Supporting Information Appendix: Experimental Evolution of Prepared Learning

1. Details of the Modeling

This model is logically similar to the model described in Dunlap & Stephens 2009, where we considered the effects of information reliability and certainty on whether selection should favor the evolution of learning or of n unlearned preference. Here we consider how the reliabilities of two modalities of stimuli affect how selection favors which modality animals should learn about best. Flies experience pairings of two modalities of stimuli: odor (denoted by roman letters), and color (denoted by greek letters). There are two odors and two colors. To make this explicitly the same as the experiment, we can say that for the odors stimulus A is amyl acetate and stimulus B is benzaldehyde, and that for the colors, α is aqua, and β is blue.

We imagine a learning scenario where individuals use quinine (Q), as an aversive unconditioned stimulus and the odor and color as potential conditioned stimuli. As with our experiment, flies experience two phases: and experience phase and a consequence phase.

There are four possible pairings with quinine in the experience phase:

$$Q+A\alpha \text{ or } Q+B\beta \text{ or } Q+A\beta \text{ or } Q+B\alpha$$

The probabilities of quinine being placed with A or B, or α or β, are independent. We can define these probabilities as $P(Q+A) = a$, and $P(Q+B) = 1\text{-}a$; and $P(Q+\alpha) = b$, and $P(Q+\beta) = 1\text{-}b$.

Now we define the possible outcomes in the consequence phase (Table S1).

Table S1. Table of all possible outcomes in the consequence phase.

Roman/Color Best		Greek/Odor Best
P(A best|Q+Aα)	Neither Predict	P(α best|Q+Aα)
P(B best|Q+Bβ)		P(β best|Q+Bβ)
P(B best|Q+Aα)	Both Predict	P(β best|Q+Aα)
P(A best|Q+Bβ)		P(α best|Q+Bβ)
P(B best|Q+Aβ)	Color (greek) predicts, but odor (roman) does not	P(α best|Q+Aβ)
P(A best|Q+Bα)		P(β best|Q+Bα)
P(A best|Q+Aβ)	Odor predicts, but color does not	P(β best|Q+Aβ)
P(B best|Q+Bα)		P(α best|Q+Bα)

These simplify into 2 independent probabilities: the probability that olfactory stimuli predict the best environment in the consequence phase, given any previous pairing in the experience phase, P(olfactory best | any pairing), and the probability that color stimuli predict the best environment, P(color best | any pairing).

We then assign a probability to the reliability of each stimulus with regards to the quinine pairing predicting the best environment, where O is the probability odor predicts the best environment, and C is the probability color predicts the best environment. We define the number of eggs oviposited by the female in the consequence phase as N, the error rate for the female as ε (with $1-\varepsilon$ being the eggs she lays in the substrate she "prefers"), and r being the proportion of eggs which survive to adulthood (with zero surviving on the "bad" media).

There are 16 different combinations possible of experience phase pairings and consequence phase outcome (Table S2).

Table S2. Possible pairings and outcomes in the two stimulus modality system.

Experience Phase	Probability of Pairings	Consequence Phase	Probability of "Best"	Selective Learning to Roman	Selective Learning to Greek
Q+Aα	ab	Roman only predicts	$O(1-C)$	$r(1-\varepsilon)N$	$r\varepsilon N$
Q+Bβ	$(1-a)(1-b)$				
Q+Aβ	$a(1-b)$				
Q+Bα	$(1-a)b$				
Q+Aα	ab	Greek only predicts	$C(1-O)$	$r\varepsilon N$	$r(1-\varepsilon)N$
Q+Bβ	$(1-a)(1-b)$				
Q+Aβ	$a(1-b)$				
Q+Bα	$(1-a)b$				
Q+Aα	ab	Both Predict	CO	$r(1-\varepsilon)N$	$r(1-\varepsilon)N$
Q+Bβ	$(1-a)(1-b)$				
Q+Aβ	$a(1-b)$				
Q+Bα	$(1-a)b$				
Q+Aα	ab	Neither Predict	$(1-C)(1-O)$	$r\varepsilon N$	$r\varepsilon N$
Q+Bβ	$(1-a)(1-b)$				
Q+Aβ	$a(1-b)$				
Q+Bα	$(1-a)b$				

We can now calculate the fitnesses for each of the two types of learners. We do this using the geometric mean. The fitness for the roman only learner (learning about odor only) is

$$ab\,O(1-C)\ln r(1-\varepsilon)N + (1-a)(1-b)\,O(1-C)\ln r(1-\varepsilon)N + a(1-b)\,O(1-C)\ln r(1-\varepsilon)N + (1-a)b\,O(1-C)\ln r(1-\varepsilon)N + ab\,C(1-O)\ln r\varepsilon N + (1-a)(1-b)\,C(1-O)\ln r\varepsilon N + a(1-b)\,C(1-O)\ln r\varepsilon N + (1-a)b\,C(1-O)\ln r\varepsilon N + ab\,CO\ln r(1-\varepsilon)N + (1-a)(1-b)\,CO\ln r(1-\varepsilon)N + a(1-b)\,CO\ln r(1-\varepsilon)N + (1-a)b\,CO\ln r(1-\varepsilon)N + ab\,(1-C)(1-O)\ln r\varepsilon N + (1-a)(1-b)\,(1-C)(1-O)\ln r\varepsilon N + a(1-b)\,(1-C)(1-O)\ln r\varepsilon N + (1-a)b(1-C)(1-O)\ln r\varepsilon N$$

The fitness for the greek-only learner (learning about color only) is

$$ab\,O(1-C)\ln r\varepsilon N + (1-a)(1-b)\,O(1-C)\ln r\varepsilon N + a(1-b)\,O(1-C)\ln r\varepsilon N + (1-a)b\,O(1-C)\ln r\varepsilon N + ab\,C(1-O)\ln r(1-\varepsilon)N + (1-a)(1-b)\,C(1-O)\ln r(1-\varepsilon)N + a(1-b)\,C(1-O)\ln r(1-\varepsilon)N + (1-a)b\,C(1-O)\ln r(1-\varepsilon)N + ab\,CO\ln r(1-\varepsilon)N + (1-a)(1-b)\,CO\ln r(1-\varepsilon)N + a(1-b)\,CO\ln r(1-\varepsilon)N + (1-a)b\,CO\ln r(1-\varepsilon)N + ab\,(1-C)(1-O)\ln r\varepsilon N + (1-a)(1-b)\,(1-C)(1-O)\ln r\varepsilon N + a(1-b)\,(1-C)(1-O)\ln r\varepsilon N + (1-a)b(1-C)(1-O)\ln r\varepsilon N$$

2. Details on the stimuli

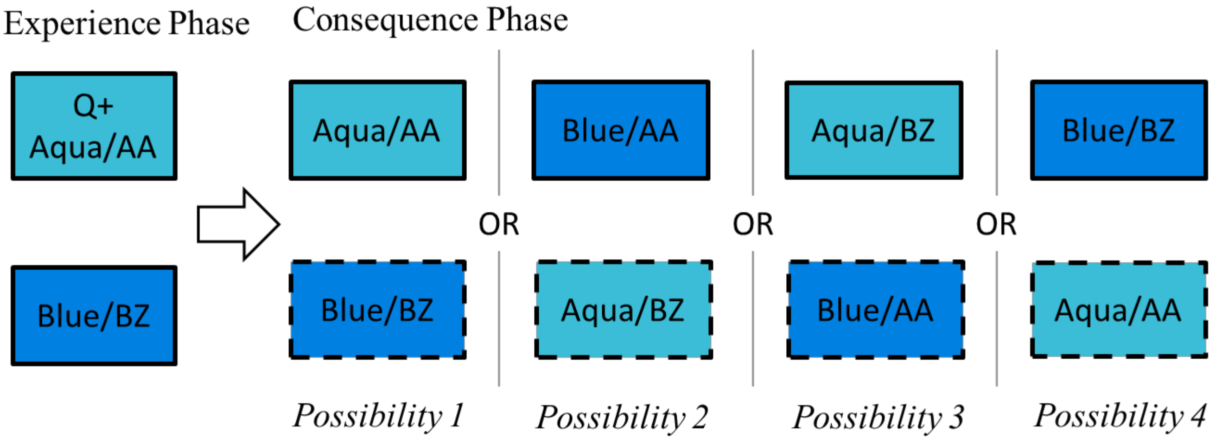
To avoid any effects of chemical detection of food coloring or dye agents, we used 100mm diameter circles of color placed under each 100mm diameter petri dish of clear agar. To choose the colors, we conducted an extensive series of pilot studies to determine that 1) flies would discriminate between the colors chosen, as evidenced by discrimination learning, and 2) flies showed no strong preference between the chosen colors. The colors we used in the experiment were aqua and cobalt blue acrylic poster paint from Dick Blick Art Supplies, which we painted onto white cardstock.

Through an additional series of pilot studies, we then titrated the amount of odorants such that learning about the olfactory cues did not completely overshadow learning about the color cues. Additionally, we balanced the preference across the two olfactory cues chosen.

3. Treatments and Generation Assignments

Prior to the experiment, we randomly assigned all the quinine pairings and selected substrates for each line. Figure S1 shows some example of how these pairings could occur using the color stimuli of aqua and blue (equivalent to A and B in the model) and the olfactory stimuli of amyl acetate (AA) and benzaldehyde (BZ) (equivalent to α and β in the model).

Figure S1. Example of possible pairings. Assume the best environment is indicated by the box with the dashed lines. In the first possibility, when paired with quinine, both aqua and AA reliably predict the best environment (Blue/BZ). In the second possibility, only the pairing of AA with quinine predicts the best environment (BZ); an individual attending to/ learning the aqua pairing would be making an incorrect choice. There are two additional possibilities: that aqua, but not AA predicts, and that neither predict.



We assigned a probability to the reliability of each stimulus with regards to how the quinine pairing predicted the best environment, where *O* was the probability the odor pairing predicted the best environment, and *C* was the probability that the color pairing predicted the best environment. By manipulating the quinine pairings and where we reared the eggs from, we created different patterns of reliability. For instance, in treatments where a modality had a reliability of 1.0, that modality always correctly predicted the best environment; in treatments where a modality had a reliability of 0.5, it predicted the best environment correctly in half of the generations (and incorrectly the other half of the generations).

To better explain this, the following tables give examples of how these assignments work in practice.

Table S3. This table shows a sample assignment for a line in the treatment where both modalities are reliable (C = 1.0, O = 1.0). The specific odor and color pairing remain identical during both phases.

Generation	Quinine with		Best Environment	
1	Aqua	BZ	Blue	AA
2	Blue	AA	Aqua	BZ
3	Aqua	BZ	Blue	AA
4	Blue	AA	Aqua	BZ
5	Blue	BZ	Aqua	AA
6	Aqua	AA	Blue	BZ
7	Aqua	AA	Blue	BZ
8	Blue	BZ	Aqua	AA
9	Blue	AA	Aqua	BZ
10	Aqua	BZ	Blue	AA

Table S4. This table shows a sample assignment where one modality is reliable, but the other is unreliable. In this case, color is unreliable and the generations where the odor+quinine pairings incorrectly indicate the best environment are indicated in red. In this case, during some generations, the specific color and odor pairs can change between the two experimental phases.

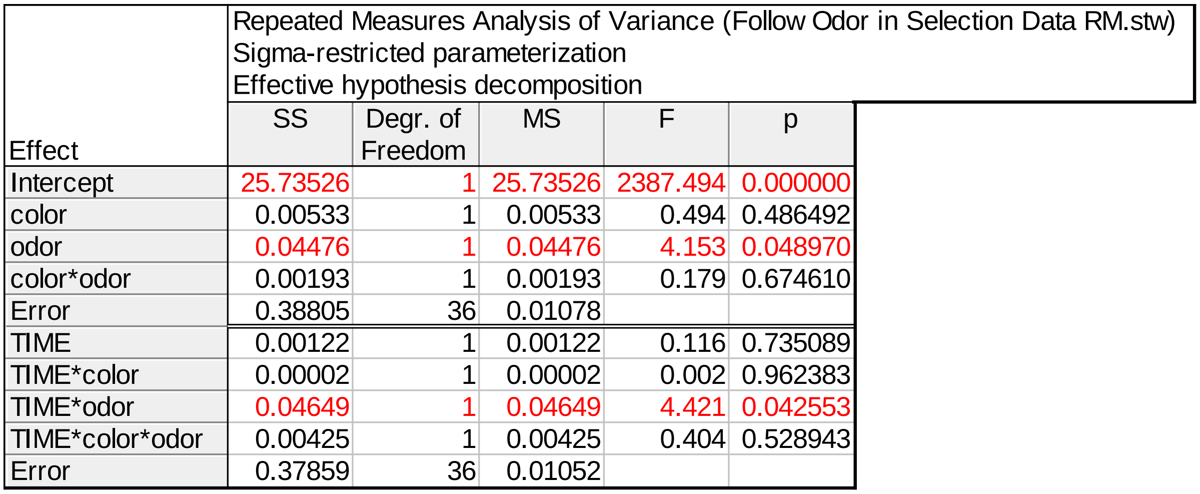
Generation	Quinine with		Best Environment	
1	Aqua	BZ	Blue	AA
2	Blue	AA	**Blue**	BZ
3	Aqua	BZ	**Aqua**	AA
4	Blue	AA	Aqua	BZ
5	Blue	BZ	**Blue**	AA
6	Aqua	AA	Blue	BZ
7	Aqua	AA	**Aqua**	BZ
8	Blue	BZ	Aqua	AA
9	Blue	AA	Aqua	BZ
10	Aqua	BZ	**Blue**	AA

We randomized assignments in blocks of two generations to ensure 1) an equal balance of the stimuli being paired with quinine and then serving as the best environment, and 2) to avoid runs of multiple generations with repeated assignments of a particular stimulus. Each of these was randomized and assigned independently.

4. ANOVA Tables for the Graphs in the Paper

The first set of tables is the full ANOVA tables corresponding to the analyses described in the paper.

Table S5. ANOVA table corresponding to Figure 2 in the paper. This is a repeated measure ANOVA on the learning scores for the **Odor**-quinine pairing, with effects of color reliability and odor reliability. The repeated measures are the learning scores for the two time points compared: the mean of the first two generations and last mean of the last two generations (denoted as Time in the table). This analysis focuses on the overall change across the entire experiment. Red values indicate statistically significant effects.



Repeated Measures Analysis of Variance (Follow Odor in Selection Data RM.stw)
Sigma-restricted parameterization
Effective hypothesis decomposition

Effect	SS	Degr. of Freedom	MS	F	p
Intercept	25.73526	1	25.73526	2387.494	0.000000
color	0.00533	1	0.00533	0.494	0.486492
odor	0.04476	1	0.04476	4.153	0.048970
color*odor	0.00193	1	0.00193	0.179	0.674610
Error	0.38805	36	0.01078		
TIME	0.00122	1	0.00122	0.116	0.735089
TIME*color	0.00002	1	0.00002	0.002	0.962383
TIME*odor	0.04649	1	0.04649	4.421	0.042553
TIME*color*odor	0.00425	1	0.00425	0.404	0.528943
Error	0.37859	36	0.01052		

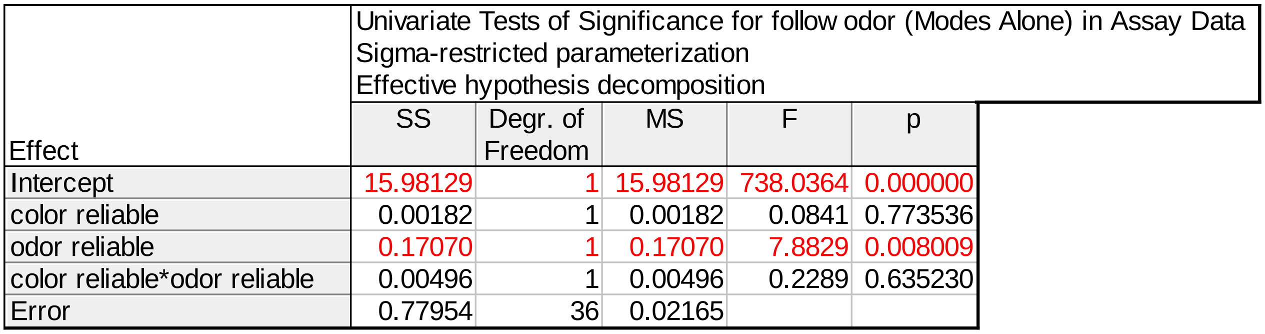
Table S6. ANOVA table corresponding to Figure 3 in the paper. This is a repeated measure ANOVA on the learning scores for the **Color**-quinine pairing, with effects of color reliability and odor reliability. The repeated measures are the learning scores for the two time points compared: the mean of the first two generations and the mean of the last two generations (denoted as Time in the table). This analysis focuses on the overall change across the entire experiment. Red values indicate statistically significant effects.

	Repeated Measures Analysis of Variance (Follow Color in Selection Data RM.stw) Sigma-restricted parameterization Effective hypothesis decomposition				
Effect	SS	Degr. of Freedom	MS	F	p
Intercept	24.72944	1	24.72944	3307.253	0.000000
color	0.05367	1	0.05367	7.177	0.011059
odor	0.00031	1	0.00031	0.041	0.841048
color*odor	0.00044	1	0.00044	0.058	0.810494
Error	0.26918	36	0.00748		
TIME	0.00036	1	0.00036	0.027	0.870091
TIME*color	0.05807	1	0.05807	4.378	0.043527
TIME*odor	0.01642	1	0.01642	1.238	0.273307
TIME*color*odor	0.00269	1	0.00269	0.203	0.655315
Error	0.47751	36	0.01326		

Table S7. ANOVA table corresponding to Figure 4 in the paper. This is a 2x2 factorial ANOVA, with factors of odor reliability and color reliability. The response variable is the proportion of eggs consistent with learning about the odor-quinine pairing. Red values indicate statistically significant effects.

	Univariate Tests of Significance for follow color (Modes Alone) in Assay Data Sigma-restricted parameterization Effective hypothesis decomposition				
Effect	SS	Degr. of Freedom	MS	F	p
Intercept	13.43271	1	13.43271	527.0866	0.000000
color reliable	0.10677	1	0.10677	4.1896	0.048026
odor reliable	0.00107	1	0.00107	0.0420	0.838851
color reliable*odor reliable	0.08755	1	0.08755	3.4354	0.072021
Error	0.91745	36	0.02548		

Table S8. ANOVA table corresponding to Figure 5 in the paper. This is a 2x2 factorial ANOVA, with factors of odor reliability and color reliability. The response variable is the proportion of eggs consistent with learning about the odor-quinine pairing. Red values indicate statistically significant effects.



Univariate Tests of Significance for follow odor (Modes Alone) in Assay Data
Sigma-restricted parameterization
Effective hypothesis decomposition

Effect	SS	Degr. of Freedom	MS	F	p
Intercept	15.98129	1	15.98129	738.0364	0.000000
color reliable	0.00182	1	0.00182	0.0841	0.773536
odor reliable	0.17070	1	0.17070	7.8829	0.008009
color reliable*odor reliable	0.00496	1	0.00496	0.2289	0.635230
Error	0.77954	36	0.02165		

5. Fine-Scaled Analysis of the Selection Data

The following analyses take into account learning scores throughout the experiment. Given the incredibly large numbers of eggs and petri dishes to count, the following data were all counted using a script we wrote for ImageJ, which assesses the total area of a petri dish covered with eggs. We then converted this area to an estimated number of eggs using a polynomial equation. To obtain this equation, we took a series of petri dishes with a known quantity of eggs, counted the area in ImageJ, and then fitted a predictive function with a high r-square value (0.997). This method accounts quite well for clumps of eggs, but is not as accurate as the human eye.

For the statistical analyses, we combine generations into blocks of 2 (corresponding to the randomization scheme of the experiment). For clarity in the graphs, we combine generations into blocks of 4 generations each.

Figure S2. This graph shows the proportion of eggs laid on the substrate consistent with learning about odor cues. We show the means across blocks of 4 generations each. Variance is high, as combinations of odors, colors, and quinine pairings are difference across the lines at any given time, due to the experimental treatments each generation. This is why we conduct separate learning tests at the end of the experiment, to give each line an identical series of tests.

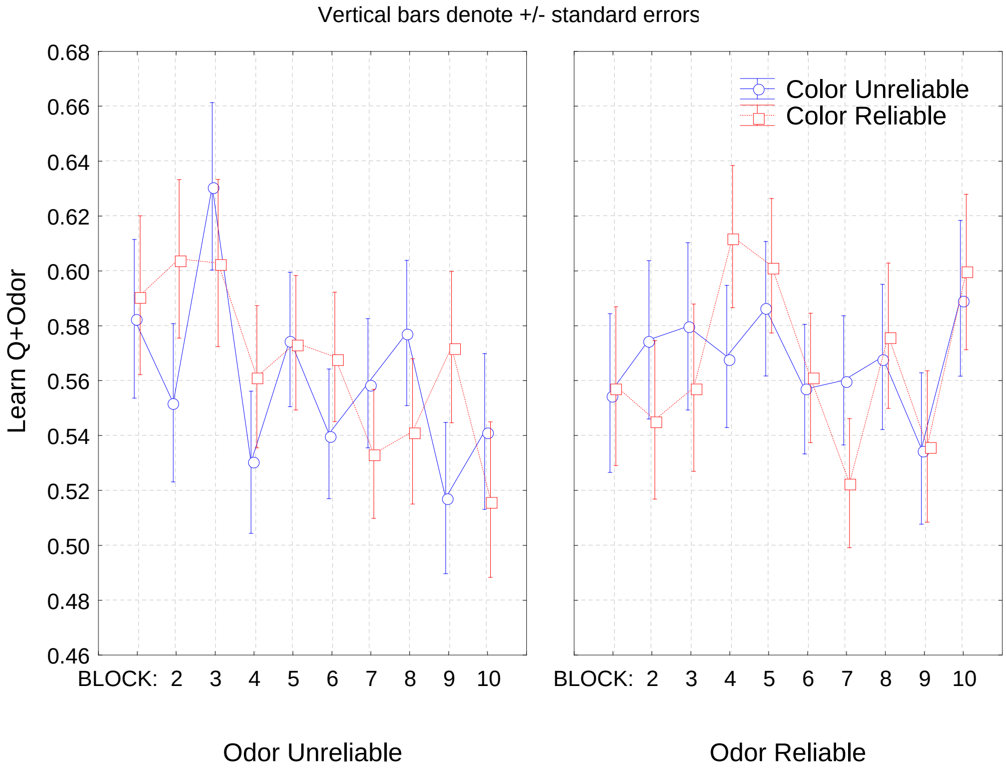


Table S9. Repeated measures ANOVA, with factors of color reliability and odor reliability. The repeated measure is of the blocks of 2 generations each (20 total blocks across 40 generations). To fit the assumptions of ANOVA, these data were arcsin transformed. Here we find a statistically significant interaction between the reliability and odor and the generations across the experiment. This is evident in Figure 2 above, where learning appears to decreases over time when odor is unreliable, but increases over time when odor is reliable. Red values indicate statistically significant effects.

	Repeated Measures Analysis of Variance (Follow Odor) in Selection Data RM Sigma-restricted parameterization Effective hypothesis decomposition				
Effect	SS	Degr. of Freedom	MS	F	p
Intercept	581.3473	1	581.3473	20105.96	0.000000
color	0.0013	1	0.0013	0.04	0.834000
odor	0.0019	1	0.0019	0.07	0.799278
color*odor	0.0023	1	0.0023	0.08	0.778011
Error	1.0409	36	0.0289		
Generation Block	0.3044	19	0.0160	1.13	0.315619
Generation Block*color	0.1746	19	0.0092	0.65	0.869988
Generation Block*odor	0.5067	19	0.0267	1.88	0.013034
Generation Block*color*odor	0.1483	19	0.0078	0.55	0.939449
Error	9.7020	684	0.0142		

Table S10. To confirm a difference between learning about odor-quinine pairings in the final generation block (generations 39 and 40), we ran a contrast within the ANOVA where we compare learning for lines in environments where odor is unreliable versus lines in environments where odor is reliable. The difference is statistically significant.

	Univariate Test of Significance for Planned Comparison (Follow Odor) in Selection Data Tests for transformed variables				
Variable	Sum of Squares	Degr. of Freedom	Mean Square	F	p
M1	0.093998	1	0.093998	9.874274	0.003347
Error	0.342703	36	0.009520		

Figure S3. This figure shows the learning about **Color**-quinine pairings, averaging across blocks of 4 generations each. This graph shows the proportion of eggs laid on the substrate consistent with learning about color cues. As with the previous figure, variance is high.

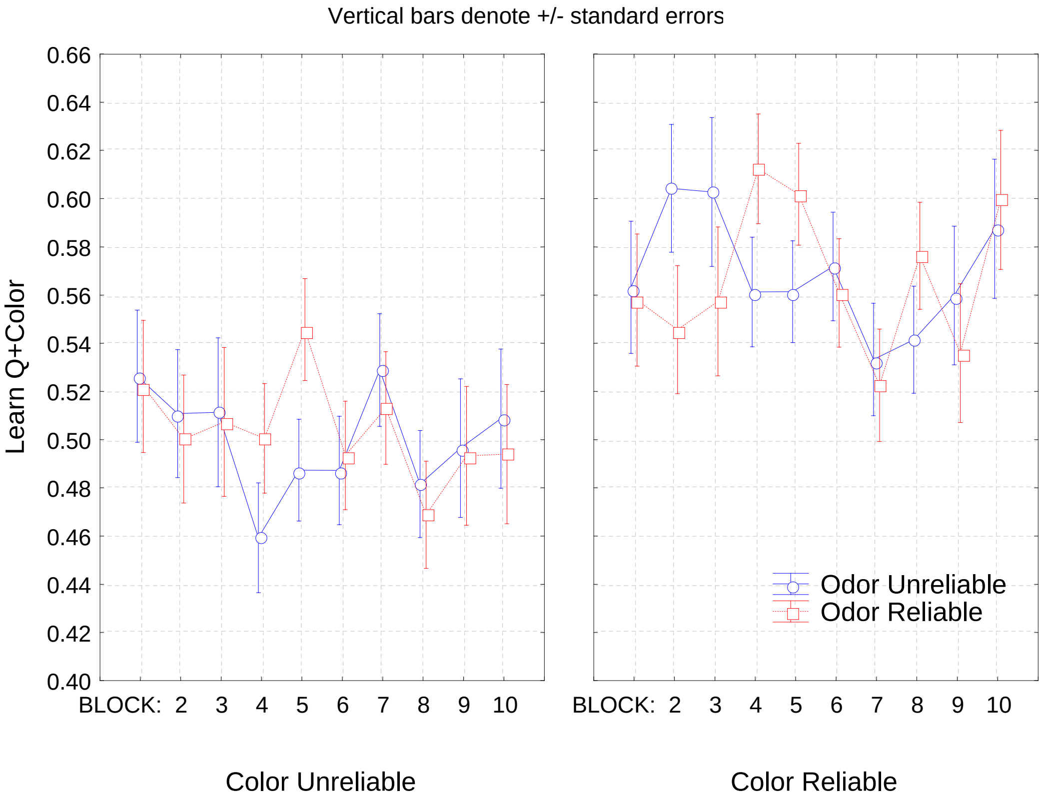


Table S11. Repeated measures ANOVA, with factors of color reliability and odor reliability. The repeated measure is on the blocks of 2 generations each (20 total blocks across 40 generations). To fit the assumptions of ANOVA, these data were arcsin transformed. Here we find a statistically significant effect of color reliability across the experiment. Unlike with the analysis with odor learning, we do not find a significant interaction with generation block, likely because for lines where color is unreliable, learning scores remain close to 50% throughout the experiment (they do not decline with time as with odor learning and odor unreliability).

	Repeated Measures Analysis of Variance (Follow Color) in Selection Data RM.stw) Sigma-restricted parameterization Effective hypothesis decomposition				
Effect	SS	Degr. of Freedom	MS	F	p
Intercept	539.6322	1	539.6322	27717.38	0.000000
color	0.9208	1	0.9208	47.29	0.000000
odor	0.0001	1	0.0001	0.01	0.942702
color*odor	0.0013	1	0.0013	0.07	0.794920
Error	0.7009	36	0.0195		
Generation Block	0.1763	19	0.0093	0.71	0.814595
Generation Block*color	0.3012	19	0.0159	1.21	0.245296
Generation Block*odor	0.2507	19	0.0132	1.00	0.453725
Generation Block*color*odor	0.2258	19	0.0119	0.90	0.577675
Error	8.9893	684	0.0131		

Table S12. Contrast analysis to confirm a difference between learning about color-quinine pairings in the final generation block (generations 39 and 40). There is a statistically significant difference for learning scores in lines where color is reliable from lines where color in unreliable.

	Univariate Test of Significance for Planned Comparison (Follow Color) in Selection Data RM.stw) Tests for transformed variables				
Variable	Sum of Squares	Degr. of Freedom	Mean Square	F	p
M1	0.117229	1	0.117229	11.05189	0.002045
Error	0.381858	36	0.010607		