Colour differences people cannot see fall more often between flowers and their backgrounds

A preregistered receptor-noise test on 728 measured Australian surfaces, with label-shuffle controls, a failed instrument check, and a replication under photon noise and on a second surface set

Zain Dana Harper

Version 1.0, 9 November 2026. Finding F81 in the findings ledger, narrowed by F88. Signal (tier B).

Abstract

Two surfaces can look the same colour to a person and differ sharply to a bee, which sees ultraviolet. We asked where such hidden differences fall among 728 measured Australian surfaces: 221 flower petals and 507 rocks, bark and leaves. Each observer was modelled with the receptor-noise-limited model and published receptor curves. Among pairs a modelled human cannot separate, modelled bees and blue tits separate flower-background pairs more often than background-background pairs, and the separation ratio exceeded its label-shuffle 95th percentile in every condition tested: two noise models on the first set, and two on a second set of Californian flowers against North American soil and leaf litter. The size of the ratio did not travel: for the bee it was 2.04 with Weber noise and 1.67 with photon noise added on the first set, and 45.6 on the second set, where the backgrounds are brown and alike. A preregistered instrument check failed under both noise models: the model resolves finer wavelength steps than bees show in behaviour, so only the direction of the effect is claimed.

1 What this means for a reader

A meadow that looks plain to us is full of colour differences we cannot see. In this collection of surfaces, those hidden differences cluster between flowers and what surrounds them, which fits the old idea that flower colours are tuned to the animals that visit them. The ultraviolet channel is textbook, and earlier work used this same data to study flower signals. What this paper adds is a direct count of the differences a person misses, with a control for chance and a check of how far the count can be trusted. The direction can be trusted on two sets of surfaces; the exact size cannot.

2 Introduction

Each animal's colour vision rests on a few receptor classes. People have three cone classes, the honeybee three with one in the ultraviolet, and the blue tit four with one in the ultraviolet. Two surfaces can give the same signals in a person's cones and different signals in a bee's, so the bee tells apart what the person cannot.

The standard way to compute such differences is the receptor-noise-limited model of Vorobyev and Osorio [1]. It treats colour as coded by comparisons between receptor classes and sets the threshold by the noise in each class. A distance of one just-noticeable difference (JND) is the threshold. The model was built to predict colour thresholds for animals whose eyes have been measured but whose thresholds have not.

The link between flower colour and pollinator vision has a long record. Chittka and Menzel [2] analysed 180 flower reflectance spectra and found that sharp steps in reflectance fall where bee-like insects discriminate best; their calculations of the best set of three receptors for telling flower colours apart matched the receptors measured in many Hymenoptera. Dorin, Shrestha, Garcia, Burd and Dyer [3] measured the surface set used here, flowers with rocks, minerals, bark and leaves from Australia, and found that the share of flowers with

sharp spectral steps was an order of magnitude above the share among background samples. They argued that insect colour vision is older than flowers and was shaped by flight among rocks and plants, and that flowers evolved colours that stand out against that background.

This paper asks a narrower question that the model can answer directly. Among the pairs of surfaces a person cannot tell apart, does a modelled bee or blue tit tell apart flower-background pairs more often than background-background pairs? Searches we ran on 9 November 2026 found the work above and no direct count of this kind; that is what the searches returned, not proof that none exists.

3 Preregistration and design

The first study was written as a preregistration on 29 October 2026 and locked at 01:27:38 UTC on 30 October. Before the lock, the code was run only on scrambled data, with every spectrum permuted across wavelength and the group labels permuted, and only keys, counts and timings were printed. The model was checked against its published closed forms on random numbers, to 4 times 10 to the power -16. No statistic of a real spectrum seen through any observer was computed before the lock.

The locked file names the data, the observers and their noise values with a grade for each source, the model, the measures, the hypotheses with their pass rules and kill conditions, the controls and the multiple-testing plan. This paper reports hypothesis H2, the flower test, with its secondary H2b and the instrument check V0. The same run tested other hypotheses about the information each observer carries; those are reported in the findings ledger and are not part of this paper.

H2 reads: among pairs one observer X cannot separate, the share another observer Y separates is larger for flower-background pairs than for background-background pairs. It was tested for four pairs of observers: human then bee, human then blue tit, dog then bee, and dog then blue tit. The kill condition: a ratio at or below 1, or inside the label-shuffle range. Each conditional share needs at least 20 pairs.

A second study, written after the first and locked at 02:40:37 UTC on 30 October, added photon noise to the model, reran the instrument check, and repeated H2 on the first set and on a second set of surfaces. Its preregistration was hash-locked in the same way. Its result narrows this finding and is reported here as the replication.

4 Data and provenance

Set 1 is the 728 reflectance spectra of Dorin and colleagues [4], under CC0: 346 rocks and minerals, 96 bark, sticks and dry plant material, 65 green leaves, 153 insect-pollinated and 68 bird-pollinated flower petals, measured from 300 to 700 nm. The files were fetched from the Zenodo mirror of the Dryad deposit, because the Dryad file interface needed a token; the archive hashes equal those Dryad publishes.

Human cones are the CIE 2006 two-degree cone fundamentals, and the light is CIE standard illuminant D65, both under CC BY-SA 4.0 [5]. Bee, dog and blue tit receptor curves, blue tit ocular media and two natural lights come from the data inside the pavo 2.10.0 R package [6], under GPL version 2 or later; they were read with a pure-Python reader and no package code was run.

Set 2, used only in the replication, joins 512 flower spectra from Californian serpentine seeps (LeCroy and colleagues [7], CC0) with 41 eastern North American forest-floor backgrounds: 20 soils, 10 dry litter and 11 wet litter samples (Kraemer and Adams [8], CC0). Flowers and backgrounds come from different deposits and places.

5 Methods

The model

For a surface with reflectance R under light I, the quantum catch of receptor class i with sensitivity S_i is

$$Q_i = \sum_{\lambda} R(\lambda) I(\lambda) S_i(\lambda) \tag{1}$$

summed over wavelength λ in photon units. With von Kries adaptation, q_i is Q_i divided by the catch from a perfect white, and the receptor signal is $f_i = \ln q_i$. For two surfaces, let d_i be the difference in f_i. For an observer with three receptor classes and noise e_1, e_2, e_3, the colour distance in JND is

$$S^2 = \frac{(e_1)^2(d_3 - d_2)^2 + (e_2)^2(d_3 - d_1)^2 + (e_3)^2(d_1 - d_2)^2}{(e_1 e_2)^2 + (e_1 e_3)^2 + (e_2 e_3)^2} \tag{2}$$

Two- and four-class observers use the general form of the same model [1]. Brightness differences are left out.

Noise

In the first study each receptor's noise is a Weber fraction, a fixed share of its signal. The human values are 0.063, 0.028 and 0.020 for the short, middle and long cones, from a long-cone fraction of 0.02 and relative cone densities of 1 to 5 to 10. The bee values are 0.13, 0.06 and 0.12. The blue tit uses 0.1 for its long-wave class with relative densities 0.37 to 0.70 to 0.99 to 1. The dog uses 0.05 for its long-wave class with densities 1 to 9. The preregistration grades these sources: the human densities and fraction are from memory with low confidence, the bee and blue tit values from memory of published sources with moderate confidence, and the dog value assumed.

The replication adds photon noise:

$$(e_i)^2 = (w_i)^2 + \frac{1}{N_i q_i} \tag{3}$$

where w_i is the Weber fraction above, q_i the catch, and N_i the photon catch from the white per receptor and integration time. N_i is N scaled by each class's relative density, with N set to 10,000. That value is an order-of-magnitude memory with low confidence. Noise is evaluated at the mean catch of the pair.

The test statistic

A human-blind pair is a pair with $S \le 1$ for the human model. Among the human-blind pairs, let P_{FB} be the share of flower-background pairs with $S > 1$ for the second observer, and P_{BB} the same share among background-background pairs. The separation ratio is

$$H = \frac{P_{FB}}{P_{BB}} \tag{4}$$

Dog-blind pairs are treated the same way. Flower-flower pairs are not counted.

Secondary test and named examples

H2b compares, for each flower, the median bee distance to all backgrounds between insect- and bird-pollinated flowers, with a one-sided label-permutation p value. Named examples were listed among human-blind flower-background pairs, with each surface's UV share.

Instrument check

The instrument check, V0, asks the model for the smallest step in wavelength between two pure spectral lights that reaches 1 JND, in steps of 0.1 nm. In the first study the human minimum had to lie between 0.5 and 5 nm and the bee minimum between 1.5 and 12 nm, against behavioural values of about 1 to 2 nm for people and 4 to 5 nm for bees. In the replication the rules were tightened to a factor of 2 of the behavioural values at named

wavelengths, for example a human step between 3 and 12 nm at 640 nm and a bee minimum between 2.25 and 9 nm. The behavioural values come from secondary sources and memory; the primary papers were not in hand. If V0 fails, absolute sizes from the model are not used.

6 Controls and planted checks

The control for H2 is a label shuffle: the flower and background labels are permuted over the same surfaces 500 times and the ratio recomputed. The null distribution of the ratio has a mean near 1, as it should. A real ratio must exceed the shuffle's 95th percentile. The p value from draws is the share of shuffles at or above the real ratio, counted with one added to both counts, so with 500 draws the smallest possible value is 0.002. The p values were then put through the Benjamini-Hochberg correction: over 7 tests in the first study and 14 in the replication.

Two code checks stand in for a planted check. The model's distances agreed with the published closed forms for two and three receptor classes to 4 times 10 to the power -16 on random inputs, and pavo's D65 equalled the CIE D65 to within 8 times 10 to the power -6. The instrument check is the known-answer test against behaviour. No surface set with a planted flower effect was run.

7 Results

The first study

Among pairs a modelled human cannot separate, modelled bees and blue tits separate flower-background pairs more often than background-background pairs, and the separation ratio exceeded its label-shuffle 95th percentile in every condition tested: two noise models on the first set, and two on a second set of Californian flowers against North American soil and leaf litter. In the first study, with Weber noise, the bee separated 0.98 of 1,108 human-blind flower-background pairs and 0.48 of 3,892 human-blind background pairs, a ratio of 2.04 against a shuffle 95th percentile of 1.08. The blue tit separated 0.99 against 0.45, a ratio of 2.22 against 1.09. Among dog-blind pairs the ratios were 1.45 for the bee and 1.16 for the blue tit, each above its shuffle percentile. All four had p 0.002 and passed the correction (Table 1).

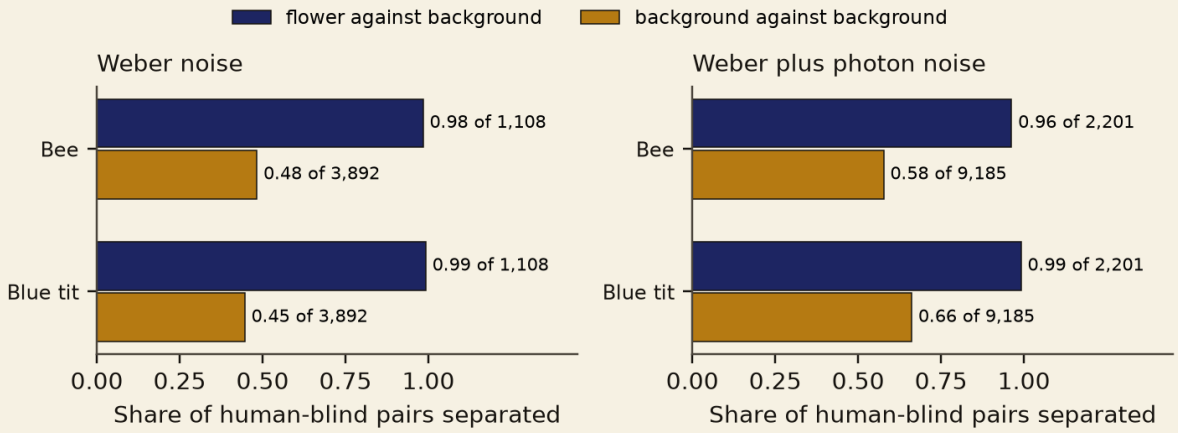


Figure 1. Share of human-blind pairs that the bee and the blue tit separate, set 1, under Weber noise and with photon noise added. Each label gives the share and the number of pairs. Drawn from the two runs' result files.

H2b passed for the bee: the median bee distance from a flower to the backgrounds was 9.47 JND for insect-pollinated flowers and 6.83 for bird-pollinated flowers (p 0.002). It is not specific to bees. The human model also separates insect-pollinated flowers more (13.7 against 9.4, p 0.010), and for the blue tit the order reverses

(9.65 against 10.16, p 0.79).

The named examples show what the count is made of. A petal of Hypericum pygmaeum against a dry, thin brown leaf sits 0.63 JND apart for the human model and 22.4 for the bee, with a UV share of 0.70 for the petal and 0.00 for the leaf. The reverse also occurs. The red, bird-pollinated Calothamnus graniticus sits 0.63 JND from fuchsite and 0.71 from white opal for the bee, and about 30 for the human model.

The instrument check

A preregistered instrument check failed under both noise models: the model resolves finer wavelength steps than bees show in behaviour, so only the direction of the effect is claimed. With Weber noise the model's smallest step was 0.4 nm for the human, at 615 nm, and 0.3 nm for the bee, at 540 nm. Weber-only noise lets a receptor with a near-zero catch report changes that photon noise would bury. With photon noise the human minimum rose to 0.8 nm at 500 nm, which passes, but the human step at 640 nm was 2.2 nm against a required 3 to 12 nm. The bee minimum was 0.7 nm at 540 nm, outside both the required range and the required place. The model is still too fine for the bee.

The replication

The size of the ratio did not travel: for the bee it was 2.04 with Weber noise and 1.67 with photon noise added on the first set, and 45.6 on the second set, where the backgrounds are brown and alike. With photon noise on set 1, more pairs became human-blind (2,201 flower-background and 9,185 background pairs), and the bee separated 0.96 against 0.58 of them. The blue tit's ratio fell from 2.22 to 1.50. Every one of the 12 replication ratios exceeded its shuffle percentile and passed the correction, with p 0.002 to 0.004. Figure 2 draws all 16 ratios against their shuffle ranges.

On set 2 the bee separated 0.98 of 1,374 human-blind flower-background pairs and only 0.022 of 465 background pairs. The 41 background spectra are brown and alike, so a bee separates few background pairs a human confuses. There the ratio measures how uniform the backgrounds are as much as anything about flowers.

Table 1. Separation ratios with their pair counts and label-shuffle 95th percentiles (500 shuffles each). FB and BB are the numbers of flower-background and background-background pairs the first observer cannot separate.

Test and condition	FB pairs	BB pairs	Ratio	Shuffle q95	p
Bee given human, set 1, Weber	1,108	3,892	2.04	1.08	0.002
Blue tit given human, set 1, Weber	1,108	3,892	2.22	1.09	0.002
Bee given dog, set 1, Weber	14,166	21,257	1.45	1.03	0.002
Blue tit given dog, set 1, Weber	14,166	21,257	1.16	1.02	0.002
Bee given human, set 1, photon	2,201	9,185	1.67	1.06	0.002
Blue tit given human, set 1, photon	2,201	9,185	1.50	1.04	0.002
Bee given dog, set 1, photon	16,557	26,035	1.38	1.03	0.002
Blue tit given dog, set 1, photon	16,557	26,035	1.14	1.01	0.002
Bee given human, set 2, photon	1,374	465	45.6	1.24	0.002
Blue tit given human, set 2, photon	1,374	465	3.85	1.21	0.002
Bee given dog, set 2, photon	3,071	595	9.80	1.06	0.002
Blue tit given dog, set 2, photon	3,071	595	2.37	1.05	0.002
Bee given human, set 2, Weber	566	316	313	1.31	0.004
Blue tit given human, set 2, Weber	566	316	52.3	1.26	0.004
Bee given dog, set 2, Weber	2,750	570	15.0	1.08	0.002
Blue tit given dog, set 2, Weber	2,750	570	2.40	1.05	0.002

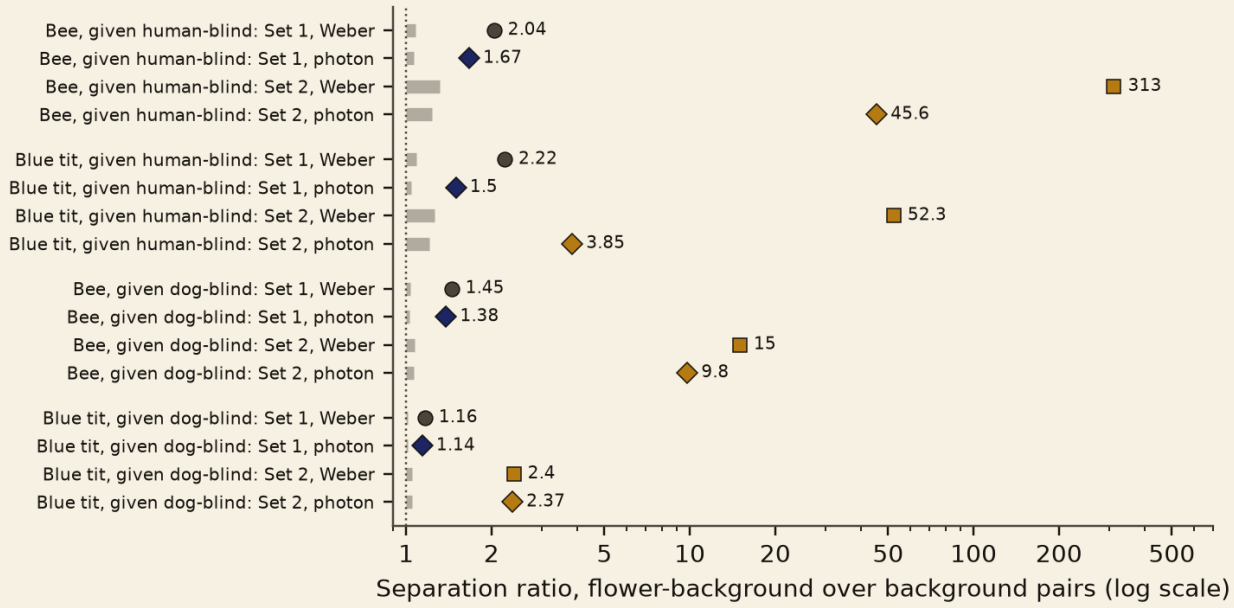


Figure 2. Separation ratios (markers) against the range from 1 to the label-shuffle 95th percentile (grey bars), for four observer pairs under two noise models on two surface sets. Every marker clears its bar; the sizes range over two orders of magnitude. Drawn from the two runs' result files.

8 Discussion

The result agrees with Chittka and Menzel [2] and with Dorin and colleagues [3], who used this same surface set: flower colours stand apart from their backgrounds in ways insect vision can see. This paper puts a different number on it. Of the colour differences a person cannot see, a larger share falls between a flower and a background than between two backgrounds, for a modelled bee and a modelled blue tit. The control for chance holds that number up in every condition.

The size of the effect is a property of the model and the surfaces together. Photon noise made more pairs human-blind and cut the bee's ratio from 2.04 to 1.67. A second set of surfaces with uniform brown backgrounds pushed it to 45.6. None of these values is the ratio in a real meadow. The direction, flowers ahead of backgrounds, survived both changes, and that is the claim this paper makes.

The instrument check marks the limit. The model resolves wavelength steps finer than bees can in behaviour. A model that is too fine for the bee would separate pairs a real bee cannot, and we expect that error to raise both shares in the ratio. The shuffle shows that the labels carry the difference between the shares, but the absolute shares should not be read as what a bee sees. Each failed check here is an honest null and is reported with the result.

H2b shows why care is needed with stories of design. Insect-pollinated flowers stand further from the backgrounds than bird-pollinated ones for the bee, and also for the human model. A colour that stands out for many eyes is not evidence of tuning to one.

9 Limitations

- The instrument check failed twice, with Weber noise and with photon noise, so the model is too fine for at least the bee. Only the direction of the ratio is claimed.
- Several receptor noise values come from memory or secondary sources, and the photon count is an order-of-magnitude memory. Absolute sizes depend on them.

- The model uses whole-surface spectra and ignores pattern, so nectar guides and other patterns inside one flower are out of reach.
- Backgrounds were not matched to each flower's own site, and set 2 joins flowers and backgrounds from different places. Two sets agreeing is not generality across the world's surfaces.
- Discrimination capacity in a model is not what the animal experiences, and no behaviour was tested.
- Dorin and colleagues studied flower signalling on set 1 before us, so set 1 was not a new sample for the general idea.
- The analysis was run by our own code and was not blind. For these reasons the finding is published at tier B, a signal worth expanding.

10 Verified

Verified: on 9 November 2026 the 728 surfaces and their group counts, every flower-background and background share with its pair count, every ratio with its shuffle percentile and p value, the H2b medians and p values, the named example distances and UV shares, the instrument-check values under both noise models, the set 2 group counts and the photon count were re-read from the two runs' result files. They match the findings ledger, including the ledger's later line that the set 1 photon ratios are 1.67 and 1.50 and that publication carries the direction only. The Calothamnus example, which the web write-up marked as not re-read, was re-read here from the first run's example list. The SHA-256 of each preregistration matches the hash recorded in its result file. Both figures were drawn by a script that reads only those two files. The studies were not re-run for this paper. The literature in the introduction was found by searches run on 9 November 2026 and read at the level of abstracts, publisher pages and search summaries; the full texts were not read.

11 Corrections

- 30 October 2026 (UTC): the replication run with photon noise lowered the set 1 ratios to 1.67 for the bee and 1.50 for the blue tit. A later dated line in the findings ledger narrows the published claim to the direction of the effect, and this paper follows it.
- 9 November 2026: first published version.

12 Reproduction

- Data: download the Dorin and colleagues spectra from Dryad (or its Zenodo mirror), the CIE 2006 cone fundamentals and D65, and the pavo 2.10.0 receptor curves; for set 2, the LeCroy and colleagues flower spectra and the Kraemer and Adams soil and litter spectra.
- Resample every spectrum to 300 to 700 nm in 1 nm steps, compute catches with eq. 1 under D65 with von Kries adaptation, and distances with eq. 2 and its general form; for the replication, use the noise of eq. 3 with N 10,000.
- Mark pairs at or below 1 JND for the human (or dog), compute eq. 4 for the bee and the blue tit, and compare it with 500 label shuffles; correct the p values by Benjamini-Hochberg at q 0.05.
- First study: preregistration SHA-256 fec1f1655ac86d3c16394e1947b9e2919225d9420680e5981e219f5a7f8fccd2, plan SHA-256 cc9c52ac2d9456618a10e1429f135574afe43466735b431b2ee002b6c9c6531f, run 20261030T012757Z.
- Replication: preregistration SHA-256 54571ccf85df9f48772bd0d03446598c3aa08cbdde1c29c9e27592895662e45d, plan SHA-256 95ba3678d3e9a2b1a3579aa1d1b1522ba59c6b699bf91761621d341517475a54, run 20261030T024058Z.
- Cost: the first run took 19.7 minutes of wall time on one desktop; the energy attributed to it by a CPU-share model, which is a model and not a measurement, was 1,645 J. The replication run took 7.3 minutes, with

1,995 J attributed the same way.

13 Appendix A. The preregistered verdicts

Each verdict is a set of rules written before the run, and it passes only when every rule holds. Values are read from the runs' result files.

Table 2. Preregistered verdicts for the flower test and the instrument check, and the rule that decided each.

Verdict	Deciding rule	Value	Holds
H2, bee given human	ratio above 1 and above its shuffle 95th percentile (1.08), corrected	2.04	yes
H2, blue tit given human	ratio above 1 and above its shuffle 95th percentile (1.09), corrected	2.22	yes
H2, bee given dog	ratio above 1 and above its shuffle 95th percentile (1.03), corrected	1.45	yes
H2, blue tit given dog	ratio above 1 and above its shuffle 95th percentile (1.02), corrected	1.16	yes
H2b, bee	insect minus bird median distance above 0, one-sided p below 0.05	2.64; p 0.002	yes
V0, first study	human minimum step 0.5 to 5 nm; bee minimum 1.5 to 12 nm	0.4; 0.3 nm	no
V0, replication	every model step within a factor of 2 of behaviour at named wavelengths	human 2.2 nm at 640; bee 0.7 nm	no
H2 replication, 12 tests	each ratio above 1 and above its shuffle 95th percentile, corrected	1.14 to 313	yes, 12 of 12

14 Glossary

Each term is defined where it is first used in the text. The same definitions, with the field each comes from and what each excludes, are collected here.

- **receptor class**. One kind of light-sensitive cell in the eye, defined by the wavelengths it responds to; humans have three cone classes, bees three including one for ultraviolet, blue tits four. Field: vision science. Excludes the rods used in dim light, and the nerve cells that compare receptor signals.
- **receptor-noise-limited model**. The model of Vorobyev and Osorio in which two colours can be told apart when the difference in log receptor signals is large compared with the noise in each receptor class; brightness differences are left out. Field: visual ecology. Excludes models of colour appearance, and any effect of pattern, motion or learning.
- **just-noticeable difference (JND)**. The unit of colour distance in the receptor-noise-limited model; a distance of 1 is treated as the threshold, so pairs at or below 1 count as indistinguishable for that observer. Field: psychophysics. Excludes a measured threshold for a real animal on a real task.
- **reflectance spectrum**. The share of light a surface reflects at each wavelength, here measured from 300 to 700 nm, which includes the ultraviolet below 400 nm. Field: optics and visual ecology. Excludes the colour a viewer sees, which also depends on the light and the eye.
- **preregistration**. A dated written plan for a study, fixed and hash-locked before any result is seen: the question, the data, the statistic, the controls, the forecasts and the kill conditions. Field: research method (metascience). Excludes a plan revised after looking at the outcome, and a protocol written only for the record after the run.
- **kill condition**. A rule written in the preregistration that, when met, decides against the claim with no room for judgement after the run. Field: research method (our usage, after Popper's falsification). Excludes reasons

found after the run for setting a result aside.

- **control**. A version of the data with the structure under test removed (for example shuffled or pooled), scored the same way as the real data, so the real score has something to beat. Field: statistics and experimental design. Excludes a comparison with a different dataset or with a published number.
- **photon noise**. The random variation in how many photons a receptor absorbs, which every real receptor has and which matters most when the catch is small; added here as a term that falls with the catch. Field: physics of light detection. Excludes noise in the nerve cells after the receptors.
- **quantum catch**. The number of photons a receptor class absorbs from a surface, the sum over wavelength of reflectance times light times the receptor's sensitivity, here divided by the catch from a perfect white under the same light. Field: vision science. Excludes the brightness a viewer reports.
- **von Kries adaptation**. Scaling each receptor class's catch by its catch from a white under the same light, a simple model of how eyes adjust to the colour of the light. Field: colour science. Excludes adaptation that mixes signals between receptor classes.
- **Weber fraction**. The noise in one receptor class as a fixed share of its signal, the same at every light level; we call a model with only this noise the Weber model. Field: psychophysics and vision science. Excludes photon noise, which grows in relative size as the catch falls.
- **human-blind pair**. A pair of surfaces whose modelled colour distance for the human trichromat is at most 1 JND, so the model says a person cannot tell them apart by colour; dog-blind pairs are defined the same way. Field: our usage, after the metamer of colour science. Excludes pairs a person cannot tell apart by brightness, which the model does not test.
- **separation ratio**. Among pairs one observer cannot separate, the share of flower-background pairs a second observer separates divided by the share of background-background pairs it separates; 1 means flowers gain nothing over the background. Field: our usage. Excludes how far apart the pairs are for the second observer.
- **insect- and bird-pollinated**. The dataset's labels for flowers whose recorded pollinators are insects only or birds only. Field: pollination ecology. Excludes flowers visited by both, which the dataset does not include in these groups.
- **UV share**. A surface's mean reflectance below 400 nm divided by its mean reflectance from 400 to 700 nm. Field: our usage. Excludes whether the ultraviolet forms a pattern on the petal.
- **instrument check**. A preregistered test in which the model must reproduce published behavioural measurements, here the smallest wavelength step each observer can detect; if it fails, only ranks and ratios from the model are used. Field: our usage; a known-answer test of the model against behaviour. Excludes a test of the model against its own output, and a planted check on synthetic data.
- **label shuffle**. A control that randomly reassigns the flower and background labels over the same surfaces, here 500 times, and recomputes the separation ratio, which shows how large it gets when the labels carry no meaning. Field: statistics (permutation test). Excludes a shuffle of the spectra themselves.
- **null distribution**. The spread of scores a control gives over many random draws; its 95th percentile is the bar a real score must clear. Field: statistics. Excludes a theoretical distribution assumed rather than drawn from the data.
- **p value (from draws)**. The share of control draws scoring at or above the real score, with one added to the count and to the number of draws, so it is never zero. Field: statistics. Excludes a p value from a textbook formula that assumes a distribution.
- **Benjamini-Hochberg correction**. A rule that adjusts a set of p values so that, among the tests called significant, the expected share of false calls stays at or below a chosen rate q, here 0.05. Field: statistics, multiple testing. Excludes the stricter Bonferroni rule, which bounds the chance of any false call.
- **planted check**. Data built so that the right answer is known, run through the same analysis before the real data, to show that the analysis can see the effect when it is there. Field: research method; called a positive

control in experimental science. Excludes any check made on the real data, which cannot show what the analysis misses.

- **honest null**. A working test that found no effect beyond a stated bound, a forecast that failed, or an error of our own found and fixed, reported with the same weight as a positive result. Field: research method (our usage). Excludes a void run, and a result left unpublished because it was negative.
- **tier**. How ready a finding is to publish: A candidate, B signal worth expanding (B* strong), C internal, D honest null worth reporting, E not reliable as it stands. Field: our publication register. Excludes a measure of how large or how true an effect is.
- **findings ledger**. The append-only, dated list of every result we have recorded, numbered F1, F2 and on; corrections are added as new dated lines and old entries are never rewritten. Field: our records. Excludes a curated list of successes.
- **verdict**. The pass or fail of one preregistered rule set against the run's numbers, read off by code. Field: research method (our usage). Excludes an interpretation written after reading the numbers.

15 References

[1] Vorobyev, M. and Osorio, D. (1998). Receptor noise as a determinant of colour thresholds. Proceedings of the Royal Society B 265, 351-358. doi:10.1098/rspb.1998.0302

[2] Chittka, L. and Menzel, R. (1992). The evolutionary adaptation of flower colours and the insect pollinators' colour vision. Journal of Comparative Physiology A 171, 171-181. doi:10.1007/BF00188925

[3] Dorin, A., Shrestha, M., Garcia, J. E., Burd, M. and Dyer, A. G. (2023). Ancient insect vision tuned for flight among rocks and plants underpins natural flower colour diversity. Proceedings of the Royal Society B 290, 20232018. doi:10.1098/rspb.2023.2018

[4] Dorin, A., Shrestha, M., Garcia, J. E., Burd, M. and Dyer, A. G. Reflectance spectra of rocks, minerals, sticks, bark, leaves and bird- and insect-pollinated flower petals (data deposit). Dryad, doi:10.5061/dryad.fn2z34v2c; Zenodo mirror record 10641891. CC0.

[5] CIE. CIE 2006 LMS cone fundamentals, 2 degree, and CIE standard illuminant D65. CIE datasets, doi:10.25039/CIE.DS.tijidesg. CC BY-SA 4.0.

[6] pavo: perceptual analysis, visualization and organization of spectral colour data. R package version 2.10.0, internal data (receptor sensitivities, ocular media and illuminants). GPL version 2 or later.

[7] LeCroy and colleagues. Flower reflectance spectra from Californian serpentine seeps. Dryad dataset v41ns1rtq, read through its Zenodo mirror. CC0.

[8] Kraemer and Adams. Soil and leaf litter reflectance spectra. Dryad dataset 4dv78, read through its Zenodo mirror. CC0.