Electronic Supplementary Materials

Predictions about reward outcomes in rhesus monkeys

Yiyun Huang, Hayoung Chang, Laurie R. Santos, & Alexandra G. Rosati

Supplemental Methods

As reported in the main text, the primary experimenter presented the stimuli to monkeys by manipulating two boxes on a stage. Supplementary Figure S1 shows photos of these demonstrations from the monkey's perspective. In the *positive error condition* shown here, one fruit was the expected outcome whereas three fruits were an unexpected outcome (more than expected). In the *negative error condition*, the trials were the same except the demonstration phase always showed three fruits rather than one. In the *number control condition*, the trials were the same except the number of fruits shown in the demonstration phase matched the final outcome.

Figure S1: Photographs of the experimenter demonstration from the monkey's perspective. Photos show the steps for trials in *positive error condition.* The experimenter opened the box flap with the visual cue to reveal one fruit in the demonstration phase. After repeating this motion four times, she opened the box to reveal either one or three fruits as the test outcome. During the test trial, the experimenter looked down after attracting the monkey's attention and then starting the trial, to avoid making eye contact with the monkey. In the first trial she manipulated the left box, and in the second trial she manipulated the right box.

Supplemental Results: Study 1

As reported in the main text, our first set of analyses for study 1 examined duration of looking across conditions and trial types. Table S1 reports the parameter estimates for the best fit model from those analyses.

Predictor	**Estimate**	**S.E.**	***t* value**	***p* value**
Sex (reference: female)	0.338	0.255	1.326	= 0.19
Age	-0.170	0.026	-6.475	< **0.0001**
Outcome order (reference: one fruit first)	0.193	0.254	0.761	= 0.45
Outcome fruit number (reference: one fruit)	0.620	0.342	1.810	= 0.07
Condition negative (reference: control)	0.761	0.395	1.926	= 0.05
Condition positive (reference: control)	-0.639	0.394	-1.619	= 0.11
Outcome fruit number : Condition negative	-0.932	0.484	-1.925	= 0.06
Outcome fruit number : Condition positive	1.137	0.484	2.348	**= 0.02**

Table S1: Predictors of looking times in Study 1. Parameters are from the best-fit model (model 3: full model); reference level is noted in the table as relevant.

As reported in the main text, our second set of analyses for study 1 examined a difference score (looking to three fruits – looking to one fruit) across conditions. Table S2 reports the parameter estimates for the best fit model from those analyses.

Predictor	**Estimate**	**S.E.**	***t* value**	***p* value**
Sex (reference: female)	0.155	0.395	0.393	= 0.69
Age (linear in years)	-0.072	0.041	-1.768	= 0.08
Outcome order (reference: one fruit first)	0.415	0.394	1.054	= 0.29
Condition negative (reference: control)	-0.918	0.484	-1.897	= 0.06
Condition positive (reference: control)	1.147	0.483	2.377	**= 0.02**

Table S2: Predictors of difference scores in Study 1. Parameters are from the best-fit model (model 2: full model); reference level is noted in the table as relevant.

Supplemental Results: Study 2

As reported in the main text, our first set of analyses for study 2 examined duration of looking across trial types and age cohorts (juveniles, adolescents, adults, and older adults as an ordinal factor). Table S3 reports the parameter estimates for the best fit model from those analyses.

Predictor	**Estimate**	**S.E.**	***t* value**	***p* value**
Sex (reference: female)	0.433	0.197	2.203	**= 0.03**
Age cohort L (linear effect)	-1.977	0.291	-6.791	**< 0.0001**
Age cohort Q (quadratic effect)	0.629	0.287	2.190	**= 0.03**
Age cohort C (cubic effect)	-0.459	0.284	-1.616	= 0.11
Trial (reference: trial 1 - expected)	1.562	0.162	9.668	**< 0.0001**
Age cohort L : Trial number	-0.101	0.327	-0.308	= 0.76
Age cohort Q : Trial number	-0.183	0.323	-0.567	= 0.57
Age cohort C : Trial number	0.979	0.319	3.072	**= 0.002**

Table S3: Predictors of looking times in Study 2. Parameters are from the best-fit model (model 3: age cohort x trial type interaction model); reference level is noted in the table as relevant.

As reported in the main text, our second set of analyses for study 2 examined a difference score (looking to three fruits – looking to 1 fruit) across age cohorts. Table S4 reports the parameter estimates for the best fit model from those analyses.

Predictor	**Estimate**	**S.E.**	***t* value**	***p* value**
Sex (reference: female)	0.246	0.267	0.918	= 0.36
Age cohort L (linear effect)	-0.105	0.328	-0.322	= 0.75
Age cohort Q (quadratic effect)	-0.171	0.324	-0.530	= 0.60
Age cohort C (cubic effect)	0.957	0.320	2.992	**= 0.003**

Table S4: Predictors of difference scores in Study 2. Parameters are from the best-fit model (model 2: age cohort model); reference level is noted in the table as relevant.

Supplemental Results: Adult performance in Study 2

In an additional set of analyses of the data from Study 2, building on the analyses reported in the main text, we then compared performance across the entire adult sample (individuals over age 6 years). Here we examined continuous age effects using linear mixed models (LMMs). We used this as an additional test of the results from our primary analyses that did not detect a difference between adults and older adults, when these ages were coded as categorical order cohorts.

We first examined duration of looking in the adult sample using an approach that paralleled our primary analyses of looking times reported in the main text. The base model accounted for *subject identity* (as a random effect), *sex*, and *trial type* (expected outcome on trial one, versus the unexpected outcome on trial two). We then added *age* (in years) as a linear predictor in the second model, which did not improved fit [$\chi^2 = 1.58$, $df = 1$, $p = 0.21$; AIC = 1173.0 compared to 1167.3 in the first model]. We added the interaction between *age* and *trial type* to test whether monkeys' responses to expected versus unexpected outcomes were modulated by age, which also did not improve fit [$\chi^2 = 0.36$, $df = 1$, $p = 0.55$; AIC = 1179.4]. AIC comparisons further supported the

base model as the best-fit model. That is, there was no statistical change in age with adults and older adults in overall looking times, in line with the results reported from the primary analyses. Table S5 reports the parameter estimates for the full model from those analyses.

Predictor	**Estimate**	**S.E.**	***t* value**	***p* value**
Sex (reference: female)	0.449	0.257	1.748	= 0.08
Age (linear)	-0.046	0.034	-1.368	= 0.17
Trial (reference: trial 1 - expected)	0.862	0.488	1.765	= 0.08
Age : Trial number	0.022	0.037	0.596	= 0.55

Table S5: Predictors of looking times in adult only sample for Study 2. Parameters are from model 3 (the age x trial type interaction model, as shown in the parallel analysis reported in Table S3); the best-fit model in this analysis was model 1 which included the main effect of trial type. Reference level is noted in the table as relevant.

We similarly examined difference scores in the adult sample using linear regression. We accounted for *sex* in our base model. We then added *age* as a linear predictor in years to the second model, which did not improve model fit compared to the base model [$\chi^2 = 0.47$, $df = 1$, $p = 0.49$; AIC = 622.6 compared to 621.1 in the first model]. Finally, we added the interaction between *age* and *sex* which also did not improve fit [$\chi^2 = 0.17$, $df = 1$, $p = 0.68$; AIC = 624.5]; AIC values also indicated that the base model was the best-fit model. That is, there was no shift with age in the adult sample, mirroring the results reported in the primary analyses. Table S6 reports the parameter estimates for the full model from those analyses.

Predictor	**Estimate**	**S.E.**	***t* value**	***p* value**
Sex (reference: female)	0.413	0.339	1.217	= 0.23
Age (linear)	0.025	0.037	0.681	= 0.50

Table S6: Predictors of difference scores in adult only sample for Study 2. Parameters are from model 2 (age model, as shown in the parallel analysis reported in Table S4); the best-fit model here did not include age (model 1). Reference level is noted in the table as relevant.

Supplemental Movie Caption

Movie S1: Experimental demonstrations and example monkey looking responses. The first two videos show an example experimental demonstration of the procedure, comprising (1) an example of the expected outcome in the *positive condition*, and (2) an example of the unexpected outcome in the *positive condition*. Across conditions, the experimenter manipulated boxes on a white stage while a monkey observed. In the positive condition, monkeys first experienced that the visual cue predicted one fruit in four initial demonstrations, and then saw either one fruit (expected) or three fruits (unexpected) in the test event. In the *negative condition*, the same visual cues predicted three fruits in the demonstration phase; the expected outcome then revealed three fruits, whereas the unexpected outcome revealed one. In the number control, the demonstrations and outcomes matched on both trials. In these example videos, the expected event was implemented as trial 1 (trial one always involved the left box from the monkey's perspective, containing oranges), and the unexpected event was implemented as trial 2 (the right box, containing apples). In Study 1, the order of these trials was counterbalanced across conditions, whereas monkeys in Study 2 always saw the positive condition in this order. For every test trial, the experimenter attracted the monkey's attention before lifting the flap, and then looked down concurrently while saying "now" to initial the 10s test trial; the experimenter held this position for the duration of the trial. Videos 3, 4, and 5 show example coding clips illustrating monkey looking responses in the task. As illustrated in these videos, coding clips start a few second before the primary experimenter said "now" and do not contain any information about the monkey's assigned condition so that coders could assess them blind to condition. On the videos, the secondary experimenter (cameraperson) then says "stop" after at least 10s have passed; clips were always coded for exactly 10s from the moment the trial started to equate total trial duration across monkeys.