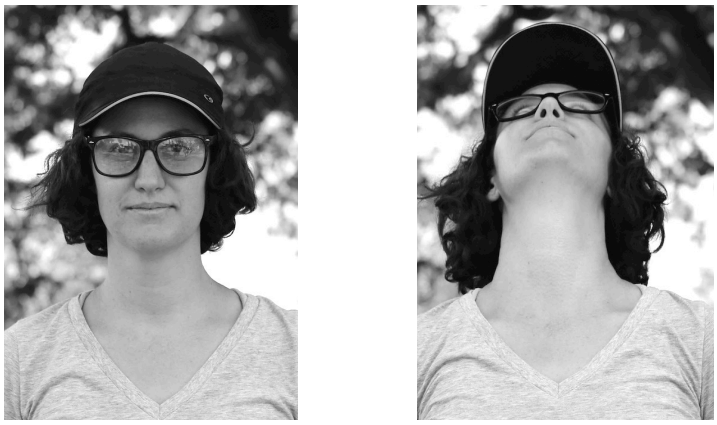
Supplemental Information
Rhesus monkeys show human-like changes in gaze following across the lifespan
A.G. Rosati, A.M. Arre, M.L. Platt, & L.R. Santos

Subjects

We tested semi-free-ranging monkeys from the Cayo Santiago population. These monkeys live in natural social groups, are provisioned with monkey chow at feeding corrals each day (in addition to access to plants growing on the island), and have ad-lib access to water. Monkeys are identifiable by unique combinations of tattoos and ear notches, and are familiar with human experimenters. Infants younger than a year do not yet have a tattoo, but their birthdates and sex are recorded in the census in association with their mother. Consequently, to test infants without tattoos, we identified babies in close proximity to an identifiable female (e.g., clinging to a female for an extended period or nursing from her) and then identified that female's baby in the census.

General Methods

As reported in the main text, two experimenters approached a calmly sitting monkey (standing 1-2 m away from the monkey). Experimenter 1 (E1) attracted the monkey's attention to her face; once the monkey was looking she said "now" and looked directly up (Figure S1). Experimenter 2 (E2) filmed the monkey's face for the 10s after E1 looked up. As required by research protocols at the Cayo Santiago site, both experimenters always wore a hat and glasses for eye protection.



(a) **(b)**

Figure S1: Experimental demonstration. (a) E1 stood 1-2m away from a calmly sitting monkey and attracted their attention. (b) Once the monkey looked at her, she looked straight up.

Subject Exclusions

In Study 1, additional monkeys were approached by the experimenters for testing but were not included in the analyses because they ran away before being successfully identified (n = 12). Monkeys had to complete at least one complete trial where they were appropriately watching the experimenter when she called "now" and looked up, and they remained visible throughout the 10s trial (e.g., a monkey who ran out of sight before the 10s was complete could not be properly coded for gaze-following). An additional 5 monkeys were excluded because they

were originally tested on only one trial, and subsequently the blind coders scored that to be a bad trial (e.g., monkey not looking when E1 said "now" or the monkey ran out of shot). If a given trial was scored as bad but that individual successfully completed other trials, we analyzed only those successful trials (an additional 27 trials were excluded from monkeys that were otherwise included in the dataset; as a total 1397 trials were included in the analyses, this means we excluded a minority of only 1.8% of trials). If monkeys were tested in more than one session (for example, because they were identified only after the test and it turned out they had been tested previously), we used only the first session where they successfully completed at least one trial for analyses.

In Study 2, we re-tested monkeys that had previously completed Study 1. Monkeys had to successfully complete at least the first two trials (e.g., one *Up* and one *Down* trial) to be included in the study due to the within-subjects design. Consequently, additional monkeys were approached by the experimenters for testing but were not included in the analyses because they only completed one trial (n = 10), or because the coders scored that the monkey was not looking when the primary experimenter called "now" on one of the first two trials (n = 4). An additional 2 trials (from the second half of the test) were excluded from monkeys who successfully completed the first two trials and were therefore included in the final sample.

Coding

We first clipped out individual trials (from longer session videos) using the program MPEG Streamclip. Each trial was clipped starting a few second before E2 called "now" and ending after E2 said "stop." Each resultant trial clip was assigned a new random identifying number that randomized the order of trials across all monkeys tested in Study 1 and Study 2. This allowed us to code each trial blind to trial number, as well as the total number of trials each monkey completed. The second phase of data collection (in 2014-15) randomized trials from Study 1 and Study 2, and therefore coding was also blind to study and condition (as Study 2 included both trials where E1 looked up, and control trials where she looked down).

All trials were scored by two independent coders. A primary coder examined all trials from the entire study. A first reliability coder scored the first half of sessions (data collected in 2013), and a second reliability coder scored the data collected in the second half (2014-2015). For each trial clip, coders independently identified the start of the word "now" and examined each trial frame-by-frame (30 frames = 1s) for the 10s after this start time. We coded: (1) whether the monkey ever looked straight upwards (using either their entire head or eyes only) in the ten second period of the trial; (2) the total duration of time the monkey spent looking up; (3) latency to initially look up; and (4) total number of discrete looks the monkey made (e.g., looking up, looking back down or at the experimenter, and then gazing up again; see Video S2).

Reliability between the primary coder and the reliability coders were high for all of these measures. For the first reliability coder, agreement for whether or not the monkey looked up was Kappa = 0.92; correlation of total time spent looking up was $r_p = 0.96$; correlation for the total number of discrete looks was $r_p = 0.87$; and correlation for latency to look up was $r_p = 0.92$. For the second reliability coder, agreement for whether or not the monkey looked up was Kappa = 0.94; correlation of total time spent looking up was $r_p = 0.95$; correlation for the total number of discrete looks was $r_p = 0.92$; and correlation for latency to look up was $r_p = 0.90$.

In Study 2, we did not additionally code whether the monkeys looked downwards, for two reasons. First, we had chosen the upward look test initially because assessing when the monkey looked straight upwards was relatively straightforward; when the monkeys tended to

turn their entire head to look up, whereas monkeys who looked down did not necessarily do so (given that the experimenter was standing in front of the monkey, it was not necessary for the monkey to turn their entire head to look in this down direction). Second, monkeys in this population more rarely produced looks directly upwards in their normal behavior, but often looked to the side or downwards in the context of foraging or scanning social group members in this naturalistic context where monkeys were tested while free-ranging. Consequently, their performance in the *Up test* versus the *Down control* is an assessment of whether the monkeys were more likely to look up after the experimenter looked in that specific direction, compared to the control where her behavior was identical but she looked in different direction.

Study 1 Supplementary Results

Statistical analyses

We analyzed data using the *glmer* function from the LME4 software package [1] in R [2]. We fit binomial models to a logit link function using maximum likelihood, including random subject intercepts to account for repeated trials within subjects. GLMM can account for unequal repeats across subjects [3], as was the case for our free ranging subjects. We conducted post-hoc Tukey comparisons of model factors using the *glht* function in the multcomp package [4], and compared the fit of different models with likelihood ratio tests [5].

Subject trial completion and performance by cohort and sex

As reported in the main text, we used GLMM modeling to account for unequal trial completion across individuals. Table S1 reported trial completion and percentage of individuals looking upward across the four trials, broken down by age cohort and sex.

	Percent individuals looking up for females, males, and overall			
	Trial 1	*Trial 2*	*Trial 3*	*Trial 4*
Infants	F = 25, 44% M = 22, 27% **n = 47, 36%**	F = 17, 35% M = 16, 44% **n = 33, 39%**	F = 11, 36% M = 11, 45% **n = 22, 41%**	F = 8, 38% M = 8, 25% **n =16, 31%**
Juveniles	F= 79, 61% M = 70 69% **n = 149, 64%**	F = 56, 55% M = 59, 44% **n = 115, 50%**	F = 45, 47% M = 45, 51% **n = 90, 49%**	F = 29, 45% M = 36, 36% **n = 65, 40%**
Adults	F = 126, 55% M = 114, 41% **n = 240, 48%**	F = 103, 48% M = 96, 41% **n = 199, 44%**	F = 73, 45% M = 85, 33% **n = 158, 39%**	F = 57, 40% M = 71, 25% **n =128, 32%**
Older Adults	F = 32, 31% M = 13, 8% **n = 45, 24%**	F = 25, 28% M = 10, 30% **n = 35, 29%**	F = 20, 30% M = 9, 0% **n = 29, 21%**	F = 18, 17% M = 8, 25% **n = 26, 19%**

Table S1: Trial completion and percentage individuals gazing upwards, by cohort and sex. F = number of female subjects in that cohort who completed that trial, M = male subjects in that cohort who completed that trial.

Analysis of habituation within cohorts

As reported in the main text, we examined what factors predicted gaze following responses within each cohort separately. For each separate cohort, we used GLMM modeling to test the importance of *trial number*, *sex*, and *age* (as a linear predictor within each cohort) by

sequentially adding predictors to a base model including only subject as a random factor, and then retained predictors that improved model fit. This allowed us to determine the specific factors that predicted gaze following within each cohort separately.

(1) Gaze following in infants (n = 47). In infants, initial models adding trial and sex did not improve model fit, compared to a base model with only subjects as a random factor [*trial model* versus *base model:* χ^2 =1.29, df = 1, p = 0.26, n.s.; *sex model* versus *base model*: χ^2 = 0.65, df = 1, p = 0.42, n.s.]. This indicates the infants did not show habituation across trials, and also did not differ by sex. However, adding in age as a linear predictor did improve fit [*age model* versus *base model*: χ^2 = 17.89, df = 1, p < 0.001], indicating that propensity to follow gaze increased as a function of age in this cohort. The parameters from a full model including all predictors are reported in Table S2.

Factor	*Estimate*	*S.E.*	*Z*	*P*
Trial number (covariate)	-0.470	0.304	-1.546	0.12
Sex (female baseline)	-0.841	1.106	-0.760	0.45
Age (covariate)	7.455	2.576	2.894	**< 0.005**

Table S2: Factors influencing propensity to follow gaze in macaque infants (Study 1). Parameters from the full model; the best fit model included only age.

(2) Gaze following in juveniles (n = 149). In juveniles, including trial increased model fit [*trial model* versus *base model*: χ^2 15.14, df = 1, p < 0.001], indicating that this age cohort showed habituation across trials. In subsequent models we then added sex and age, but neither of these terms further improved model fit [*sex-model* versus *trial-only model*: χ^2 = 0.01, df = 1, p = 0.95, n.s.; *age-model* versus *trial-only model*: χ^2 = 1.13, df = 1, p = 0.29, n.s.]. Thus, within juveniles there was not a sex difference in responses, and no major shifts by age within this cohort. The parameters from a full model including all predictors are reported in Table S3.

Factor	*Estimate*	*S.E.*	*Z*	*P*
Trial number (covariate)	-0.421	0.112	-3.769	**< 0.001**
Sex (female baseline)	0.058	0.290	0.200	0.84
Age (covariate)	-0.134	0.124	-1.078	0.28

Table S3: Factors influencing propensity to follow gaze in macaque juveniles (Study 1). Parameters from the full model; the best fit model included only trial number.

(3) Gaze following in adults (n = 240). In adults, including trial increased model fit as in juveniles [*trial model* versus *base model*: χ^2 = 13.76, df = 1, p < 0.001], indicating that this age cohort also showed habituation across trials. In addition, adding sex improved model fit as well [*sex-model* versus *trial-only model:* χ^2 = 6.92, df = 1, p < 0.01]. However, adding age did not improve model fit [*age-model* versus *sex and trial model*: χ^2 = 0.40, df = 1, p = 0.52, n.s.]. Thus, adults showed flexible habitation like juveniles, but sex differences in responses also emerged in this age group. The parameters from a full model are reported in Table S4.

Factor	*Estimate*	*S.E.*	*Z*	*P*
Trial number (covariate)	-0.287	0.082	-3.495	**< 0.001**
Sex (female baseline)	-0.604	0.226	-2.670	**< 0.01**
Age (covariate)	-0.028	0.045	-0.632	0.53

Table S4: Factors influencing propensity to follow gaze in macaque adults (Study 1). Parameters from the full model; the best fit model included trial number and sex.

(4) Gaze following in older adults (n = 45). In older adults, in contrast, none of these predictors improved model fit compared to the base model [trial model versus base model: χ^2 = 0.30, df = 1, p = 0.58, n.s.; sex-model versus base-model: χ^2 = 1.91, df = 1, p = 0.17, n.s.; age-model versus base-model: χ^2 = 0.15, df = 1, p = 0.70, n.s.;]. Thus, older adults showed less flexible control over the responses as evidenced by the lack of habituation across trials. The parameters from a full model including all predictors are reported in Table S5.

Factor	*Estimate*	*S.E.*	*Z*	*P*
Trial number (covariate)	-0.108	0.202	-0.536	0.59
Sex (female baseline)	-0.833	0.646	-1.291	0.20
Age (covariate)	0.004	0.092	0.041	0.97

Table S5: Factors influencing propensity to follow gaze in macaque older adults (Study 1). Parameters from the full model; no predictors improved model fit compared to the base model.

Trial Completion

One possible explanation for our main results is that the differences in habituation responses across cohorts were due to age differences in the total number of trials that monkeys completed. We compared number of completed trials using an ANOVA with age cohort as a between-subjects factor. In fact, infants completed fewer total trials than adults [$F_{3,477}$ = 2.79, p < 0.05; posthoc Tukey test revealed only infant-adult comparison p < 0.05; infant mean 2.51 ± 0.18 trials; juveniles 2.81 ± 0.10; adults 3.02 ± 0.08; older adults 3.00 ± 0.19 trials]. This may reflect that infants were sometimes displaced or picked up by their mother during the test. However, this difference in trial completion cannot account for our overall results, as we found that while juveniles and adults exhibited habituation, but both infants and older monkeys did not.

Baseline tendency to look

A second possible explanation for the lifespan patterns of gaze following is that cohorts differed in their baseline tendency to look up. In particular, juveniles may have looked up more often than other cohorts in our experiment simply because they made more baseline upward looks in general. If this was the case, we predicted that juveniles would exhibit differences in their *latency* to look up after the experimenter did so, as well as the total *duration* of their looks. In terms of response latency, baseline looks should be equally likely to occur at any time during the 10s trial, whereas gaze-following responses would be more likely to occur following the experimenter's action. Thus, if juveniles made more baseline looks than other groups, they should show longer response latencies. To address this issue, we examined trial one performance for those individuals who did follow gaze. In fact, in an ANOVA comparing performance across age cohorts, we found no significant difference in response latency in the different groups [mean response latency for infants: 3.09 ± 0.74s; juveniles: 1.88 ± 0.20s; adults: 2.13 ± 0.22s; older

adults: 2.33 ± 0.55s; $F_{3,236} = 1.44$, $p > 0.2$, n.s.]. Indeed, if anything juveniles exhibited a slightly faster response than other groups. Furthermore, baseline looks upward are not oriented towards finding a specific target, and therefore should have a shorter duration than gaze-following responses. However, we similarly found no difference in the total duration of gazing upwards across cohorts [infant mean: 2.04 ± 0.44; juveniles: 2.14 ± 0.18; adults: 1.59 ± 0.14; older adults: 2.05 ±0.71; $F_{3,236} = 1.95$, $p > 0.12$, n.s]. Thus, these results do not support the possibility that differences in baseline looking rates account for our main findings.

Study 2 Supplementary Methods and Results

As described in the main text, we examined how responses on *Up* and *Down trials* related to the subject's age by modeling responses for each trial type separately. For each trial type, we first fitted a base model with random *subject* intercepts, and *condition order* (first trial up or down, counterbalanced across subjects) as predictors as in the previous analyses; we did not include trial number as a covariate in this analysis as the trials alternated between the two possible types (up and down). In a second model, we then added *age* as an additional predictor to test its importance. We found that including age significantly improved model fit when modeling responses to *Up trials* [$\chi^2 = 14.62$, df = 1, $p < 0.001$]: monkey's gaze following responses declined with age, replicating the basic results from Study 1 (see Table S6 for parameters from the full model).

Factor	*Estimate*	*S.E.*	*Z*	*P*
Condition order (*Down* baseline)	-1.127	0.528	-2.137	**< 0.05**
Age (covariate)	-0.211	0.070	-3.000	**< 0.005**

Table S6: Factors influencing propensity to follow gaze in *Up trials* (Study 2). Parameters from the full model; including age improved model fit compared to the base model.

In contrast, including age did not increase model fit for *Down trials* [$\chi^2 = 1.61$, df = 1, $p > 0.20$, n.s.]: baseline rates of looking upwards remained at similar low rates across ages (see Table S7 for parameters from the full model). Consequently, the main results concerning the developmental changes in gaze-following from Study 1 are unlikely to be due to shifts in baseline reactivity.

Factor	*Estimate*	*S.E.*	*Z*	*P*
Condition order (*Down* baseline)	-0.507	1.404	-0.361	0.72
Age (covariate)	-0.238	0.209	-1.135	0.26

Table S7: Factors influencing propensity to follow gaze in *Down trials* (Study 2). Parameters from the full model; including age did not improve model fit compared to the base model.

References

[1] Bates, D. 2010 The LME4 package: linear mixed-effects models using S4 classes. See http://www.R-project.org.

[2] R Development Core Team. 2014 A language and environment for statistical computing. See http://www.R-project.org.

[3] Baayen, R.H. 2008 *Analyzing linguistic data. A practical introduction to statistics*. Cambridge, MA, Cambridge University Press.
[4] Hothorn, T., Bretz, F. & Westfall, P. 2008 Simultaneous inference in general parametric models. *Biometrical Journal* **50**, 346--363.
[5] Bolker, B.M., Brooks, M.E., Clark, C.J., Geange, S.W., Poulsen, J.R., Stevens, M.H.H. & White, J.S.S. 2008 Generalized linear mixed models: a practical guide for ecology and evolution. *Trends Ecol Evol* **24**, 127-135.

Supplemental Video Captions

Video S1: Infant following gaze. This monkey was categorized in the infant cohort (less than one years of age). The demonstrator (E1) and camera person (E2) stood next to eat other, approximately 1-2 m away from the monkey. E1 can be heard calling the monkey's attention. When the monkey looks at her face, she says "now" and looks straight up, holding that position for the rest of the trial. E2 films the monkey's face and times the trial, saying "stop" after a full 10s has passed.

Video S2: Multiple discrete looks. In this trial, the adult male monkey makes three discrete looks upwards: looking up, looking away in a different direction, and then looking up again. Multiple discrete looks was one index of the cognitive processes underlying monkey gaze following, as it reflects attempts to locate the (absent) target.