Supplementary Materials for

Infants make more attempts to achieve a goal when they see adults persist

Julia A. Leonard,* Yuna Lee, Laura E. Schulz

*Corresponding author. Email: jlnrd@mit.edu

This PDF file includes:

Materials and Methods

Supplementary Text

Table S1

Caption for movie S1

Other supplementary material for this manuscript includes the following:

Movie S1

Supplementary Materials:

Materials and Methods

Experiment 1

Participants

Infants were recruited at an urban children's museum and tested individually in a quiet testing room off the museum floor. A power analysis assuming a large effect size ($d = 0.8$, power $= 0.9$) indicated that 34 infants per condition would allow a high probability of finding any differences between conditions in planned t-test comparisons. A total of 24 infants were excluded from the experiment (11 in Effort condition, 7 in No Effort condition and 6 in Baseline). Infants were excluded for the following reasons: never pressing the button on the toy ($n = 8$; 4 in Effort Condition, 4 in No Effort condition), experimental error due to stimuli breaking or not getting child's date of birth ($n = 3$; 1 in each condition type), and parental interference ($n = 13$; 6 in Effort condition, 2 in No Effort condition and 5 in Baseline). Parental interference consisted of: 1) demonstrating how the toy worked by pressing the button themselves ($n = 11$; 5 in Effort condition, 2 in No Effort condition, and 4 in Baseline); 2) handing the child a toy other than the test toy ($n = 1$ in Effort condition) and 3) not giving back the toy to the child once they tossed it off their table ($n=1$ in Baseline). The remaining 102 infants (mean: 15.36 months; range: 13-18 months; 50 Female, 52 Male) were randomly assigned to the Effort, No Effort, and Baseline conditions ($n = 34$/ condition; ages were matched between conditions, $\beta = -0.15$, 95% CI [-0.51, 0.20]). All research was approved by the MIT Institutional Review Board and conducted with the informed consent of parents.

Materials

Two toys were used by the adult model. One toy was a tomato container with a rubber frog inside. The tomato container looked as though it could be opened by removing a plastic lid on the bottom of the container, but actually opened by peeling off a sticker at the top of the container. The other toy was a carbineer with a cow key chain attached. The key chain lit up and made mooing sounds when a button was pressed and could be removed by twisting and then squeezing the carabineer. The toy used for the infant test task was a square music box (6.35^3 cm) covered in felt with a large red button with a musical note (3.81 cm) on the top. The button was easy to press but inert; a button concealed under the felt at the bottom of the toy actually activated the music. The bottom of the toy needed to be pressed firmly on a hard surface to trigger the button. The trigger was intended to be too difficult for the infants to activate (although 7% of infants succeeded, excluding these children did not change our results; see "additional results" in supplemental text). Additionally, three warm-up toys (a rattle, a stuffed elephant, and a toy that lit up and vibrated) were used to familiarize the infant to the high chair and testing room.

Procedure

During the experiment, infants sat in a high chair or booster seat next to their parent. When the experimenter introduced the child to the music toy, she placed the toy out of the infant's sight (underneath the tray on the infants' seat) and activated the toy using the hidden button. The toy played a musical tune for approximately 5 seconds. The experimenter then handed the toy to the infant and left the room. Parents had been instructed during the consent

process to refrain from interacting with their infants except to return the toy to them up to three times if the children dropped or handed off the toy. At the end of the experiment, the experimenter helped the infant successfully activate the music toy.

Coding and analyses

   Button presses were operationalized as a hand pushing down the button. Button presses were coded from videotape by two coders blind to hypotheses and condition (inter-rater reliability $r = .99, p < .001$). Data from a single coder was used for analyses but all results held with the second coder's data. Coders agreed with the experimenter's judgment on the termination of the experiment 100% of the time. Additionally, a coder blind to condition and hypotheses coded the tapes for potential confounds. No difference was found across the conditions for whether the parent talked to the child ($X^2(2, N=102) = 1.03, p = .60$), parents' proximity to the infant (as distance in inches, $H(2) = 4.32, p = .12$), and parents' encouragement to the infant ($X^2(2, N = 102) = 5.56, p = .06$). Additionally, because the experimenter might have conveyed more enthusiasm in handing the toy to the child in the Effort than the No Effort condition, a second coder rated the experimenter's enthusiasm at the start of the test trial on a Likert scale, blind to conditions ($W = 525, p = .50$). For linear models, the dependent variables were transformed to the 0.5 power so that the distribution would adhere better to a normal distribution. The 95% confidence intervals reported throughout were obtained from a bootstrap with 10,000 samples.

Replication

Participants

   Infants were recruited in the same fashion as in experiment 1. A power analysis using the effect sizes for the dependent measures in the Effort and No Effort conditions in experiment 1 ($d = .63$, power $= .8$) indicated that we needed 40 infants per condition in the replication. A total of 30 infants were excluded from the experiment (14 in Effort condition and 16 in No Effort condition). Infants were excluded for the following reasons: never pressing the button on the toy (n = 7; 6 in Effort Condition, 1 in No Effort condition), experimental error due to stimuli breaking, incorrectly demonstrating condition, or ending experiment early (n = 4; 1 in Effort Condition, 3 in No Effort condition), and parental interference (n = 16; 4 in Effort condition, 12 in No Effort condition). Again, parental interference consisted of: 1) demonstrating how the toy worked by pressing the button themselves (n = 13; 4 in Effort condition, 9 in No Effort condition); 2) handing the child a toy other than the test toy (n = 1 in No Effort condition) and 3) not giving back the toy to the child once they tossed it off their table (n = 2 in No Effort condition). Additionally, we excluded infants for being fussy (operationalized as parents ending the experiment early due to perceived child fussiness) in the replication (n = 3, all in Effort condition). Although we did not exclude (two) fussy infants in experiment 1, we had pre-registered the exclusion criteria to exclude fussy infants in the replication (and all results of experiment 1 hold when excluding the two fussy infants).

   The remaining 80 infants (mean: 15.21 months; range: 13-18 months; 44 Female, 36 Male) were randomly assigned to the Effort, and No Effort conditions (n = 40/ condition; ages were matched between conditions, $\beta = -0.43$, 95% CI [-1.14, 0.30]).

Materials

All Materials were the same as in the experiment 1 except that some additional felt was added to cover the concealed switch at the bottom of the toy to make it harder for infants to activate.

Procedure
The procedure was the same as in experiment 1.

Coding and analyses
All data were coded from videotape by the first author and a coder blind to hypotheses and condition; data from the coder blind to hypotheses and condition were used throughout but all results remained the same with the first author's coding (button pressing inter-rater reliability $r = .99, p < .001$). A third coder blind to hypotheses and conditions rescored 30% of the button pressing data (inter-rater reliability with the other blind coder was $r = .99, p < .001$ for both coders). Coders agreed with the experimenter's judgment on the termination of the experiment 100% of the time. The coder blind to hypotheses and conditions also coded the tapes for potential confounds. No difference was found across the conditions for whether the parent talked to the child ($X^2(1, N = 80) = 0.0, p = 1.0$), parents' proximity to the infant (as distance in inches, $W = 768, p = .76$), parents' encouragement to the infant ($X^2(1, N = 80) = 0.33, p = .57$), and tone of voice of the experimenter when they handed the toy to the infant (on a Likert scale from 1, not encouraging, to 7, very encouraging: $W = 788, p = .87$). As in experiment 1, the dependent variables were transformed to the 0.5 power to better adhere to a normal distribution for linear models.

Experiment 2

Participants
Infants were recruited in the same fashion as in experiment 1 and the replication. A total of 17 infants were excluded from the experiment (9 in Effort condition and 8 in No Effort condition). Infants were excluded for the following reasons: never pressing the button on the toy ($n = 4$; 2 in Effort Condition, 2 in No Effort condition), experimental error due to stimuli breaking and not getting child's date of birth ($n = 3$; 2 in Effort Condition, 1 in No Effort condition), and parental interference ($n = 6$; 2 in Effort condition, 4 in No Effort condition). In this experiment, the only form of parental interference was demonstrating how the toy worked by pressing the button themselves ($n = 6$; 2 in Effort condition, 4 in No Effort condition). As in the replication, we also excluded infants for being fussy (operationalized as parents ending the experiment early due to perceived child fussiness; $n = 4$; 3 in Effort condition, 1 in No Effort condition).

The remaining 80 infants (mean: 15.55 months; range: 13-18 months; 41 Female, 39 Male) were randomly assigned to the Effort, and No Effort conditions ($n = 40$ in Effort, 40 in No Effort; ages were matched between conditions, $\beta = -0.35$, 95% CI [-1.11, 0.41]).

Materials
All Materials were the same as in the replication.

Procedure

The procedure was the same as in the replication except the experimenter did not make eye contract, say the infants' name or use infant-directed speech during the Effort and No Effort demonstrations.

Coding and analyses

All data were coded from videotape by the first author and a coder blind to hypotheses and condition (button press inter-rater reliability $r = .99, p < .001$); data from the coder blind to hypotheses and condition were used throughout but all results remained the same with the first author's coding. Coders agreed with the experimenter's judgment on the termination of the experiment 100% of the time. The coder blind to hypotheses and conditions also coded the tapes for potential confounds. No difference was found across the conditions for whether the parent talked to the child ($X^2(1, N = 80) = 0.21, p = .65$), parents' proximity to the infant (as distance in inches, $W = 886.5, p = .39$), parents' encouragement to the infant ($X^2(1, N = 80) = 1.47, p = .23$), and tone of voice of the experimenter when they handed the toy to the infant (on a Likert scale from 1, not encouraging, to 7, very encouraging: $W = 713, p = .28$). As in both previous experiments, the dependent variables were transformed to the 0.5 power to better adhere to a normal distribution for linear models.

To make sure infants' attention to the experimenter demonstration did not differ by pedagogical context, a blind coder coded the number of seconds infants were looking at the experimenter during the two demonstrations in the first 32 participants with data where infants eyes were visible from the replication and experiment 2. Infant attention did not differ across the replication and experiment 2 (Mean (SD): replication = 64.6 (6.3) seconds, experiment 2 =63.9 (4.4); $t(62) = -0.56, p=.58$).

Supplemental text

Experiment 1

Additional Results

Along with ANOVAs, non-parametric Kruskal-Wallis tests revealed that both the total number of times infants pressed the button and the number of times they pressed the button before first handoff differed by condition (total button presses: $H(2) = 8.13, p = .02$; presses before first handoff: $H(2) = 8.02, p = .02$).

There were no differences in overall playtime between conditions (mean play time in seconds, Effort: 88.26, 95% CI [78.06, 98.41], No Effort: 85.00, 95%CI [73.97, 96.21]; Baseline: 71.71, 95%CI [60.88, 82.32]; $F(2,99) = 2.47, p = .09, \eta^2 = .05$) or tendency to hand-off or discard the toy between conditions (mean number of hand-offs, Effort: 2.09, 95% CI [1.71, 2.53], No Effort: 2.09, 95%CI [1.68, 2.53]; Baseline: 2.56, 95%CI [2.27, 2.91]; $X^2(6, N=102) = 4.55, p = .60$).

Further analyses revealed a correlation between age in months and total number of button presses ($r_s(100) = 0.24, p = .02$), but not button presses before handoff ($r_s (100)= 0.12, p = .22$). No age by condition interactions were found for either total button presses or button presses before first handoff (both $p > .6$).

We opted to limit the exclusion criteria to those specified above, however, the results were robust to the inclusion criteria; all results held if we additionally excluded children 1) who successfully activated the toy (n = 7, 4 in Effort Condition, 3 in No Effort condition) 2) whose parents verbally encouraged them to press the button (n = 19, 9 in Effort Condition, 8 in No Effort condition and 2 in Baseline), 3) whose parents physically encouraged their child to press

the button by pointing to the toy or pushing the toy toward the child (n = 9, 4 in Effort Condition, 3 in No Effort condition, and 2 in Baseline) and 4) whose parents asked to terminate the experiment early because their child was fussy (n = 2, 1 in Effort Condition, 1 in No Effort condition).

Further, results were robust to outliers (defined as more than 1.5 interquartile range above the third quartile). When the 1 Effort and 2 Baseline outliers for total button presses were excluded, the ANOVA and Kruskal-Wallis test ($F(2, 96) = 6.33$, $p = .003$, $\eta^2 = .12$; $H(2) = 9.26$, $p = 0.01$) and post-hoc comparisons that were significant remained significant (Effort vs Baseline: $\beta = 1.37$, $t(63) = 3.54$, $p = .0008$, 95% CI [0.63, 2.13]; $W = 757.5$, $p = .003$, $r = -.37$; Effort vs No Effort: $\beta = 1.11$, $t(65) = 2.50$, $p = .02$, 95% CI [0.24, 2.00]; $W = 729$, $p = .04$, $r = -.26$). This was also true when excluding the 2 Effort, 2 No Effort, and 1 Baseline outliers for button presses before first hand off ($F(2, 94) = 5.04$, $p = .008$, $\eta^2 = .10$; $H(2) = 8.22$, $p = 0.02$; Effort vs Baseline: $\beta = 1.00$, $t(63) = 2.70$, $p = .009$, 95% CI [0.27, 1.74]; $W = 336.5$, $p = .01$, $r = -.31$; Effort vs No Effort: $\beta = 1.05$, $t(62) = 2.55$, $p = .01$, 95% CI [0.24, 1.85]; $W = 330.5$, $p = .01$, $r = -.30$).

Replication

Additional Results

To further understand the patterns in our results, we ran simulations to calculate the power of our results using our own data. We ran 10,000 linear models and Wilcoxon rank-sum tests with 40 bootstrapped samples from the effort and no effort group respectively. We then coded each significant model as 1 and each insignificant model as 0. Dividing the number of significant models by the total number of simulations gives us an estimate of our studies' power. For the total number of presses, we got a power of .55 for the linear models and a power of .66 for the Wilcox rank-sum test. For the presses before first handoff, we got a power of .48 for the linear models and .59 for the Wilcox rank-sum test. This confirms that we had sufficient power to detect an effect and suggests that the Wilcox rank-sum test is a more powerful model for this data over the linear models (presumably because of the non-normality of our data).

As in experiment 1, there were no differences in overall playtime between conditions (mean play time in seconds, Effort: 88.25, 95% CI [78.42, 98.47], No Effort: 76.15, 95%CI [66.03, 85.90]; $t(78) = -1.66$, $p = .10$, $d = -.37$) or in tendency to discard or hand off the toy between conditions (mean number of hand-offs, Effort: 2.13, 95% CI [1.75, 2.55], No Effort: 2.48, 95%CI [2.18, 2.83], $X^2(3, N=80) = 3.96$, $p = .27$).

In contrast to the experiment 1, Spearman correlations reveal no relationship between age in months and total number of button presses ($r_s(78) = 0.08$, $p = .47$). This relationship was also not found between age and button presses before handoff ($r_s(78) = 0.06$, $p = .59$). No age by condition interactions were found for either total button presses or button presses before first handoff (both $p > .3$).

Additionally, all results held when excluding children with possible confounds of 1) parents verbally encouraging their children to press the button (n = 15, 9 in Effort Condition, 6 in No Effort condition) and 2) parents physically encouraging their child to press the button by pointing to the toy or pushing the toy toward the child (n = 19, 8 in Effort Condition, 11 in No Effort condition). Note that no children successfully activated the toy in the replication because after experiment 1, we added more felt to the bottom of the toy to make it more difficult for the infants to activate.

Although we did not pre-register removing outliers from our analyses, we explored the robustness of our results when taking them out at the suggestion of a reviewer. When removing the 3 Effort and 2 No Effort total button presses outliers, the significant difference in total button presses by condition remained significant ($\beta = 0.80$, $t(73) = 2.33$, $p = .02$, 95% CI [0.13, 1.49]; $W = 478$, $p = .02$, $r = -.27$). Similarly, when removing the 1 Effort and 2 No Effort button presses before first handoff outliers, the button presses before first handoff in the Effort condition was significantly more than the No Effort condition ($\beta = 0.95$, $t(75) = 2.56$, $p =. 01$, 95% CI [0.22, 1.71]; $W = 494.5$, $p = .01$, $r = -.29$).

Experiment 2

Additional Results

The tendency to hand off the toy did not differ between conditions (mean number of hand-offs, Effort: 1.93, 95% CI [1.55, 2.30], No Effort: 2.20, 95%CI [1.83, 2.60], $X^2(3, N=80) = 5.17$, $p = .16$). Unlike experiment 1 and the replication, in experiment 2 there was a difference in overall playtime between conditions, with children in the Effort condition playing for longer than children in the No Effort condition (mean play time in seconds, Effort: 100.65, 95% CI [93.20, 108.40], No Effort: 81.83, 95%CI [71.48, 92.45]; $t(78)= -2.80$, $p=.007$, $d = -.63$).

Just as in experiment 1, Spearman correlations revealed a relationship between age in months and total number of button presses ($r_s(78) = .24$, $p = .03$). This relationship was not found between age and button presses before handoff ($r_s(78) = .15$, $p = .19$). No age by condition interactions were found for either total button presses or button presses before first handoff (both $p > .2$).

When excluding children with possible confound of parents verbally encouraging their children to press the button (n = 13, 9 in Effort Condition, 4 in No Effort condition), there was no effect of condition on total number of button presses ($W = 686.5$, $p = .11$, $r = -.20$). Further, when excluding children with the possible confound of parents physically encouraging their child to press the button by pointing to the toy or pushing the toy toward the child (n = 25, 12 in Effort Condition, 13 in No Effort condition), the total number of button press difference by condition was only a trend ($W = 485$, $p = .07$, $r = -.24$).

We looked at whether the results were robust to removing outliers (although as noted, we planned to analyze all the data that met inclusion criteria and this was not a pre-registered analysis). There were 2 outliers in the Effort condition and 1 in the No Effort condition (with outliers defined as infants whose button pressing was in the 1.5 quartile interval above the third quartile). Without these outliers, the results became a non-significant trend ($\beta = .72$, $t(75) = 1.86$, $p = .07$, 95% CI [-0.04, 1.47]; $W = 561$, $p = .07$, $r = -.21$) and the effect of outliers on Experiment 2 can also be seen in the substantial overlap in the 95% confidence intervals for the median (a degree of overlap not present in experiment 1 or the replication; see Table 2.)

Table S1. Wilcoxon rank-sum tests for the two main outcome measures of interest by each condition in experiment 1, the replication, and experiment 2.

	Experiment 1	**W**	**p**	**r**
Effort vs. Baseline	Total button presses	793	.008	-.32
	Presses before first handoff	367.5	.01	-.31
Effort vs. No Effort	Total button presses	763	.02	-.27
	Presses before first handoff	390.5	.02	-.28
No Effort vs. Baseline	Total button presses	593	.86	-.02
	Presses before first handoff	582.5	.96	-.01
	Replication			
Effort vs. No Effort	Total button presses	556.5	.02	-.26
	Presses before first handoff	573.5	.03	-.24
	Experiment 2			
Effort vs. No Effort	Total button presses	599	.05	-.22
	Presses before first handoff	684	.27	-.12

Movie S1
Effort and No Effort conditions demonstrated on two infants each from experiment 1 and the replication.