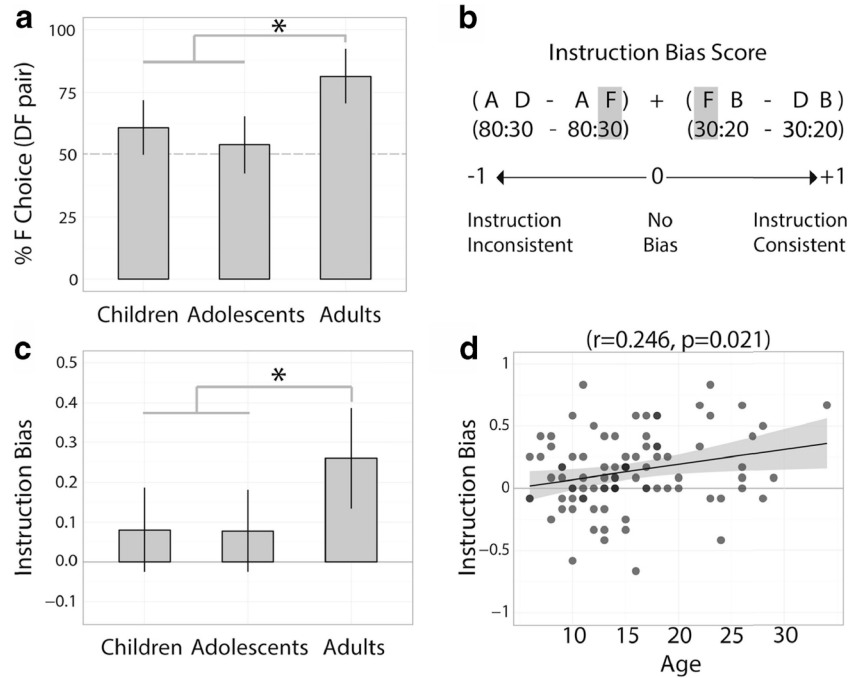
Fig. 3.

Test-phase performance. (A) Percentages of F choices when seeing the DF pair (30:30 instructed). (B) The instruction bias was the average of two difference scores between stimulus pairs of equal probability that were differentially instructed. (C) Adults showed a significantly larger instruction bias than did both children and adolescents. (D) A significant increase in instruction bias occurred with age (darker circles indicate two data points). Error bars (panels A and C) are 95 % confidence intervals; shading (panel D) indicates *SEM*s

Table 1

Reinforcement learning model parameter fits

Model	Parameter	Median			KW	MWW		
		Child	Adolescent	Adult	Chi-Square (*df* = 2)	Child vs. Adolescent	Child vs. Adult	Adolescent vs. Adult
Standard RL	α	.429	.083	.036	*H* = 12.87 *p* = .002	*W* = 579 *p* = .10	*W* = 621 *p* = .0002	*W* = 504 *p* = .11
	β	0.80	1.25	3.89	*H* = 16.26 *p* = .0003	*W* = 361 *p* = .14	*W* = 150 *p* < .0001	*W* = 238 *p* = .009
	AIC	131.2	125.2	117.2				
Modified bias RL	α	.289	.046	.054	*H* = 8.82 *p* = .013	*W* = 622 *p* = .023	*W* = 565 *p* = .004	*W* = 396 *p* = .92
	β	1.32	2.89	4.55	*H* = 18.71 *p* < .0001	*W* = 316 *p* = .031	*W* = 127 *p* < .0001	*W* = 262 *p* = .024
	α_I	1.42	1.44	3.31	*H* = 14.25 *p* = .0009	*W* = 430.5 *p* = .62	*W* = 156.5 *p* = .0002	*W* = 237 *p* = .021
	AIC	135.8	128.3	100.0				

α, learning rate. β, softmax parameter, α_I, bias parameter, AIC, Akaike information criterion. The nonparametric Kruskal–Wallis (KW) and Mann–Whitney–Wilcoxon (MWW) tests were used to compare the group parameter estimates