Cognitive Psychology 150 (2024) 101650

Contents lists available at ScienceDirect

Cognitive Psychology

journal homepage: www.elsevier.com/locate/cogpsych

The structure and development of explore-exploit decision making

Madeline B. Harms [a,*], Yuyan Xu [a], C. Shawn Green [a], Kristina Woodard [a], Robert Wilson [b], Seth D. Pollak [a]

[a] Department of Psychology, University of Wisconsin – Madison, 1202 West Johnson Street, Madison, WI 53706, United States
[b] Department of Psychology, University of Arizona, 1503 E. University Blvd. (Building 68), Tucson, AZ 85721, United States

ARTICLE INFO

Keywords:
Exploration
Exploitation
Adolescence
Decision-making
Principal component analysis

ABSTRACT

A critical component of human learning reflects the balance people must achieve between focusing on the utility of what they know versus openness to what they have yet to experience. How individuals decide whether to explore new options versus exploit known options has garnered growing interest in recent years. Yet, the component processes underlying decisions to explore and whether these processes change across development remain poorly understood. By contrasting a variety of tasks that measure exploration in slightly different ways, we found that decisions about whether to explore reflect (a) random exploration that is not explicitly goal-directed and (b) directed exploration to purposefully reduce uncertainty. While these components similarly characterized the decision-making of both youth and adults, younger participants made decisions that were less strategic, but more exploratory and flexible, than those of adults. These findings are discussed in terms of how people adapt to and learn from changing environments over time. Data has been made available in the Open Science Foundation platform (osf.io).

1. Introduction

It is common for people to be in situations that require them to decide between a familiar option with a known value or to choose a new option with unknown (but perhaps advantageous) value. Examples of this type of decision problem include choosing between selecting a familiar meal at the cafeteria versus trying a new food; choosing between staying with a familiar peer group versus pursuing a different social opportunity; or sticking with a current job versus making an employer or career change. Decisions involve tradeoffs, and the optimal choice is often not clear at the time a decision is made. Familiar options might afford less stress, anxiety, and avoiding an outcome that is worse than the status quo. However, staying with the familiar may prevent a person from discovering and learning new information about the world. Acquisition of new information is particularly important in childhood and adolescence.

For this reason, the development of healthy decision-making involves flexibility navigating between exploration and exploitation, depending on the context and relative risks involved (Hills et al., 2015; Mekern et al., 2019; Mehlhorn et al., 2015; Schulz & Gershman, 2019; Wilson et al., 2021). Despite an emerging literature that is examining how humans learn to manage these decisions (Addicott et al., 2017; Giron et al., 2022; Gopnik, 2020; Meder et al., 2021; Schulz et al., 2019; Somerville et al., 2017; Wilson et al., 2014), little is currently understood about the specific cognitive processes underlying these behaviors or the extent to which these processes change across development. Here, we identify processes that contribute to explore-exploit decision making using multiple common measures

* Corresponding author at: Department of Psychology University of Minnesota Duluth, 320 BohH, 1207 Ordean Court, Duluth, MN 55812, United States.

E-mail address: harms124@d.umn.edu (M.B. Harms).

https://doi.org/10.1016/j.cogpsych.2024.101650
Received 4 January 2023; Received in revised form 5 March 2024; Accepted 6 March 2024
Available online 10 March 2024
0010-0285/© 2024 Elsevier Inc. All rights reserved.

in the field and examine whether these components change between early adolescence and adulthood.

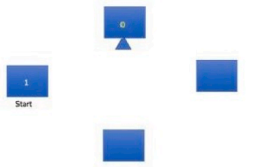
1.1. Measuring exploration and exploitation

We define exploration as seeking new information, and exploitation as utilizing existing knowledge at the expense of learning something new. Both processes can be deployed to seek rewards; however, these fundamental motivations are often, though not always, opposed to one another. Laboratory paradigms designed to measure explore-exploit decision-making are similar in that they create scenarios where an individual must choose to either explore or exploit on a given trial. But, as shown in Fig. 1, these paradigms vary in their emphasis on factors such as working memory, cognitive flexibility, learning from previous outcomes, and uncertainty tolerance (Gershman, 2018; van den Bos & Hertwig, 2017; von Helversen et al., 2018).

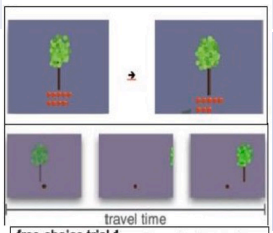
Two related components of exploration/exploitation involve learning about which aspects of an environment will be rewarded and then keeping track of the likelihood and magnitude of those various rewards. A class of methods that tap these aspects of explore-exploit decision making are called *bandit tasks* (Daw et al., 2006; Wilson et al., 2014); see Fig. 1D. In this type of task, individuals choose between several "bandits" (e.g., slot machines) that vary in the rewards they pay out. Individuals learn through exploration which bandits seem most profitable, allowing them to maximize their rewards. In bandit tasks, explore and exploit decisions probably require similar effort, since each subsequent choice could reveal new key information that should affect the next decision, making the tasks less susceptible to response sets. But they also rely heavily upon working memory because an individual must keep track of both the amount of information they have gathered about each bandit and the magnitude of rewards received, while ignoring irrelevant information (Brown, Hallquist, Frank, & Dombrovski, 2022). Dimensions such as uncertainty can be manipulated in these paradigms

A

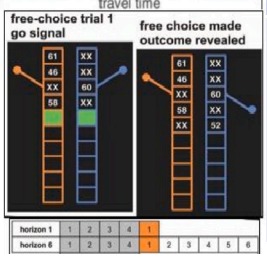
B



C



D



	Behavioral Choice	Key Variables	Other Demands
Grid Task Type of Paradigm: Sequential search Cost of exploration: Loss of choices	Choose between new or previously explored boxes to earn points	Key Measure: # of unique locations searched	Generalize from earlier trials to predict reward distributions; cognitive flexibility to switch between explore and exploit (Dale, Sampers, Loo, & Green, 2018)
Chain Task Type of Paradigm: Sequential search Cost of exploration: Loss of choices	Participants choose between exploring a new space (which all end up having 0 points) or returning to the starting location, where 1 point is consistently rewarded.	Key Measure: # of unique locations searched	Generalize from earlier trials to predict reward distributions; cognitive flexibility to switch between explore and exploit (Dale, Sampers, Loo, & Green, 2018)
Orchard Task Type of Paradigm: Patch foraging Cost of exploration: Time	Participants harvest apples from trees with a gradually dwindling supply. At any point, they may choose to travel to a new tree, which costs time, but may have a replenished supply of apples.	Key Manipulation: Travel time Key Measure: Exit Threshold: Point at which participant decides to move to a new patch	Cognitive flexibility to switch between move and stay strategies (Lloyd, McKay, & Furl, 2022; Hills & Dukas, 2012)
Horizon Task Type of Paradigm: Bandit Cost of exploration: None	After learning about rewards associated with two bandits, participants choose a high or low information option.	Key Manipulation: Uncertainty Key Measures: Directed Exploration: Changes in information-seeking Random Exploration: Decision noise	Working memory needed to track reward patterns and suppress irrelevant information (Brown, Hallquist, Frank, & Dombrovski, 2022).

Fig. 1. Depiction of the four explore-exploit tasks. A. Grid task: Participants choose between unexplored boxes (represented by question marks) or exploiting previously visited boxes to earn points. The number of points in a box is revealed and added to the participant's total when the box is clicked. B. Chain task: Participants choose between exploring a new space (which all end up having 0 points) or returning to the starting location, where 1 point is consistently rewarded. C. Orchard task: Participants harvest apples from trees with a gradually dwindling supply. At any point, they may choose to travel to a new tree, which costs time, but may have a replenished supply of apples. D. Horizon task. After learning about rewards associated with two bandits through four forced choice decisions, participants have the option of choosing a high or low information option on their fifth choice (in this example, this is the bandit on the right). Games last either five or ten trials.

by making bandit payoffs more or less variable, and by providing more or less information before participants make a choice (Gershman, 2018).

The ability to generalize from prior experiences is another component of decisions about whether to explore. This type of learning is captured in *sequential choice tasks* (Dale, Sampers, Loo, & Green, 2018; von Helversen et al., 2018); see Fig. 1 A/B. Here, participants explore options with varying rewards until they find one with a sufficiently high payoff that leads them to choose to exploit this option for the rest of the task (some tasks allow an individual to go back to a previous option; others do not). Typically, participants are given minimal information about the highest payout available. Therefore, these types of tasks require that individuals generalize from previous outcomes to determine an expected range or distribution of reward outcomes (Dale et al., 2018). Thus, the nature of prior expectations regarding the task environment influences how individuals behave across these situations (von Helversen et al., 2018).

Patch foraging types of tasks require individuals to gather resources in various "patches" (Constantino & Daw, 2022; Lenow et al., 2017); see Fig. 1C. These types of tasks may have a higher demand on cognitive flexibility than others because on a trial-by-trial basis individuals need to flexibly navigate the environment by switching between exploration and exploitation strategies (Hills & Dukas, 2012). They do so by choosing to continue exploiting their current location or deciding to leave in order to explore a new location. This paradigm introduces the tension between exploring and exploiting in two ways. First, the value of a current patch diminishes as it is exploited: For example, in a task that emulates foraging in an apple orchard, continuing to pick apples from the same tree results in fewer apples in that tree available for subsequent picking. Yet, there is a cost in time associated with moving to a new patch because no apples can be picked while searching for a new tree. Typically, the total duration of the game is fixed; thus, to maximize reward earning within a finite amount of time, one needs to increase the harvest per time unit. The harvest per time unit will drop if one either switches too often or stays with one patch for too long. As such, one must strike a balance between staying and switching. Importantly, this 'sweet spot' is dependent on one's estimate of the average reward rate in the given environment – one should increase switching in a generous environment and reduce switching in a scarcer environment. So, the explore-exploit trade-off in this foraging task is not just about information seeking, but also about flexibly adjusting one's decisions based on the estimate of "richness" of the environment and avoiding getting "stuck" in an exploitation pattern (i.e., staying at the same location too long). A final feature of this type of task structure is that the least effortful behavior in a foraging environment is to exploit the current patch, whereas a decision to move to a new patch may require greater effort or motivation (see Fig. 1).

Although these various approaches all purport to measure the same construct, they differentially tap processes that may contribute to exploration. Working memory, prior expectations, and cognitive flexibility likely play some role in each of these tasks, but the degree to which each is relevant differs based on task structure. Here, we harness these task differences to examine the structure and development of exploration.

1.2. Developmental differences in exploration/exploitation

Evolutionary theories propose that childhood is a period of learning about the world via exploration (Gopnik, 2020). Indeed, numerous studies show that young children explore more than adults during computerized explore-exploit tasks (Blanco & Sloutsky, 2021; Giron et al., 2022; Schulz et al., 2019; Sumner et al., 2019). Moreover, the complexity and efficiency of children's exploration increases from early childhood to early adolescence (Pelz & Kidd, 2020), and exploration becomes less random and more directed towards reducing uncertainty from age four to nine (Meder et al., 2021). However, less is known about developmental change in explore/exploit decision-making between early adolescence and adulthood. Using a paradigm (Horizon task; Wilson et al, 2017) that could mathematically separate explore-exploit decisions into two components—random exploration (gathering information by chance) and directed exploration (intentional exploring to reduce uncertainty)–Somerville et al., (2017) found that early adolescents and adults were equally likely to engage in random exploration. However, younger individuals engaged in less *strategic directed* exploration to intentionally reduce uncertainty, as compared with older adolescents and adults. Lower directed exploration among the youngest age group was partially explained by a preference for immediate reward over information gathering. Contrary to other work suggesting that younger children explore more than adults, this study suggests that adolescents might explore *less* than adults (in some contexts) because of heightened reward drive and/or impulsivity. However, these conclusions are limited because they are derived from just a single task with a particular structure. Here, we investigated the concepts of random and directed exploration across multiple task structures.

At the same time, there is converging evidence that adolescence is a developmental period characterized by *qualitative* changes in decision-making strategies (Hartley & Somerville, 2015; Shulman et al., 2016). Heightened risk-taking in adolescence often occurs in circumstances where the probability of positive versus negative outcomes is unknown. Adolescents appear to be comfortable with taking risks when they perceive outcomes as highly uncertain (Tymula et al., 2012). This tendency has been reflected in *lower* levels of sampling and information search prior to decision during a bandit task among adolescents, relative to children and adults (van den Bos & Hertwig, 2017). In addition, children and adolescents tend not to perform as well as adults during probabilistic reward tasks, such as the Iowa Gambling Task, which, like bandit tasks, require information search combined with weighing potentials for risk and reward (Almy et al., 2018; Cassoti et al., 2014). Inferring from developmental trends in these similar tasks, we might expect adolescents to show *less* information-driven exploration than adults in contexts such as the highly structured bandit task used by Somerville et al, but, similar to patterns seen in younger children, *more* exploration in ambiguous environments where an optimal strategy is less clear. Conceptual differences between random and directed exploration will be further explained in the next section.

1.3. Current study

The current study is the first to addresses two broad issues about exploratory behavior. First, we examined the structure of explore-exploit decision-making by contrasting prominent paradigms in the extant literature that differentially rely on a variety of cognitive processes. We hypothesized that explore-exploit decisions would be comprised of two components that vary depending upon the availability of information available to the learner. The first would resemble a noisy or random form of exploration that emerges when information availability is low. An individual engaging in random exploration explores without a specific goal in mind, but because it is more interesting or engaging to sample something new. We hypothesized that this type of exploration would be reflected in performance on sequential decision-making tasks where an individual's behavior is dependent upon expectations about the environment, novelty seeking, and tendencies to generalize from previous experience.

The second component, directed exploration, would reflect a more intentional, goal-directed gathering of information that we predicted would be reflected in the patch foraging context. As with other task structures, the goal in this task is to obtain the highest total reward possible. This form of exploration relies upon cognitive flexibility, as an individual's hypotheses about reward availability must be constantly updated and revised in light of higher degrees of available information (richness of the environment, depletion rate, switch costs, etc). As found by Wilson et al (2014), we reasoned that decisions in a bandit task would contain elements of both random and directed exploration due to a moderate degree of uncertainty, but also substantial available information.

The second issue we tested concerned developmental differences in exploration between early adolescence (age 10–13) and young adulthood. We focused on early adolescence because this period involves increasing exposure to novel social environments (e.g., transitioning from elementary to middle school; forming new peer groups) that afford opportunities to explore. Because executive processes like working memory and cognitive flexibility undergo substantial development from early adolescence to early adulthood (Ferguson, Brunsdon & Bradford, 2021), and previous work shows different exploration strategies in adolescents vs. adults, we hypothesized that the structure (sub-components) involved in exploration described above could change with age. For example, prior research shows that some cognitive constructs (e.g., sub-components of executive function) change in their level of differentiation with development (Howard, Okely, & Ellis, 2015). Finally, we predicted that age-related differences in exploration would depend on the task environment: we predicted that early adolescents would engage in less directed exploration during a bandit task than adults, replicating Somerville et al. (2017); but that younger individuals would engage in more exploration during sequential decision-making tasks that involve high uncertainty (and lend themselves to random exploration). Because exploration is tied to learning, we reasoned that the information gathered from our more comprehensive approach to studying developmental differences in explore-exploit decision making would have implications for the types of environments where adolescents can learn most efficiently.

2. Material and methods

Below, we report on our participants and procedure, along with how we determined our sample size, all data exclusions, all manipulations, and our overall analysis plan.

2.1. Participants

One hundred and twenty-one participants are included in this study. Sixty-two early adolescents ages 10–13 years ($M = 11.1$, $SD = 0.81$, 32 female) and 59 young adults ages 18–32 years ($M = 19.3$, $SD = 2.63$, 45 female) were recruited from the Madison metropolitan area to participate. The total sample size was determined based on available resources and aimed to be similar to previous published developmental studies examining exploration (e.g., Somerville et al., 2017; van den Bos & Hertwig, 2017), which included about 105–150 participants. Among the early adolescent sample, 74 % identified as white, 2 % as Black, 3 % as Hispanic, 11 % as mixed, 7 % as Asian/Pacific Islander, and 3 % did not report. Among the adult sample, 73 % identified as white, 24 % identified as Asian, and 3 % did not report. Parents of child participants and young adult participants gave informed consent for their child/themselves, and the university IRB approved all procedures. To keep children engaged in the tasks, child participants were told that earning more points would be tied to a more desirable prize. In the end, all children were given a small gift of their choice for participating, and parents were paid $20. Young adult participants were recruited through a Psychology participant pool and flyers around campus, and were compensated with their choice of $10 cash or extra credit in their Psychology course. Analysis code, raw data, and stimuli are available at https://osf.io/ytc3n/.

2.2. Procedure

Participants completed four computerized tasks in a counterbalanced order, followed by a test of fluid cognition (digit span for early adolescents or Wechsler Abbreviated Scale of Intelligence (WASI) matrix reasoning subscale for adults) to examine the extent to which generalized cognitive ability may influence decision-making behavior. The four explore-exploit tasks are summarized in Fig. 1.

Grid task. This sequential decision-making task was adapted from Dale, Sampers, Loo, & Green (2018). The screen was divided into a 20 × 20 Grid of rectangular boxes. Each box initially contained 3 question marks which, when clicked, revealed a point value underneath (see Fig. 1A). Point values were generated randomly but programmed such that participants encountered the same sequence of values for each *unique location* clicked, regardless of where they clicked (e.g., the third new box clicked would yield value × regardless of where it was located). In other words, there was no spatial correlation among the boxes. Participants were informed that each box held a consistent reward (i.e., if they went back to a box they had selected previously, they would get the same reward as

before). The values simulated a normal distribution and were then exponentiated to produce a log-normal distribution (with some very high values in the tail of the distribution). 25 % of the boxes were randomly set to 0.

Participants were told that each box contained some number of points but were not given any information regarding the reward distribution or the highest reward available. The experimenter demonstrated seven unique box clicks, followed by three repeated box clicks, selecting the highest value observed, to reveal that clicking a given box repeatedly yielded the same point value each time it was selected. Participants had the option to click as many different boxes as they liked within 200 trials and could return to high reward boxes as many times as they liked. In other words, for all 200 trials, participants had the choice to either explore a new box or return to any previously selected box. At the end of each trial, participants could see their current total score, as well as the number of trials remaining. There was no time limit for this task, or any other tasks. The exploration score for the Grid task was calculated by summing the total number of unique box clicks across all trials.

Chain Task. Also a sequential choice task, but one that is a purer measure of persistence in the context of limited reward, this task was adapted from Dale et al., (2018; Wolpert & Macready, 1997). Participants were presented with a single yellow box with the number '0′ in the center. At the top of the screen, a counter indicated the number of turns left, and a second counter indicated the number of points earned. Participants were not provided with any further information regarding reward values available. If participants pressed the 'C' key, they would remain in the initial box and the '0′ in the box would change to a '1′ (i.e., they would be awarded with a single point). If they pressed the 'I' key, there was an 80 % chance of moving to a new box elsewhere on the screen (unbeknownst to the participant, all new boxes awarded 0 points). The remaining 20 % of the time, the participant was moved to the first box and was awarded one point (see Fig. 1B). After 7 boxes, if the participant continued pressing 'I', they would progress through the same sequence again. Thus, if the participant pressed the 'I' key while in the first box, there was an 80 % chance of moving to the second box and a 20 % chance of staying in the first box. If the participants pressed the 'I' key while in the second box, there was an 80 % chance of moving to the third box and a 20 % chance of moving back to the first box, and so forth until they reached the seventh box. Participants were given 100 turns in total. Their exploration score was calculated as the total number of times that they clicked the 'I' key. Thus, participants who tended to exploit the small reward that came from pressing the 'C' key received a low exploration score, and those who persisted in pressing the unrewarded 'I' key had high exploration scores.

Orchard task. In the Orchard patch foraging task (Constantino & Daw, 2022; Lenow et al., 2017), participants spent 14 min harvesting apples in a series of four Orchards (see Fig. 1C). Participants were told they should try to collect as many apples as possible, and that apples would later be converted to points. The initial supply of apples at each try was randomly drawn from a Gaussian distribution with mean of 10 and standard deviation of 1. The apple supply at each tree gradually dwindled with repeated harvests, by a mean of 0.88 ($sd = 0.07$) for each successive harvest. At each trial, participants chose via key press to either continue harvesting at their current tree (exploit) or move to a different tree (explore). Within each Orchard, the "travel time" between trees was either long (12 s) because trees were far apart, or short (6 s), reflecting differing levels of opportunity cost for moving (since participants could not travel and harvest apples at the same time). Travel time conditions were counterbalanced in an ABAB/BABA block design. Correspondingly, participants were informed that in some Orchards trees were spread out, so it would take longer to walk to a new tree, and in other orchards trees were closer together. In all Orchards, the "harvest" time was 3 s, and the mean depletion rate was 0.88 (each harvest yielded 88 % of the apples in the previous harvest). Participants were required to harvest at least once at each tree before continuing. The dependent variable was the average of the last two harvests before moving to a different tree–the "exit threshold." A low exit threshold indicated a higher exploration rate, while a high exit threshold indicated a lower exploration rate. We excluded the first exit threshold in each block, given that participants did not know if the travel time was "short" or "long" until they traveled to a new tree. Subsequent exit thresholds were then averaged across each environment type (short or long travel time) to calculate the total exploration score.

Horizon task. In the Horizon task (Wilson et al., 2014), participants chose between two one-armed bandits that paid out differing point values (see Fig. 1D). In contrast to the grid task, participants were informed that each bandit was relatively consistent in its payoff amount within each game. After selecting a bandit, participants saw only the points awarded by their chosen bandit. Each game was either five or ten choices in length (Horizon), and the computer determined the first four selections (the participant had to select the bandit that was highlighted in order to continue). Thus, games were categorized as H1 (one free choice) or H6 (6 free choices). The "forced" choices controlled which information participants were exposed to before making their first free choice, which was the main dependent variable. Within the first four choices, the computer always selected one bandit three times and the other bandit once, so that there was an imbalance of information about the two bandits (this differed from the original task, and was done to shorten the task). Choosing the bandit with fewer previous payouts (i.e., "high information choices") was categorized as exploration because by doing so, the participant would gain more information about the payoff amounts of the two bandits. Participants played 80 games in total. Each game was a fixed length, so participants choices did not influence the number of trials they completed. As expected, participants learned to choose the high mean option and to decrease information seeking as each game progressed (see Online Supplement for further information and figures depicting this manipulation check).

Other variables within the task–mean payout amount and Horizon length–influenced the relative advantage of exploring in each game, and yielded information about participants' exploration strategies. We distinguished between directed exploration (intentional information search) and random exploration (random decision noise that was not value-based, i.e., random). In each Horizon condition, directed exploration was quantified in a model-based manner as the "information bonus" – the additional value given to the more informative option. Random exploration was quantified as the decision noise, which can be used to compute the effective probability of choosing the low-mean option (in the original paper using this task (Wilson et al., 2014) the decision noise parameter was used as additional confirmation of the presence of random exploration). The equation for the random exploration model is below, stating that the probability of choosing left is a softmax function of Delta R (difference in observed mean reward between left and right

options), Delta I (difference in information between left and right options; +1 if left is more informative, i.e. played less in forced trials; −1 if right is more informative), A (information bonus), B (side bias), Sigma (decision noise).

This model allows fitting of noise (sigma) separate from information bonus (A).

$$p(left) = \frac{1}{1 + \exp(\frac{\Delta R + A\Delta I + B}{\sigma})}$$

Although in our version of the Horizons task, all trials presented "unequal information," random exploration can be calculated from the slopes of choice curves across all trials in a game. Graphs of choice curves showing random and directed exploration are provided in the Supplemental Materials. Directed exploration is reflected by a shift in slope, while random exploration is indicated by a flatter slope (Feng et al., 2021).

For ease of interpretation, Model-based measures were then standardized from 0 to 1. We used choices in H1 games as baseline measures of random and directed exploration tendencies, and the first choice in H6 games as a measure of more strategic, information driven exploration (since information learned in these games could be used in subsequent trials). This resulted in four measures of exploration: H1 directed, H6 directed, H1 random, and H6 random.

2.3. Analysis plan

For each task, we planned to remove any participants from analysis who had a neurological condition or who showed statistically extreme patterns (>2 *SD*s above or below the mean) of nearly always exploring or exploiting, as these patterns could reflect misunderstanding of the task goals. We first planned to examine the extent to which exploration task performance was correlated with general cognitive ability (WASI or digit span). Then we planned to analyze main effects of age group, task, and their interaction in the key measure(s) of exploration using a mixed ANOVA and post-hoc independent *t*-tests. For the Horizon and Orchard task, we conducted additional mixed ANOVAs to examine effects of condition by age within these tasks. Finally, we addressed the main hypothesis of the study—that explore-exploit decision-making would be explained by separable components–by exploring the relationships between the four tasks for adolescents and adults using Principal Components Analysis (PCA). PCA is a data reduction technique that orthogonally transforms observed data into linear combinations (components) that capture most of the variance in the data (Jolliffe, 2002). It is most suitable for when correlations between measures are relatively high (>0.3), as observed in the current study. Bivariate correlations were also included to provide additional detail regarding relationships between each of the exploration measures. Analyses were completed using SPSS version 25 and R version 4.3.0.

3. Results

In the results below, we first report overall effects of age group and task, followed by findings at the level of individual tasks to examine differing patterns of decision making between youth and adults. Then we report findings regarding bivariate correlations among tasks, and finally, findings regarding the overall structure of explore-exploit decision making. Data from one adolescent participant diagnosed with autism were not included in analyses.

3.1. Exploration and general cognitive ability

To examine the extent to which general cognitive ability explained individual variation in exploration, we examined bivariate correlations between standardized WASI/digit span and measures of exploration. Digit span/WASI were negatively correlated with Horizon random exploration for adolescents and adults (r = -0.39, p =.002). In addition, WASI was positively correlated with Tree task exploration for adults only (r = 0.34, p =.008), while digit span had a marginal negative correlation with grid exploration among adolescents only (r = -0.25, p <.06). There were no significant correlations between WASI/digit span and chain task exploration or directed exploration in Horizon.

Table 1
Task performance (Mean & SD) by age group.

	Youth	Adults
Grid: # of boxes	116 (39.4)	102 (50.4)+
Chain: # of moves	40.7 (23.1)	37.6 (23.6)
Orchard: short travel time exit threshold	7.48 (1.8)	7.59 (1.7)
Orchard: long travel time exit threshold	6.75 (2.2)	7.04 (2.08)
Horizon: random exploration H1	0.17 (0.11)	0.17 (0.11)
Horizon: random exploration H6	0.25 (0.11)	0.25 (0.10)
Horizon: directed exploration H1	0.48 (0.07)	0.51 (0.07)*
Horizon: directed exploration H6	0.51 (0.10)	0.56 (0.11)*

Note: *indicates significant group difference (p <.05). + indicates marginal group difference (p <.1).

3.2. Effects of age group and task in explore-exploit decision making

Table 1 lists descriptive statistics by age group for the primary dependent variables in each task. A two (age group) by five (exploration measure) mixed ANOVA on Z-scored exploration scores revealed no main effects of age group or task. However, there was a significant age group x task interaction, $F(4,440) = 3.76$, $p = .005$. Simple effects *t*-tests indicated a significant age group difference in directed exploration in the Horizon task, $F(1,110) = 9.1$, $p = .003$, but no significant effects of age in other exploration measures.

Grid task. Early adolescents ($M = 116$, $SD = 39.4$) explored more than adults ($M = 101$, $SD = 50.4$), although the effect did not reach significance, $t(117) = 1.7$, $p < .1$, $d = 0.3$; See Fig. 2. No outliers were identified in this task.

Chain task. One child and one adult who moved on every trial (i.e., never exploited) were removed from analysis. No significant effects of age group were found, $t(114) = 0.57$, $p = .28$.

Orchard Task. Participants who left a tree after harvesting just once over 80 % of the time or who never explored a new tree in two or more Orchards were identified as statistical outliers and removed; these criteria resulted in two adolescent and two adult exclusions.

An additional 2 (travel time) x 2 (age group) mixed ANOVA was performed to examine group differences in exploration (average exit threshold) in the Orchard task. As expected, there was a main effect of travel time, $F(1,116) = 32.3$, $p < .001$, $\text{eta}^2 = 0.22$, indicating that participants chose to move between trees more frequently in the short travel time condition, where the "cost" of exploring a new tree was lower. However, there was no significant effect of age group or interaction between age group and travel time. We also examined foraging behavior in terms of the maximum value threshold (MVT) for each condition, which reflects the exit threshold that would maximize the long-run average reward rate (i.e., when the expected number of apples from one more harvest is smaller than the number of apples that one would expect on average). This value was 6.52 in the short travel time condition and 5.31 in the long travel time condition. With reference to these thresholds, one-sample *t*-tests showed that participants in both age groups explored with higher than optimal frequency, which reduces the harvest per time unit. (means = 6.75–7.59; $ps < 0.06$ - <0.001); See Fig. 2.

Horizon task. Two follow-up 2 (horizon: long or short) x 2 (age group) mixed ANOVAs tested 1) directed exploration and 2) random exploration (decision noise) in the Horizon task. As expected, participants engaged in more directed exploration in H6 games (when they were afforded more free choices) than in H1 games (which afforded fewer choice options), $F(1,118), = 22.7$, $p < .001$, $\text{eta}^2 = 0.18$. Adults engaged in more directed exploration than adolescents, $F(1,118), = 7.55$, $p < .01$, $\text{eta}^2 = 0.10$. There was no significant age group x horizon interaction for directed exploration. Participants also engaged in more random exploration in H6 games versus H1 games, $F(1,118), = 78.0$, $p < .001$, $\text{eta}^2 = 0.40$. There was no significant main effect of age group or age group x horizon interaction for random exploration. See supplement for more details and manipulations checks regarding the Horizon task; see Fig. 2.

3.3. Relationships among tasks

Tables 2 and 3 summarize the correlations (Pearson R) for each age group across tasks. Given strong positive correlations between exploration among short and long travel time conditions on the Orchard task ($r = 0.82$, $p < .001$), we averaged those variables in subsequent analyses. In both age groups, exploration during the chain and grid tasks were significantly correlated, and random exploration was significantly correlated with grid exploration.

3.4. Components of explore-exploit decision making

Next, we conducted a Principal Component Analysis (PCA) with varimax rotation to examine the factor structure of combined

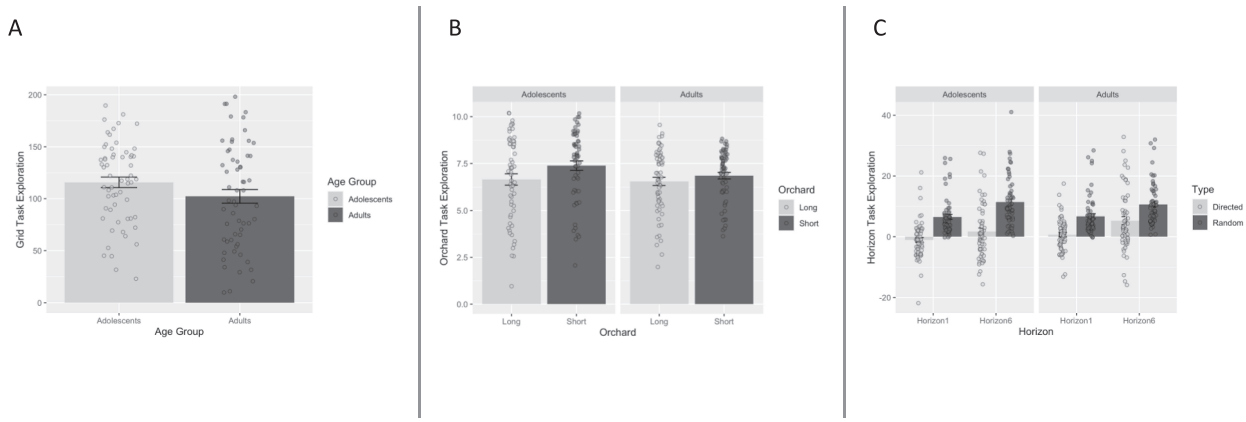


Fig. 2. Exploration during the Grid, Orchard, and Horizon task by age group. Panel A: number of unique boxes clicked in Grid task. Panel B: Orchard task exploration calculated via mean exit threshold. Panel C: Directed and Random Exploration in the Horizon task. Exploration values represented by information bonus (directed exploration) and noise (random exploration).*Note*: Error bars indicate +/- 1 SE. Dots represent individual participants.

Table 2
Correlation matrix for explore/exploit measures among adolescents.

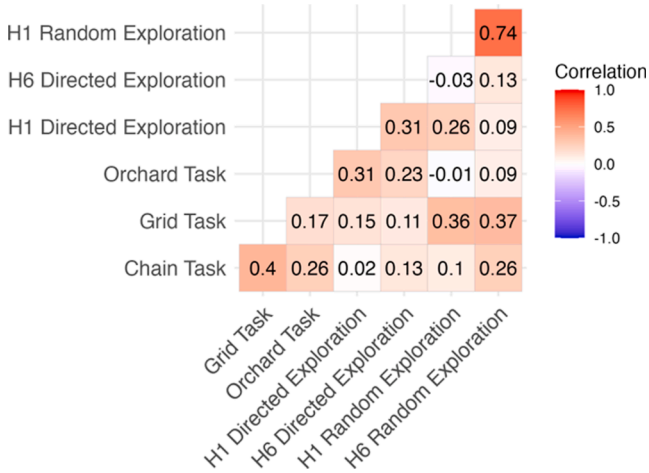
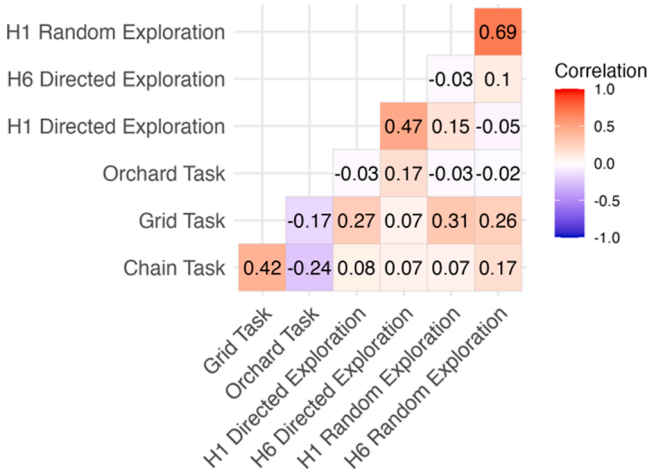


Table 3
Correlation Matrix for explore/exploit measures among adults.



performance on each exploration task. PCA was performed using the primary outcome measure of exploration for each task: (1) Orchard task average exit threshold, (2) Grid task total number of unique boxes clicked, (3) Chain task total number of moves, (4) Horizon directed exploration, and (5) Horizon random exploration (for the last two variables, we took the combined average for H1 and H6 games, which were strongly positively correlated; including all four variables separately for Horizon would lead to a less reliable PCA model due to an imbalance in the number of measures from each task). All these variables were standardized before analysis. To detect potential group differences in factor structure, we conducted this analysis separately in each age group. Given our sample size, it was inadvisable to examine age group as a separate variable in the PCA. As shown in Table 4, PCA revealed a similar composition of underlying factors among youth and adults.

Among early adolescents, PCA revealed that performance across tasks converged on two components; see Table 4. The first component accounted for 37.7 % of the variance across all tasks, and was composed of grid, chain, and Horizon random exploration. The second component accounted for 21.7 % of the variance and was composed of Horizon directed exploration and Orchard task exit threshold. The same component compositions were found among adults, with Component 1 accounting for 33.9 % of variance and Component 2 accounting for 22.7 % of variance. Additional information, including Scree plots, is available in the online supplement.

Table 4
Rotated Component Matrices for early adolescents and adults.

Early Adolescents	Component 1	Component 2
Grid # boxes	**0.82**	0.11
Chain # moves	**0.69**	−0.07
Orchard Mean Exit Threshold	0.09	**0.82**
Horizon Directed Exploration	0.07	**0.74**
Horizon Random Exploration	**0.76**	−0.07

Adults	Component 1	Component 2
Grid # boxes	**0.83**	0.06
Chain # moves	**0.74**	−0.12
Orchard Mean Exit Threshold	−0.34	**0.68**
Horizon Directed Exploration	0.19	**0.83**
Horizon Random Exploration	**0.53**	−0.03

Note: The primary loading for each measure is bolded.

4. Discussion

We sought to examine the structure across the types of explore-exploit decision making tasks used most frequently in the extant literature. All four tasks were designed to measure tendencies to make explore versus exploit decisions. The tasks differed, however, in terms of which factors influenced participants' decision-making behavior. We reasoned that decision making performance differences across four tasks would be accounted for by the differential emphasis of each task on various cognitive processes and on the amount of uncertainty in the environment. We also aimed to examine effects of age on the structure of explore-exploit decision making and on exploratory behavior within each task.

Consistent with our predictions, two underlying components accounted for explore-exploit decisions. Contrary to prediction, these components did not differ between adolescents and adults. For both early adolescents and adults, the first component underlying exploratory behavior is comprised of exploration on sequential choice tasks and Horizon random exploration. This component might represent pursuit of novelty, regardless of reward value, expectations/perceptions about potential reward in the environment based on generalizing past experiences, or other unmeasured processes—all potential underlying aspects of "random exploration." The second component, comprised of exploration during patch foraging and Horizon directed exploration, might represent strategic information seeking to obtain larger rewards, subserved by cognitive flexibility and working memory.

Exploration in sequential choice tasks, such as Grid and Chain, tend to be positively correlated (Dale et al., 2018). Exploration in these tasks is influenced by the extent to which the participant believes that a very large reward may be present; if so, they should explore many different boxes or locations. Because there is so little information about the environment available, any hypothesis testing that occurs is expected to depend on an individual's tendency to generalize from past experiences and may therefore be more "noisy." Another interesting aspect of these tasks is that more exploration was associated with lower total rewards (because there was in fact no "outlier" high value in the task environment); in other words, the utility of exploration from a pure reward maximization standpoint is diminished). Speculatively, measures of exploration in the Grid and Chain tasks may be similar to random variability (decision noise) in the Horizon task. Thus, both the sequential choice and Horizon tasks require a certain amount of random exploration; while neither strategic nor information-driven, random responses are computationally less costly than directed exploration in these paradigms (Wilson et al., 2014). In contrast, Orchard task exploration relies heavily upon strategic search to maximize reward in the context of limited resources, and where information about the environment can be learned quickly. One way to interpret this task is that participants identify an optimal threshold based on the average rate of return, and switch patches once the current option falls below the estimated value of alternatives (i.e., a purely exploitative perspective). However, the fact that participants tended to leave patches earlier than optimal for reward maximization suggests the alternative interpretation that participants did in fact engage in information seeking. This was also the only task that was positively correlated with general cognitive ability, supporting an interpretation that exploration here is somewhat dependent on strategy and planning. An optimal strategy on this task requires flexibly switching between explore and exploit strategies as more information is accumulated. These features lend themselves to more strategic information search and have parallels to directed exploration in the Horizon task.

Our findings here illustrate the overall complexity of explore-exploit decision making. At the outset we defined exploration as information seeking and exploitation as seeking immediate rewards at the expense of new information. But there are several reasons why an individual might choose to explore. For example, one might make a mistake, or simply act randomly or out of boredom–these phenomena may be captured by our "random exploration" component. A person might have optimistic prior expectations about the environment, leading them to attribute a higher expected value to the novel option, regardless of information known. These alternative explanations make it difficult to measure "pure" information seeking behavior and support the use of multiple methods in assessing explore-exploit behavior.

The present data also suggest that some features of exploration change over development. Relative to adults, early adolescents

showed less directed exploration during the Horizon task, as consistent with previous research (Somerville et al., 2017), but also somewhat higher tendencies to explore in the low-information Grid task environment. This suggests that earlier development may be characterized by approaches to learning that are exploratory and flexible, but less strategic than those of adults. This phenomenon has been documented in young children, who tend to show higher rates of exploration overall, and more decision noise in their exploration relative to adults (Gopnik et al., 2017; Schulz et al., 2019). Findings that adolescents engage in less information-driven search than adults are consistent with data suggesting that this age group is more comfortable with ambiguity and uncertainty (Somerville et al., 2017; Tymula et al., 2012; van den Bos & Hertwig, 2017).

The Grid task has very low known information regarding reward distributions. Therefore our findings are consistent with the idea that early adolescence is characterized by a "resampling" of one's environment, resulting in neurobehavioral recalibration that optimizes an adolescent's behavior for their current environment, regardless of past environments (Gunnar et al., 2019). Exploratory learning at this stage of development, even if more random than information-driven, would facilitate flexible changes in decision-making, whereas more directed forms of exploration may be less flexible and become more prominent with prefrontal cortex maturation. In contrast, we found no age-group differences in the Chain task, a measure of persistence without reward; or in the Orchard task, which measured strategic allocation of limited resources. These tasks both include more reliable and consistent information about the environment, suggesting that early adolescents and adults use such information in similar ways to inform their decision-making.

4.1. Limitations and future directions.

Because PCA is by definition exploratory, our interpretation of the components yielded in this study is speculative. This study examined a novel question regarding the structure of explore-exploit decision making, and should be considered a first step in this line of inquiry. Future studies could extend knowledge in this area by using paradigms that more systematically disentangle cognitive processes such as memory, cognitive flexibility, expectations, and learning. A limitation of our study is the relatively small sample size and lack of racial diversity in our sample. Although our sample size was similar to other behavioral studies of cognitive development, a larger sample would have allowed us to examine more sophisticated factor analysis models. Future work should focus on recruiting larger and more racially and socio-economically diverse samples, given research showing differences in learning and decision-making related to family SES (e.g., Palacios-Barrios et al., 2021). Future research might also examine factors associated with individual differences in explore-exploit decision-making, such as SES, early stress history, and experience with unpredictability, as well as how explore-exploit decision-making relates to constructs such as risk-taking, reward processing, and reinforcement learning. Finally, it would be useful to examine how developmental differences in decision making contribute to aspects of psychosocial development in adolescence, such as identity formation and social learning.

4.2. Conclusions

In the current study, we examined the overall structure of explore-exploit decision making using four common paradigms. We found that patterns of decision-making were represented by two components among both youth and adults: we interpret these components to represent 1) random exploration that is not explicitly goal-directed, and 2) directed exploration to purposefully reduce uncertainty. Given this two-component model, along with differing patterns of developmental differences between tasks, it would be useful for researchers going forward to use multiple paradigms to capture multiple aspects of explore-exploit decision making; or to clearly define which aspect of this process they wish to study, and carefully choose a paradigm that aligns with that aspect. Overall, further exploring *how* we learn to explore or exploit will better our understanding of fundamental questions in developmental science regarding how individuals adapt to changing environments over time.

Funding

This work was supported by the National Institute of Mental Health (R01MH61285) to S. Pollak and a core grant to the Waisman Center from the National Institute of Child Health and Human Development (P50HD105353). M. Harms was supported by an Emotion Research Training Grant (T32MH018931-30) from the National Institute of Mental Health. M. Harms collected this data at University of Wisconsin-Madison and is now at the University of Minnesota Duluth.

CRediT authorship contribution statement

Madeline B. Harms: Writing – review & editing, Writing – original draft, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Yuyan Xu:** . **C. Shawn Green:** Methodology, Formal analysis. **Kristina Woodard:** Project administration. **Robert Wilson:** Methodology, Formal analysis. **Seth D. Pollak:** Writing – review & editing, Supervision, Resources, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

To make our work transparent to other scientists, we have deposited our experimental paradigm, de-identified data, and analysis scripts on the Open Science Framework at: https://osf.io/ytc3n/

Acknowledgements

We thank the families who participated in this study, research assistants who helped collect and analyze data, and Dr. Daniel Bolt for statistical consulting. Experimental paradigm, de-identified data, and analysis scripts are available on Open Science Framework: https://osf.io/ytc3n/. Madeline Harms ORCID: 0000-0002-6703-2571,

Author note

Data and analysis code are available at https://osf.io/ytc3n/

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cogpsych.2024.101650.

References

Addicott, M. A., Pearson, J. M., Sweitzer, M. M., Barack, D. L., & Platt, M. L. (2017). A primer on foraging and the explore/exploit trade-off for psychiatry research. *Neuropsychopharmacology, 42*(10), 1931–1939. https://doi.org/10.1038/npp.2017.108

Almy, B., Kuskowski, M., Malone, S. M., Myers, E., & Luciana, M. (2018). A longitudinal analysis of adolescent decision-making with the Iowa Gambling Task. *Developmental psychology, 54*(4), 689–702. https://doi.org/10.1037/dev0000460

Blanco, N. J., & Sloutsky, V. M. (2021). Systematic exploration and uncertainty dominate young children's choices. *Developmental Science, 24*(2), e13026.

Brown, V. M., Hallquist, M. N., Frank, M. J., & Dombrovski, A. Y. (2022). Humans adaptively resolve the explore-exploit dilemma under cognitive constraints: Evidence from a multi-armed bandit task. *Cognition, 229*, Article 105233.

Cassoti, M., Aite, A., Osmont, A., Houde, O., & Borst, G. (2014). What have we learned about the processes involved in the Iowa Gambling task from developmental studies? *Frontiers in Psychology, 5*, 915. https://doi.org/10.3389/fpsyg.2014.00915

Constantino, S. M., & Daw, N. D. (2022). Learning the opportunity cost of time in a patch-foraging task. *Cognitive, Affective & Behavioral Neuroscience, 15*(4), 837–853. https://doi.org/10.3758/s13415-015-0350-y

Dale, G., Sampers, D., Loo, S., & Green, C. S. (2018). Individual differences in exploration and persistence: Grit and beliefs about ability and reward. *PLoS One1, 13*(9), e0203131.

Daw, N. D., O'Doherty, J. P., Dayan, P., Seymour, B., & Dolan, R. J. (2006). Cortical substrates for exploratory decisions in humans. *Nature, 441*(7095), 876–879. https://doi.org/10.1038/nature04766

Feng, S. F., Want, S., Zarnescu, S., & Wilson, R. C. (2021). The dynamics of explore–exploit decisions reveal a signal-to-noise mechanism for random exploration. *Scientific Reports, 11*, 3077.

Ferguson, H. J., Brunsdon, V. E. A., & Bradford, E. E. F. (2021). The developmental trajectories of executive function from adolescence to old age. *Scientific Reports, 11*, 1382.d. https://doi.org/10.1038/s41598-020-80866-1

Gershman, S. J. (2018). Deconstructing the human algorithms for exploration. *Cognition, 173*, 34–42. https://doi.org/10.1016/j.cognition.2017.12.014

Giron, A.P., Ciranka, S.K., Schulz, E., van den Bos, W., Ruggeri, A., Meder, B., & Wu, C.M. (2022). Developmental changes in learning resemble stochastic optimization. *PsyArXiv*.

Gopnik, A. (2020). Childhood as a solution to explore–exploit tensions. *Philosophical Transactions of the Royal Society B: Biological Sciences, 375*(1803), 20190502. https://doi.org/10.1098/rstb.2019.0502

Gopnik, A., O'Grady, S., Lucas, C. G., Griffiths, T. L., Wente, A., Bridgers, S., Aboody, R., Fung, H., & Dahl, R. E. (2017). Changes in cognitive flexibility and hypothesis search across human life history from childhood to adolescence to adulthood. *Proceedings of the National Academy of Sciences, 114*(30), 7892–7899. https://doi.org/10.1073/pnas.1700811114

Gunnar, M. R., DePasquale, C. E., Reid, B. M., Donzella, B., & Miller, B. S. (2019). Pubertal stress recalibration reverses the effects of early life stress in postinstitutionalized children. *Proceedings of the National Academy of Sciences, 116*(48), 23984–23988. https://doi.org/10.1073/pnas.1909699116

Hartley, C. A., & Somerville, L. H. (2015). The neuroscience of adolescent decision making. *Current Opinion on Behavioral Sciences, 5*, 108–115. https://doi.org/10.1016/j.cobeha.2015.09.004

Hills, T. T., & Dukas, R. (2012). The evolution of cognitive search. *Cognitive search: Evolution, algorithms, and the brain*, 11-24.

Hills, T. T., Todd, P. M., Lazer, D., Redish, A. D., & Couzin, I. D. (2015). Exploration versus exploitation in space, mind, and society. *Trends in Cognitive Sciences, 19*(1), 46–54. https://doi.org/10.1016/j.tics.2014.10.004

Howard, S. J., Okely, A. D., & Ellis, Y. G. (2015). Evaluation of a differentiation model of preschoolers' executive functions. *Frontiers in Psychology, 6*, 285.

Jolliffe, I. T. (2002). Choosing a subset of principal components or variables. *Principal Component Analysis*, 111–149.

Lenow, J. K., Constantino, S. M., Daw, N. D., & Phelps, E. A. (2017). Chronic and acute stress promote overexploitation in serial decision making. *The Journal of neuroscience: the official journal of the Society for Neuroscience, 37*(23), 5681–5689. https://doi.org/10.1523/JNEUROSCI.3618-16.2017

Meder, B., Wu, C. M., Schulz, E., & Ruggeri, A. (2021). Development of directed and random exploration in children. *Developmental Science, 24*, e13095.

Mehlhorn, K., Newell, B. R., Todd, P. M., Lee, M. D., Morgan, K., Braithwaite, V. A., Hausmann, D., Fiedler, K., & Gonzalez, C. (2015). Unpacking the exploration–exploitation tradeoff: A synthesis of human and animal literatures. *Decision, 2*(3), 191–215. https://doi.org/10.1037/dec0000033

Mekern, V. N., Sjoerds, Z., & Hommel, B. (2019). How metacontrol biases and adaptivity impact performance in cognitive search tasks. *Cognition, 182*, 251–259. https://doi.org/10.1016/j.cognition.2018.10.001

Palacios-Barrios, E. E., Hanson, J. L., Barry, K. R., Albert, W. D., White, S. F., Skinner, A. T., … Lansford, J. E. (2021). Lower neural value signaling in the prefrontal cortex is related to childhood family income and depressive symptomatology during adolescence. *Developmental Cognitive Neuroscience, 48*, Article 100920.

Pelz, M., & Kidd, C. (2020). The elaboration of exploratory play. *Philosophical Transactions of the Royal Society B: Biological Sciences, 375*(1803), 20190503. https://doi.org/10.1098/rstb.2019.0503

Schulz, E., & Gershman, S. J. (2019). The algorithmic architecture of exploration in the human brain. *Current Opinion in Neurobiology, 55*, 7–14. https://doi.org/10.1016/j.conb.2018.11.003

Schulz, E., Wu, C. M., Ruggeri, A., & Meder, B. (2019). Searching for rewards like a child means less generalization and more directed exploration. *Psychological Science, 30*(11), 1561–1572. https://doi.org/10.1177/0956797619863663

Shulman, E. P., Smith, A. R., Silva, K., Icenogle, G., Duell, N., Chein, J., & Steinberg, L. (2016). The dual systems model: Review, reappraisal, and reaffirmation. *Developmental Cognitive Neuroscience, 17*, 103–117. https://doi.org/10.1016/j.dcn.2015.12.010

Somerville, L. H., Sasse, S. F., Garrad, M. C., Drysdale, A. T., Abi Akar, N., Insel, C., & Wilson, R. C. (2017). Charting the expansion of strategic exploratory behavior during adolescence. *Journal of Experimental Psychology: General, 146*(2), 155–164. https://doi.org/10.1037/xge0000250

Sumner, E., Li, A. X., Perfors, A., Hayes, B., Navarro, D., & Sarnecka, B. W. (2019). The Exploration Advantage: Children's instinct to explore allows them to find information that adults miss. *PsyArXiv*. https://doi.org/10.31234/osf.io/h437v.

Tymula, A., Rosenberg Belmaker, L. A., Roy, A. K., Ruderman, L., Manson, K., Glimcher, P. W., & Levy, I. (2012). Adolescents' risk-taking behavior is driven by tolerance to ambiguity. *Proceedings of the National Academy of Sciences of the United States of America, 109*(42), 17135–17140. https://doi.org/10.1073/pnas.1207144109

van den Bos, W., & Hertwig, R. (2017). Adolescents display distinctive tolerance to ambiguity and to uncertainty during risky decision making. *Scientific Reports, 7*, 40962. https://doi.org/10.1038/srep40962

von Helversen, B., Mata, R., Samanez-Larkin, G. R., & Wilke, A. (2018). Foraging, exploration, or search? On the (lack of) convergent validity between three behavioral paradigms. *Evolutionary Behavioral Sciences, 12*(3), 152–162. https://doi.org/10.1037/ebs0000121

Wilson, R. C., Bonawitz, E., Costa, V. D., & Ebitz, R. B. (2021). Balancing exploration and exploitation with information and randomization. *Current Opinion in Behavioral Sciences, 38*, 49–56. https://doi.org/10.1016/j.cobeha.2020.10.001

Wilson, R. C., Geana, A., White, J. M., Ludvig, E. A., & Cohen, J. D. (2014). Humans use directed and random exploration to solve the explore-exploit dilemma. *Journal of Experimental Psychology. General, 143*(6), 2074–2081. https://doi.org/10.1037/a0038199

Wolpert, D., & Macready, W. (1997). Macready, W.G.: No free lunch theorems for optimization. *IEEE Transactions on Evolutionary Computation, 1*(1), 67–82. https://doi.org/10.1109/4235.585893