Supplemental Materials

*to accompany*

**Characterizing Age-Related Change in Learning the Value of Cognitive Effort**

Phaneuf-Hadd, C. V., Jacques, I. M., Insel, C., Otto, A. R., & Somerville, L. H. (2025)

**Contents**

# I. Supplemental Methods

## A. Measurement Precision of Online Testing Platform

Given that this study was hosted online, it was important to control for as many sources of experimental setup variability as possible and collect data through platforms that maximized the measurement fidelity of key behavioral metrics of interest. One source of variability for capturing response times through remote, browser-based data collection are recording delays (Anwyl-Irvine et al., 2021; Bridges et al., 2020; Semmelmann & Weigelt, 2017). However, such timing offsets were abated by our within-subjects design: even if an operating system or browser was systematically slower on one machine or another, the lags were present across every reward-difficulty cue for a participant. Nonetheless, we heeded the recommendations of Anwyl-Irvine et al. (2021) and limited participants to a single browser for completing the study – Google Chrome. We chose Google Chrome because it is freely available and widely used, and it has performed well in previous stress tests of online testing platforms (Anwyl-Irvine et al., 2021; Bridges et al., 2020). In fact, in a rigorous examination of numerous operating system and browser combinations, Bridges et al. (2020) found that PsychoPy (Peirce et al., 2019) achieved response time precision under 3.5 ms across all candidate browsers, including Google Chrome. While Bridges et al. (2020) used a button box instead of computer keyboards, Anwyl-Irvine et al. (2021) tested desktop and laptop keyboards and found their sensitivity to be similar, within a few ms of each other on both macOS and Windows operating systems. In sum, our implementation choices were supported by prior research.

## B. Self-Report Questionnaires

Following the practice phase, learning task, and post-task ratings, participants completed a pair of self-report questionnaires measuring individual variability in motivational processes relevant to the application of reward-difficulty cues to behavior: the Need for Cognition (NFC) scale and the Behavioral Activation System Drive (BAS-Drive) subscale. Five minor participants reported having their parent/guardian help them read the items and response options on these questionnaires. One adolescent participant exited the study early and did not submit the self-report questionnaires (age = 17.17 years).

### *Need for Cognition*

To assess each participants' intrinsic motivation to engage in cognitively demanding tasks, an age-appropriate version of the NFC scale was administered. The adult version (Cacioppo et al., 2013) had 18 items (e.g., "I really enjoy a task that involves coming up with new solutions to problems.") that were endorsed on a five-point Likert scale ranging from one (extremely uncharacteristic of me) to five (extremely characteristic of me). The minor version (Keller et al., 2019) had 14 items (e.g., "I like to work on problems that require a lot of thinking."). To mirror the adult response options, the minor items were also endorsed on a five-point Likert scale ranging from one (not at all like me) to five (very much like me). The NFC scales were sum-scored according to their respective forward and reverse scoring rules, then were divided by their respective number of items, such that proportional adult and minor scores could be compared.

### *Behavioral Activation System Drive*

To quantify each participants' motivation to pursue reward, an age-appropriate version of the BAS-Drive subscale was administered. The adult version (Carver & White, 1994) had four items (e.g., "When I want something, I usually go all-out to get it.") that were endorsed on a four-point Likert scale ranging from one (Very true for me.) to four (Very false for me.). The minor version (Muris et al., 2005) also had four items (e.g., "I do everything to get the things that I want.") that were endorsed on a four-point Likert scale ranging from one (Not True) to four (Very True). The

BAS-Drive subscales were sum-scored according to their respective forward and reverse scoring rules.

**C. Random Slopes in Mixed-Effects Model**

The necessity of associating *time*, *difficulty*, and *incentive* with random slopes was evaluated by comparing conditional $R^2$ values between nested accuracy mixed-effects models containing and lacking random slopes for *time*, *difficulty*, and *incentive*. The r2 function from the *performance* package in R (version 0.9.1; Lüdecke et al., 2021) was used to achieve these comparisons. The conditional $R^2$ value for the *accuracy* model without random slopes was 0.837. Meanwhile, the conditional $R^2$ values for the *accuracy* model with a random slope for *time* was 0.844, with a random slope for *difficulty* was 0.836, and with a random slope for *incentive* was 0.840. Given the rather small degree of additional explained variance, if any, in the more complex models, *time*, *difficulty*, and *incentive* were not associated with random slopes in the reported *accuracy* mixed-effects model. *Time*, *difficulty*, and *incentive* were not considered as random slopes in the *reaction time* mixed-effects model, as to keep the random effect structures parallel across both measures of cognitive effort.

**D. Robustness of Age-Related Change**

To evaluate the importance of development for cognitive effort value learning, we considered how individual differences in baseline task-switching ability, intrinsic motivation to engage in cognitively demanding tasks, and motivation to pursue reward shaped response accuracy. We again implemented mixed-effects models with *time*, *difficulty*, and *incentive* as interacting fixed main effect predictors for *accuracy*, but diverting from the primary analyses reported in the main text, we entered *individual differences* as an additional interacting predictor and fashioned *age* as a control covariate. We continued to associate *participant ID* with a random intercept to account for repeated measurements. As alluded to before, *individual differences* were in baseline task-switching ability (in mean accuracy for the last block of each difficulty context in the practice phase; calculated continuously and treated as a metric variable), intrinsic motivation to engage in cognitively demanding tasks (in proportion NFC scores; treated as a metric variable), or motivation to pursue reward (in BAS-Drive sum-scores; treated as a metric variable). The *accuracy* models were fit as binomial generalized linear mixed-effects analyses with the glmer function. As always, the distributional assumptions of the models were appraised with posterior predictive checks (using the posterior_predictive_check function from the *performance* package in R; version 0.9.1; Lüdecke et al., 2021). While we found significant main effects and interactions involving some of the individual difference measures (**Supplemental Table S4**), importantly, *age* consistently persisted as a significant predictor of *accuracy*, even in its role as a control covariate.

### E. Transparency and Openness

There were minimal updates made to the pre-registration for the primary experiment. These updates included: 1) administering the matrix reasoning subtest in a manner that more closely mirrored an in-person assessment; 2) scoring the Need for Cognition scale with an analogous but more intuitive computation; 3) not mean-centering age in our mixed-effects models (as originally planned), as the transformation was ultimately unnecessary for interpreting the results; 4) not excluding participants for responding with the same answer to every item on a questionnaire (as originally planned), as there was no evidence for inattention on these scales; and 5) correcting an oversight in planned analyses for responses to “How hard was this galaxy?” (we aimed to test how perceptions of demand varied by reward *and* difficulty; only difficulty was originally written). The fourth and fifth minimal updates, listed above, were also made to the pre-registration for the secondary experiment.

## II. Supplemental Results

### A. Reward and Difficulty Guided Cognitive Effort Differently Across Time and Age

To further investigate how cognitive effort exertion is tuned across time and age according to accessible reward and learnable difficulty information, we next characterized the interacting influences of *age*, *time*, *difficulty*, and *incentive* on the *reaction time* of responses during the learning task. Mirroring the *accuracy* mixed-effects model reported in the main text, from the *reaction time* mixed-effects model, we found significant main effects of *age* ($\chi 2(1, N = 150) = 17.01$, $p < 0.001$) and *difficulty* ($\chi 2(1, N = 150) = 71.73$, $p < 0.001$), such that *reaction time* decreased with age and was faster for the easy relative to hard difficulty context. However, we also found a significant main effect of *time* ($\chi 2(2, N = 150) = 22.48$, $p < 0.001$), whereby participants responded more rapidly as the task progressed.

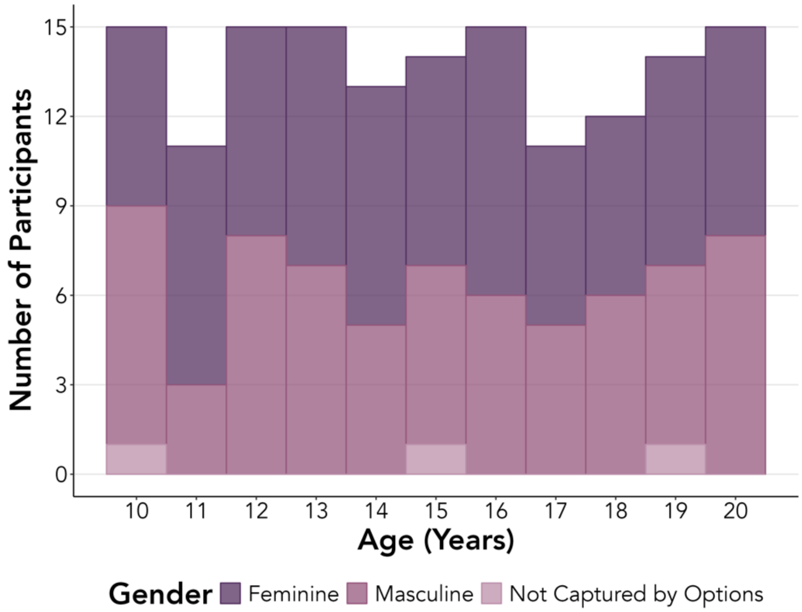
Building from these main effects, and like the *accuracy* mixed-effects model, we found a marginally significant two-way interaction between *age* and *difficulty* ($\chi 2(1, N = 150) = 2.88$, $p < 0.10$; **Supplemental Figure S7A**); *reaction time* accelerated for both the easy and hard difficulty contexts with increasing age, but slightly more steeply for the hard difficulty context. This result was qualified by a marginally significant three-way interaction between *age*, *difficulty,* and *incentive* ($\chi 2(1, N = 150) = 2.96$, $p < 0.10$; **Supplemental Figure S7B**). The trend – that *reaction time* declined at a faster rate with age in the hard difficulty context – was especially true for lower incentives. There were no other significant main effects or interactions than those listed here ($p$s > 0.104). Together, these results suggest that although advance difficulty information was seemingly more influential on age-related changes in cognitive effort exerted in service of speed than advance reward information, and as revealed by the *accuracy* mixed-effects model, both cue types shaped cognitive effort.

**B. Metacognitive Awareness of Cognitive Effort Across Age**

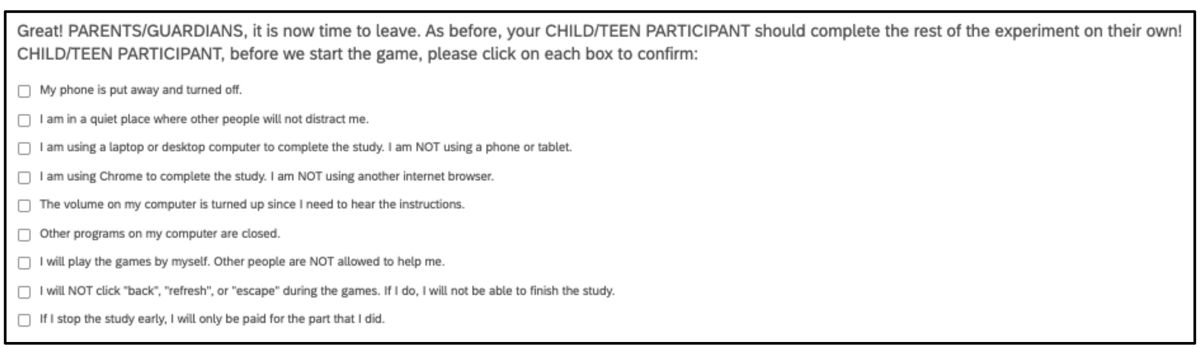
In addition to characterizing metacognitive awareness from the question, “How hard did you try in this galaxy?” in the main text, to evaluate whether participants’ subjective experience of difficulty varied across galaxies, we also examined responses to the question, “How hard was this galaxy?” using a separate model fit as a linear mixed-effects analysis with the lmer function. Here, the dependent variable was participants’ 7-point Likert *difficulty rating*s and the fixed main effect predictors were *age* and *galaxy*.

Perplexingly, in this *difficulty rating* mixed-effects model, we again found a significant main effect of *galaxy*, such that participants of all ages reported greater difficulty for the galaxies associated with higher than lower incentives in both difficulty contexts ($\chi 2(3, N = 149) = 75.38$, $p < 0.001$). Even though participants performed better for the easy condition (as revealed by the paired, one-tailed t-tests and *difficulty* main effects in the *accuracy* and *reaction time* mixed-effects models), suggesting that the hard galaxies were in fact harder, they endorsed the higher incentive galaxies as being more demanding in their *difficulty rating*s. There were no other significant main effects or interactions than those listed here ($p$s > 0.214).

## C. Supplemental Figures



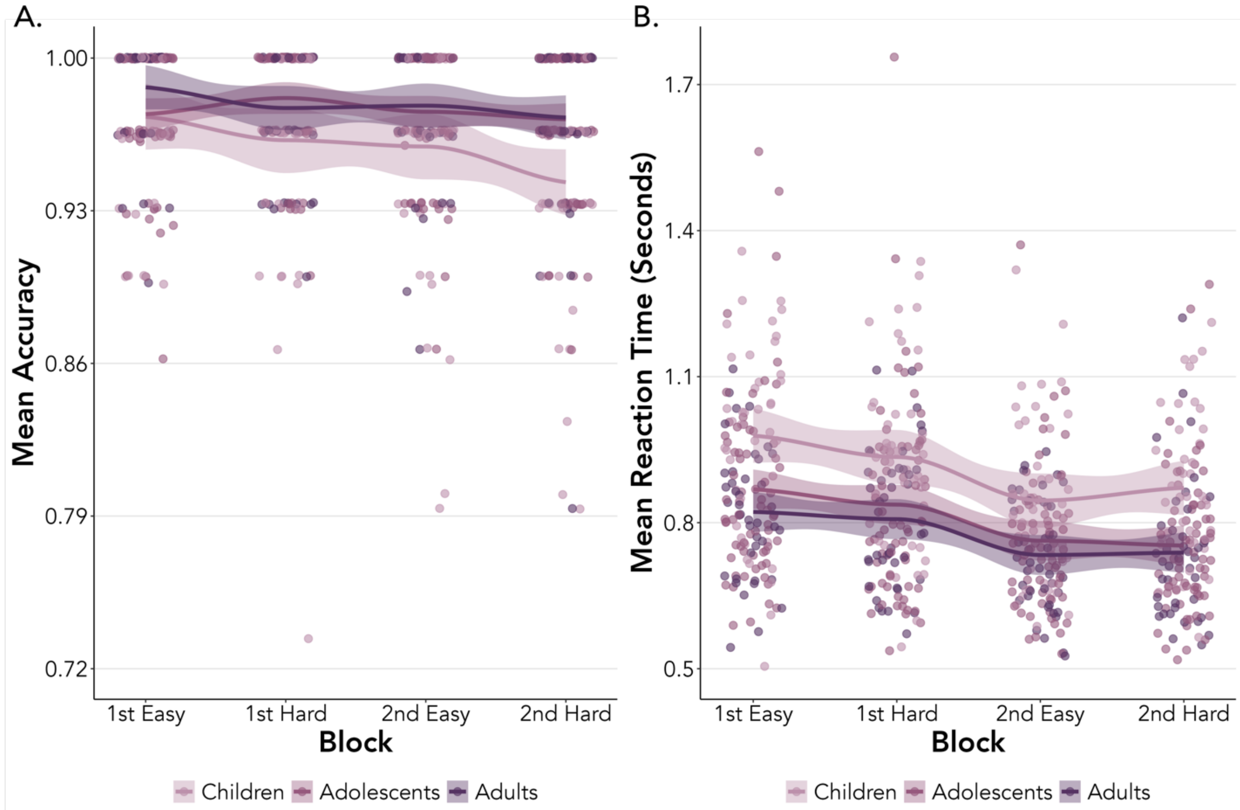

**Supplemental Figure S1.** Visualization of the participant sample (*N* = 150). Age is binned into years for data summarization purposes, but age was treated as a continuous variable in all analyses.



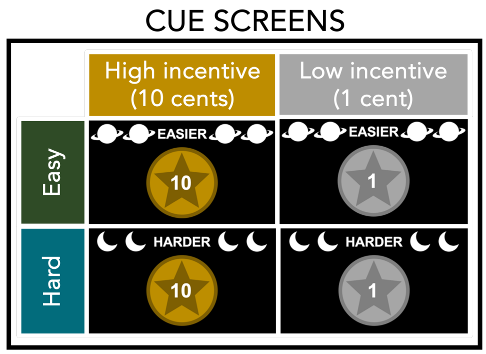
Great! PARENTS/GUARDIANS, it is now time to leave. As before, your CHILD/TEEN PARTICIPANT should complete the rest of the experiment on their own! CHILD/TEEN PARTICIPANT, before we start the game, please click on each box to confirm:

- ☐ My phone is put away and turned off.
- ☐ I am in a quiet place where other people will not distract me.
- ☐ I am using a laptop or desktop computer to complete the study. I am NOT using a phone or tablet.
- ☐ I am using Chrome to complete the study. I am NOT using another internet browser.
- ☐ The volume on my computer is turned up since I need to hear the instructions.
- ☐ Other programs on my computer are closed.
- ☐ I will play the games by myself. Other people are NOT allowed to help me.
- ☐ I will NOT click "back", "refresh", or "escape" during the games. If I do, I will not be able to finish the study.
- ☐ If I stop the study early, I will only be paid for the part that I did.

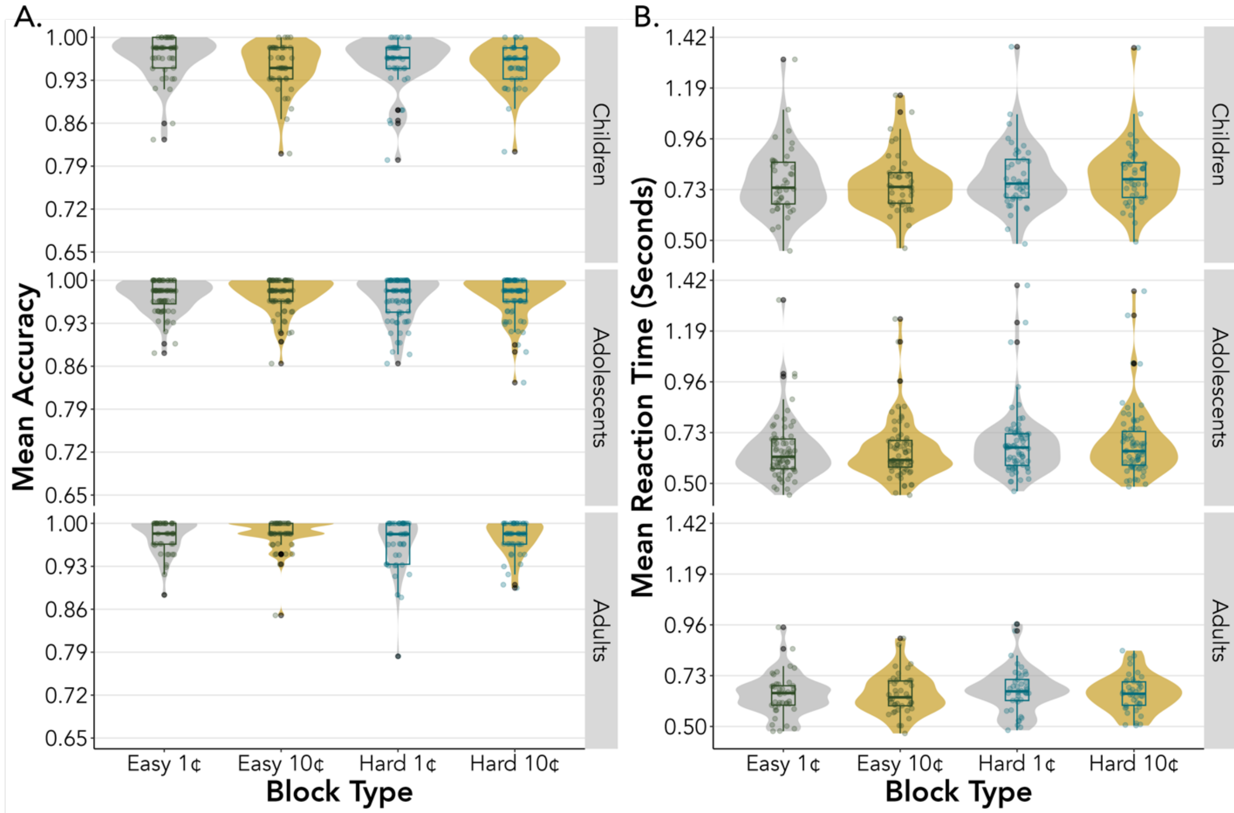
**Supplemental Figure S2.** After completing the consent process, minor participants were required to endorse these items in order to proceed to the study. Participants could not continue unless all buttons were clicked. Adult participants were shown analogous items.



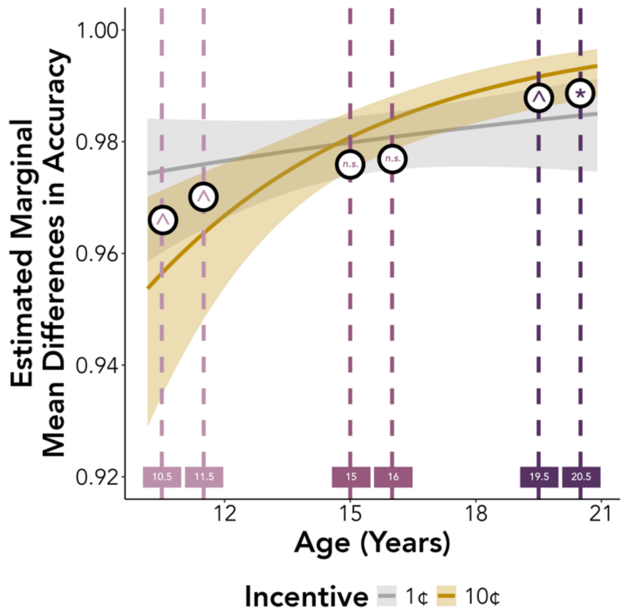

**Supplemental Figure S3.** The practice phase effectively exposed participants to the demands of the task-switching paradigm. **(A)** Across age, accuracy was consistently high, with many participants performing at or near 100% correct. **(B)** Additionally, reaction time plateaued by the end of the practice trials for individuals of all ages. Age is binned into children (10–12 years), adolescents (13–17 years), and adults (18–20 years) for data visualization purposes, but age was treated as a continuous variable in all analyses. Shading represents 95% confidence intervals around fitted lines.



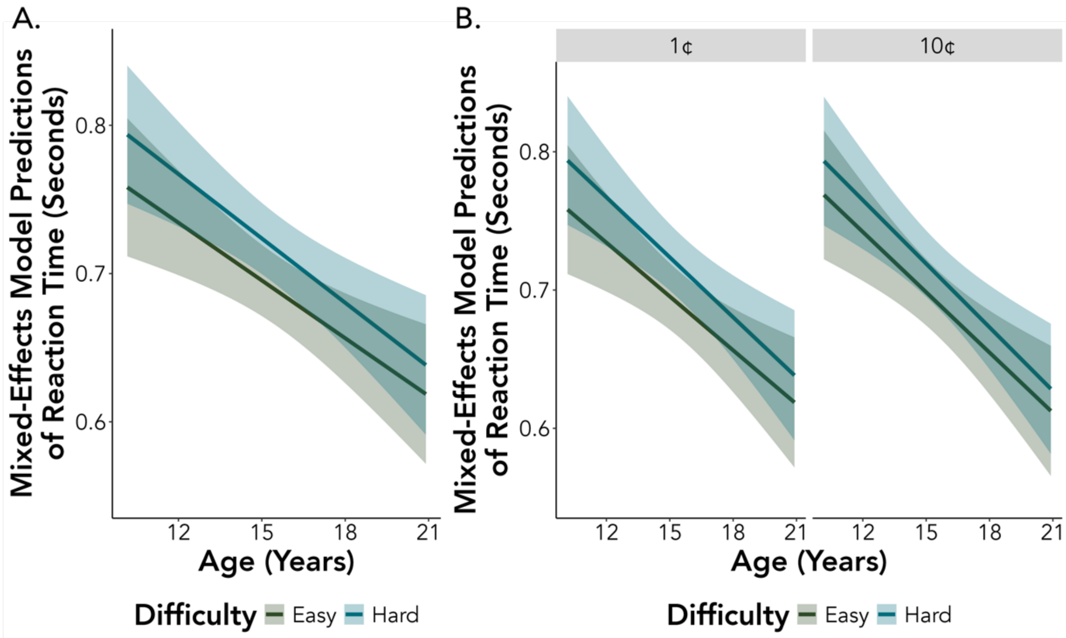

**Supplemental Figure S4.** In the secondary experiment, and unlike in the primary, the mini reward blocks were cued with screens that explicitly displayed full information about the upcoming trials. The incentive cue (1 or 10 cents) was signaled with gold and silver coins, respectively. The difficulty cue (planet or moon galaxy) was signaled with space-themed icons and the words ‘easier’ and ‘harder’, respectively. While the cue screens in the secondary study provided two sources of instructed information, they were designed to be as visually similar to the cue screens in the primary study as possible (Figure 1, Cue Screens); the only difference between the primary and secondary experiments was the explicitness of the difficulty information.




**Supplemental Figure S5.** For all age groups, performance variance in accuracy **(A)** and reaction time **(B)** was distributed fairly homogeneously across each mini reward block. Namely, consistency in performance variance across conditions was observed through the similar spreads of the violin plots relative to each other within an age group. Age is binned into children (10–12 years), adolescents (13–17 years), and adults (18–20 years) for data visualization purposes, but age was treated as a continuous variable in all analyses. Black points correspond to outliers.




**Supplemental Figure S6.** Accuracy was moderated by the incentives offered for correct responses across age. Post-hoc tests of estimated marginal mean differences between reward amounts clarified the divergence and convergence of the higher and lower incentive lines at systematic points along the age range. These tests revealed that younger participants (captured by two values near the youngest end of our age range) trended towards elevated accuracy for lower than higher incentives, adolescents (captured by two values near the middle of our age range) did not demonstrate incentive-based differences in accuracy, and older participants (captured by two values near the oldest end of our age range) displayed elevated accuracy for higher than lower incentives. Shading represents 95% confidence intervals around fitted lines. *P*-values (* $p < 0.05$, ^ $p < 0.10$) were corrected for multiple comparisons according to the Holm method.




**Supplemental Figure S7.** Reaction time was marginally moderated, across age, by the difficulty of the task-switching context and the incentives offered for correct task-switching performance. **(A)** Conditioned on the other main effects and interactions in the mixed-effects model, reaction time decreased more rapidly for the hard than easy difficulty context with age. As a result of the differential steepness of these slopes, younger participants responded somewhat slower in the more demanding blocks while older participants were similarly fast across the more and less demanding blocks. **(B)** The pattern of results depicted in (A) was specific to the lower incentive amount. Conditioned on the other main effects and interactions in the mixed-effects model, reaction time decreased at comparable rates for the hard and easy difficulty contexts with age for the higher incentive amount. Shading represents 95% confidence intervals around fitted lines.

## D. Supplemental Tables

**Supplemental Table S1.** The participant samples for the primary and secondary experiments matched across key demographic variables. No participant completed both studies.

| **Demographic Variables** | **Primary study** *(i.e., instructed reward but learnable difficulty information)* | **Secondary study** *(i.e., instructed reward and instructed difficulty information)* |
|---|---|---|
| $N_{usable}$ | 150 | 150 |
| $M_{age}$, $SD_{age}$ | 15.49 years, 3.17 years | 15.34 years, 3.20 years |
| $N_{feminine}$ | 79 | 79 |
| Asian | 11.33% | 14.00% |
| Black/African American | 6.67% | 3.33% |
| Native Hawaiian or Other Pacific Islander | 0.00% | 0.67% |
| White/Caucasian | 60.67% | 64.67% |
| Race was not captured by previous options | 21.33% | 17.33% |
| Hispanic | 12.00% | 14.00% |

**Supplemental Table S2.** Cells include statistically significant ($p < 0.05$) and marginally significant ($p < 0.10$) main effects and interactions from the type II analysis of deviance performed on the *accuracy* mixed-effects model. This model corresponds to the primary analysis, containing the key dependent variable, and reveals that *reward* and *difficulty* guided cognitive effort differently across *time* and *age*.

| **Main effects and interactions** | **Results from *accuracy* mixed-effects model** |
|---|---|
| *age* | $\chi 2(1, N = 150) = 10.42, p < 0.01$ |
| *difficulty* | $\chi 2(1, N = 150) = 9.14, p < 0.01$ |
| *age x incentive* | $\chi 2(1, N = 150) = 8.41, p < 0.01$ |
| *age x difficulty* | $\chi 2(1, N = 150) = 3.01, p < 0.10$ |
| *age x incentive x time* | $\chi 2(2, N = 150) = 9.18, p < 0.05$ |

**Supplemental Table S3.** Sensitivity analysis confirmed that speed-accuracy trade-offs could not account for differences in the effects (as detected by type II analyses of deviance) produced by the *accuracy* and *reaction time* mixed-effects models. The sole difference between the *reaction time* model run on the complete set of trials and the *reaction time* model run on correct trials only was that the marginally significant interaction between *age* and *difficulty* became non-significant. Otherwise, all other results held, including – crucially – the statistically significant main effects.

| **Main effects and interactions** | **Results from *reaction time* mixed-effects model with complete set of trials**[1] | **Results from *reaction time* mixed-effects model excluding incorrect trials** |
|---|---|---|
| *age* | $\chi2(1, N = 150) = 17.01$, $p < 0.001$ | $\chi2(1, N = 150) = 17.59$, $p < 0.001$ |
| *difficulty* | $\chi2(1, N = 150) = 71.73$, $p < 0.001$ | $\chi2(1, N = 150) = 74.53$, $p < 0.001$ |
| *time* | $\chi2(2, N = 150) = 22.48$, $p < 0.001$ | $\chi2(2, N = 150) = 21.35$, $p < 0.001$ |
| *age x difficulty* | $\chi2(1, N = 150) = 2.88$, $p < 0.10$ | $\chi2(1, N = 150) = 2.40$, $p = 0.121$ |
| *age x difficulty x incentive* | $\chi2(1, N = 150) = 2.96$, $p < 0.10$ | $\chi2(1, N = 150) = 3.25$, $p < 0.10$ |

[1] Reported in the Supplemental Results

**Supplemental Table S4.** Analyses of individual differences (*ind_diff*) in baseline task-switching ability (*ind_diff: baseline*), intrinsic motivation to engage in cognitively demanding tasks (*ind_diff: NFC*), and motivation to pursue reward (*ind_diff: BAS-Drive*) verified age as crucially important for understanding variability in cognitive effort value learning. Regardless of the individual difference measure employed, age explained meaningful variance in all mixed-effects models (top row of table). Cells include statistically significant ($p < 0.05$) and marginally significant ($p < 0.10$) main effects and interactions from the type II analyses of deviance performed on the models.

| **Main effects and interactions** | **Results from *accuracy* mixed-effects model**[2] | **Results from *baseline* mixed-effects model** | **Results from *NFC* mixed-effects model** | **Results from *BAS-Drive* mixed-effects model** |
|---|---|---|---|---|
| *age* | $\chi 2(1, N = 150) = 10.42, p < 0.01$ | $\chi 2(1, N = 150) = 8.31, p < 0.01$ | $\chi 2(1, N = 149) = 9.37, p < 0.01$ | $\chi 2(1, N = 149) = 7.28, p < 0.01$ |
| *ind_diff* | n/a | *baseline* $\chi 2(1, N = 150) = 19.09, p < 0.001$ | *NFC* $\chi 2(1, N = 149) = 3.59, p < 0.10$ | n.s. |
| *difficulty* | $\chi 2(1, N = 150) = 9.14, p < 0.01$ | $\chi 2(1, N = 150) = 5.45, p < 0.05$ | $\chi 2(1, N = 149) = 8.95, p < 0.01$ | $\chi 2(1, N = 149) = 8.89, p < 0.01$ |
| *ind_diff x incentive* | *age* $\chi 2(1, N = 150) = 8.41, p < 0.01$ | n.s. | n.s. | n.s. |
| *ind_diff x difficulty* | *age* $\chi 2(1, N = 150) = 3.01, p < 0.10$ | n.s. | n.s. | n.s. |
| *ind_diff x incentive x time* | *age* $\chi 2(2, N = 150) = 9.18, p < 0.05$ | n.s. | n.s. | *BAS-Drive* $\chi 2(2, N = 149) = 4.66, p < 0.10$ |
| *ind_diff x difficulty x time* | n.s. | n.s. | n.s. | *BAS-Drive* $\chi 2(2, N = 149) = 5.99, p < 0.10$ |

[2] Reported in the main text

| | | | | |
|---|---|---|---|---|
| *ind_diff x incentive x difficulty* | n.s. | *baseline* $\chi 2(1, N = 150) = 5.89, p < 0.05$ | n.s. | n.s. |

**Supplemental Table S5.** Cells include statistically significant ($p < 0.05$) main effects from the type II analysis of deviance performed on the *accuracy* mixed-effects model from the secondary study. For comparison purposes between the two studies, cells also include non-significant interactions from the secondary study, which were found to be significant in the primary study. The secondary study's mixed-effects model reveals that *reward* and *difficulty* did not economically guide cognitive effort into adulthood in an environment without learning demands.

| **Main effects and interactions** | **Results from *accuracy* mixed-effects model** |
|---|---|
| *age* | $\chi 2(1, N = 150) = 9.08, p < 0.01$ |
| *difficulty* | $\chi 2(1, N = 150) = 5.55, p < 0.05$ |
| *age x incentive* | $\chi 2(1, N = 150) = 2.70, p = 0.101$ |
| *age x difficulty* | $\chi 2(1, N = 150) = 0.28, p = 0.597$ |
| *age x incentive x time* | $\chi 2(2, N = 150) = 3.25, p = 0.197$ |

## III. References

Anwyl-Irvine, A., Dalmaijer, E. S., Hodges, N., & Evershed, J. K. (2021). Realistic precision and accuracy of online experiment platforms, web browsers, and devices. *Behavior Research Methods, 53*, 1407-1425.

Bridges, D., Pitiot, A., MacAskill, M. R., & Peirce, J. W. (2020). The timing mega-study: Comparing a range of experiment generators, both lab-based and online. *PeerJ, 8*, e9414.

Semmelmann, K., & Weigelt, S. (2017). Online psychophysics: Reaction time effects in cognitive experiments. *Behavior Research Methods, 49*, 1241-1260.