

Paradoxical choice in rats: subjective valuation and mechanism of choice

Andrés Ojeda^{1*}, Robin A. Murphy² and Alex Kacelnik¹

Abstract:

Decision-makers benefit from information only when they can use it to guide behavior. However, recent experiments found that pigeons and starlings value information that they cannot use. Here we show that this paradox is also present in rats, and explore the underlying decision process. Subjects chose between two options that delivered food probabilistically after a fixed delay. In one option (“*info*”), outcomes (food/no-food) were signaled immediately after choice, whereas in the alternative (“*non-info*”) the outcome was uncertain until the delay lapsed. Rats sacrificed up to 20% potential rewards by preferring the *info* option, but reversed preference when the cost was 60%. This reversal contrasts with the results found with pigeons and starlings and may reflect species’ differences worth of further investigation. Results are consistent with predictions of the Sequential Choice Model (SCM), that proposes that choices are driven by the mechanisms that control action in sequential encounters. As expected from the SCM, latencies to respond in single-option trials predicted preferences in choice trials, and latencies in choice trials were the same or shorter than in single-option trials. We argue that the congruence of results in distant vertebrates probably reflects evolved adaptations to shared fundamental challenges in nature, and that the apparently paradoxical overvaluing of information is not sub-optimal as has been claimed, even though its functional significance is not yet understood.

Highlights:

- Rats sacrifice potential rewards by preferring an option that signals outcomes.
- Results match the predictions of the Sequential Choice Model (SCM).
- Latencies in single-option trials predict preferences in choice trials.
- Latencies in choice trials are the same or shorter than in single-option trials.

Keywords:

foraging; reaction time; sequential choice model; suboptimal choice

1. Introduction

Organisms frequently face tradeoffs between maximising immediate reward and gathering information. From a functional perspective this problem has been addressed in terms of the fitness contributions of exploration and exploitation (Berger-Tal et al., 2014; Krebs et al., 1978), while on a mechanistic level the focus has resided on the potential reinforcing properties of gaining information (Bennett et al., 2016; Blanchard et al., 2015; Igaya et al., 2016). The biological significance of maximising the rate of gains in dimensions like food is obvious, but that of seeking information is not, and in particular it is not clear whether information *per se* can act as a primary reinforcer in learning processes. When information is relevant to the subject's current problem and serves to act more efficiently, for instance by increasing feeding rate, its biological relevance is derived from the increase in efficiency (Krebs et al., 1978) but in situations when the available information cannot be used to modify action, any information-seeking behaviour is hard to interpret and hence interesting. Consider a choice between two stochastic sources of delayed reward, where the outcome of each individual choice can be obtained at a price, either before or immediately after one of the options is chosen. If outcomes are signalled before each choice, it will often be worth investing in information as this allows to choose the better option. However, if the outcome is signalled after the agent makes its choice, the information would appear to be worthless, as the subject cannot change its choice. Post-choice information, however, is considered significant both theoretically (Loomes and Sugden, 1982) and empirically (Palminteri et al., 2017) in humans, and, as we discuss below, has also proven influential in experimental choice experiments with non-human species. From a reward-maximizing perspective, for information that cannot be used to be valued is paradoxical, and hence worth exploring in some detail.

Several recent studies have found experimental conditions under which birds (pigeons and starlings) prefer a lower-reward but more informative option (Smith et al., 2016; Stagner and Zentall, 2010; Stagner et al., 2011; Vasconcelos et al., 2015; Zentall and Stagner, 2011). In these experiments (hereafter referred to as paradoxical choice) animals are given choices between two options that deliver reward probabilistically a fixed delay after being chosen. During the delay that follows the choice, in one of them (the informative option) signals are provided that anticipate whether a reward will be delivered or not, while in the alternative (non-informative) the outcome remains uncertain until the end of the delay. The critical

finding is that both bird species show a strong preference towards the informative option. Remarkably, this preference for the informative option can lead birds to lose up to (and perhaps more than) 80% of available food rewards (Vasconcelos et al., 2015).

Although several studies have now firmly established the existence of such paradoxical choice (for review see McDevitt et al., 2016; Zentall, 2016), its theoretical interpretation remains under debate. Broadly, the phenomenon can be interpreted in two ways. The first is that it reflects a maladaptive property of the psychological mechanism producing it. Indeed the phenomenon has been frequently labelled as *suboptimal behaviour* and proposed as an analogue of pathological human gambling (Zentall, 2011). It has also been claimed (Smith et al., 2016) that this phenomenon challenges Optimal Foraging Theory, which proposes that predictive models of animal behaviour can be produced by assuming that animals behave as if they were designed by natural selection to maximize Darwinian fitness (Stephens and Krebs, 1986). When the only variable affecting fitness is feeding rate, OFT does not predict any loss of food in exchange for unusable information. However, the predictions of OFT address natural, rather than laboratory, circumstances, and this can be a factor of major significance.

The second way to interpret this phenomenon takes into account the laboratory-natural world distinction, arguing that while preference for the post-choice informative option leads to sub-maximal payoffs under artificial conditions, the mechanisms underlying such preferences could be beneficial in the environment in which the animals evolved (Vasconcelos et al., 2015). It is hard to imagine natural circumstances in which information about forthcoming food events cannot be used to modify behaviour adaptively. In particular, a forager that perceived a signal telling that no reward would follow from its last action would not pay the cost of waiting for the null outcome; it would instead divert its attention to exploit alternative opportunities, something subjects cannot do in the typical laboratory experiments. This hypothesis has been formalized by postulating that in the experiments birds edit out any time waiting for sure no-reward when computing reward rates (Vasconcelos et al., 2015).

A question of importance for the theoretical interpretation of paradoxical choice is whether this phenomenon is restricted to birds or is more taxonomically widespread. If the same phenomenon were to be found in a mammal, this would support its interpretation as an adaptation to a common characteristic of natural foraging scenarios, rather than being a

maladaptive constraint on the cognitive competence of birds. Attempts to reproduce paradoxical choice in rats have so far yielded conflicting results. One research group has repeatedly failed to find evidence supporting paradoxical choice (Martínez et al., 2017; Trujano et al., 2016; Trujano and Orduña, 2015), leading the authors to conclude that a genuine species difference exists. Another group has found positive evidence (Chow et al., 2017) but only under certain specific experimental conditions that might not be equivalent to those used with the birds (Martínez et al., 2017). Because several protocol features varied between experiments, it is not clear what is driving the different results. Here we adapt the paradoxical choice protocol to rats incorporating novel technical details aimed at resolving this conflict: 1) We use auditory stimuli to signal reward contingencies, as this modality is well suited to the rat's discriminatory capabilities (Heffner and Heffner, 1985); 2) We use different stimuli to signal the four possible contingencies: Two for the informative option (one for sure reward and another for sure no reward) and two for the non-informative option (two signals, each followed by 50% chance of reward); 3) Before testing for preferences, we ensure that the rats discriminate between the contingencies; 4) We vary the magnitude of reward loss associated with the informative option.

Another modification that we introduce here is aimed at investigating whether a well-established psychological mechanism underlying choice in starlings and pigeons (the so-called Sequential Choice Model), is also present in rats. To this effect, in addition to collecting data on choice behaviour in simultaneous encounters, we record latencies to respond (reaction times) in both single option and choice trials. While many models of choice processes treat decisions as the result of a 'tug-of-war' between the stimuli cueing for options present at the time of each choice, the Sequential Choice Model (SCM) suggests that animals' decision mechanisms are adapted for sequential rather than simultaneous choices (Kacelnik et al., 2010; Shapiro et al., 2008). According to this model, the subjective value of each option is reflected in the latency to respond towards it when encountered in isolation, and choices in simultaneous encounters result from a race between these latencies. We examine this issue with our data, as a means to place the phenomenon of paradoxical choice in the more general framework of adaptive animal behaviour.

2. Methods.

2.1 Subjects

All experiments were carried out in compliance with the UK Animal (Scientific Procedures) Act (1986) and its associated guidelines. 16 male Lister Hooded rats (provider Envigo), 9 weeks old at the start of the experiment served as subjects. They were housed in groups of four. Throughout the experiment they were food deprived to 85-90% of their free-feeding weight using growth curves from the provider [Initial weight: $260\text{g} \pm 12$. Final Weight: $330\text{g} \pm 20$ (Mean \pm std. dev.)]. Water was provided ad-libitum in their home cages and they were maintained in a 12-hour dark/light cycle with lights on starting at 6 AM.

2.2 Apparatus

Testing was carried out in eight operant chambers (Med Associates, USA.) Each chamber contained three retractable levers: one in the back panel (centre) and two in the front panel, left and right of a central food magazine delivery tray equipped with an infrared beam and a sensor to record head entry. Each reward delivery consisted of two 45mg sucrose pellets (TestDiet, USA). A speaker (3.5" in diameter) was positioned above the back lever in the back panel. Each chamber was also equipped with a house light and a fan, which were switched on for the duration of the session. The chambers were controlled via custom-written Med-State Notation programs running on MED-PC V (Med Associates, USA).

2.3 General procedure.

The experiment was based on a trial-based chain procedure. On *choice* trials (**Figure 1**) responding on an attention manipulandum (the initial link in the response chain)) led to a choice between two other manipulanda (the options). Responding to either of these levers caused the onset of an acoustic cue that persisted for 10 seconds until an outcome occurred. Pressing the “*info*” (informative) option resulted in either the *i-plus* (sure reward) or *i-minus* (sure no-reward) cues. Responding to the “*non-info*” (non-informative) option resulted in either the *n1* or *n2* cues (both associated with 0.5 probability of reward). On *forced* trials only one of the options was available. We varied across treatments the probability of reward in *info*, with $p(i\text{-plus})$ being either 0.2, 0.4 or 0.5. As a consequence, the *info* option had a reward probability that was 60% lower, 20% lower, or the same as the *non-info* option, respectively. For clarity, we describe first the cues and then proceed to detail the training procedures and the final protocol.

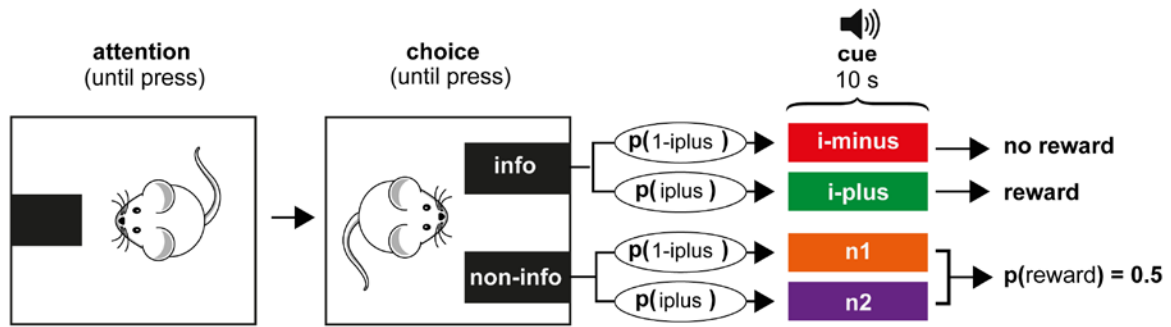


Figure 1. Choice Trial Structure

2.4 Cues

The four cues consisted of sounds, all with a duration of 10 seconds, and each associated to a reward probability. There were two cues for the informative option: *i-plus* (100% reward probability and *i-minus* (0% reward probability), and two cues for the non-informative option (*n1* and *n2*, both with 50% reward probability). The four sounds were: a pure tone (3 kHz, 78 dB), a buzzing sound (74 dB), white noise (74 dB), and a clicking sound (74 dB).

Assignment of sounds to reward probabilities was counterbalanced across subjects to avoid the possibility of option preferences being generated by any intrinsic aversive or attractive properties of the sounds.

2.5 Pre-training

2.5.1 Magazine training:

To train the rats to collect reward in the apparatus training began with a single session with 60 reward deliveries. The interval between reward deliveries was sampled from a Poisson distribution with a mean of 30 seconds.

2.5.2 Lever training:

Over the next three sessions the rats were trained to press the two front levers for food. The left lever was available on the first session, the right lever on the second and either lever (with equal probability across trials) on the third. Once a lever extended into the chamber, a single press resulted in its retraction and immediate reward delivery (fixed ratio 1 schedule). In the first two sessions, the lever became available again after 10 seconds. In the third session, this delay was composed of a constant element plus a variable one. The constant one was 5 seconds, and the variable one was sampled from a Poisson distribution with a mean of

5 seconds. All three sessions concluded after 60 reward deliveries.

2.5.3 Cue training

To ensure that in the main experiment the subjects were tuned to the reward probabilities of the four cues (*i-plus*, *i-minus*, *n1* and *n2*) we preceded the choice phase with a Pavlovian protocol in which the rats were passively exposed to the four cues, each with its associated reward contingency. Like in the main phase that was to follow, each cue (sound) was presented for 10s and reward was delivered (probabilistically) at the end of the stimulus, but in this phase no action was required to produce the cues. These cue-training sessions consisted of 120 trials, with 30 trials for each cue. To avoid large deviations from the expected reward probability of each cue, proportions of rewards were fixed for every 10 trials with each cue. These trials were then randomly intermixed. Thus, if for example a 50% reward probability was associated with a cue, exactly 5 out of every 10 trials with that cue ended in reward, but the order was randomised. The ITI (inter-trial interval) was composed by sampling a Poisson distribution with a mean of 15 seconds and adding 10 seconds. The subjects performed one daily session of this phase for 11 days.

2.6 Main experiment

2.6.1 Treatments

A within-subjects design was used, with each rat participating in the three treatments. In all treatments pressing the *info* (informative) lever resulted in the presentation of either the *i-plus* (sure reward) or *i-minus* (sure no-reward) cues, with the probability of *i-plus* differing between treatments (0.2, 0.4 and 0.5), while responding on the *non-info* (non-informative) lever caused either of the *n1* or *n2* cues, with the same probabilities as *i-plus* and *i-minus* in that treatment. Because both *n1* and *n2* delivered reward with 0.5 probability, the probability of reward in the *non-info* option was always 0.5, and treatments differed only on the probability of reward for the *info* option. Treatments were named according to the probabilities of reward in the two options: 1) 0.2 *info* vs 0.5 *non-info*, 2) 0.4 *info* vs 0.5 *non-info* and 3) 0.5 *info* vs 0.5 *non-info*. In all cases, the probability of reward in *info* (the probability of the cue *i-plus*) was equal to or lower than 0.5, hence the *info* option was at best neutral.

Subjects were split into two groups that performed the treatments in two different orders: “0.5/0.4/0.2” and “0.2/0.5/0.4” according to the probability of reward in the informative

option. Both groups performed the experiment concurrently. Subject assignment to treatment was counterbalanced such that there was no correlation between treatment and any of the following parameters: side of the informative option; hour of testing; cage in which the animals were housed, or sound-reward contingencies. For each treatment, the subjects performed one daily session for 14 consecutive days. Each rat was trained at the same time of day; one cohort of rats began the experiment at 7:30 A.M. while the other started at 11:30 A.M. A session finished after 120 trials or 3 hours, whichever occurred first. Between treatments, the *info* and *non-info* option lever sides were reversed for half of the rats, to control for carry-over effects of the spatial position of the options in the previous treatment.

2.6.2 Trial Structure

There were two kinds of trials: *choice* and *forced*. All trials started with the rear (attention) lever extending. Pressing this lever resulted in its retraction, and either one (forced trials) or both (choice trials) of the front levers to be presented (Figure 1). The purpose of requiring a back-lever press to start a trial was to ensure task engagement, to position the rat centrally in the chamber with respect to the options, and to enable measurement of latency to respond starting from a central position. Pressing a front lever initiated a cue (the terminal link) and also caused that lever (forced trials) or both levers (choice trials) to retract. The auditory cue was presented for 10 seconds, after which reward delivery could occur. Trials were separated by an ITI (inter-trial interval) generated by sampling from of a Poisson distribution with a mean of 15 seconds and adding 10 seconds.

2.6.3 Session Structure

Each session was composed of 120 trials: 80 *forced* (half *info* and half *non-info*) and 40 *choice* (randomly intermixed). Other trial sequence details were determined as previously described in the *cue discrimination training* section

2.7 Data analysis

All data processing and statistics was done in MATLAB 2017a. Before statistical analysis proportion data was arc-sine square root transformed to normalize the residuals (Grafen and Hails, 2002). When performing repeated measures analysis-of-variance the data were tested for sphericity with Mauchly's test, and if the compound symmetry assumption was not satisfied corrected p-values were calculated using the epsilon adjustment.

2.7.1 Cue training phase

The focus of this study is choice between actions leading to different cues and outcomes. It is thus important to establish if subjects were tuned to the programmed contingencies of each

cue. To this effect we quantified the total duration of head entries into the food magazine during cue presentations. To check for statistical significance of differences between cues we pooled the data from the last five sessions and performed a repeated-measures ANOVA with cumulative head poking time as the dependent variable and cue as a within-subject factor. Post-Hoc tests between all cue pairs were then performed using the Bonferroni correction for multiple comparisons.

2.7.2 Option preferences

To examine preference in choice trials, defined as the proportion of choice trials in each session allocated to the *info* or *non-info* options, we performed a repeated-measures ANOVA with treatment (categorical) and session (continuous) as within-subject factors, and preference for *info* as the dependent variable. To establish if asymptotic preference differed from indifference we pooled preference through the last five sessions of each treatment and compared it against a 0.5 expectation of indifference, using one sample t-tests.

2.7.3 Latency analysis

Testing the Sequential Choice Model required examining latencies to respond and predicting proportions for each option in choice trials using latency to respond in forced trials. To do this, for each session, we compared the latencies of 40 pairs of forced trials. Pairs were composed of a forced *info* trial and a forced *non-info* trial, following the order in which they were presented. As these were unique pairs, each latency was used only once for the prediction of choice proportions in that session. Predicted preference for *info* was defined as the proportion of forced trial latency pairs in which *info* had a shorter latency than *non-info* (for an example see Table 1 in Supplementary Methods). We then analysed statistically the predicted preferences, following the same procedure as for observed preferences.

To establish the statistical significance of the match between predicted and observed preferences we performed a permutation test (Golland et al., 2005), as follows. Pooling all data, the whole experiment yielded a total of 672 sessions, each one with a predicted and observed preference (*info* or *non-info*). We shuffled the predicted preferred options for each session in 10^6 random permutations, and determined for each permutation the percentage of sessions in which the predicted and observed preferred option matched. The probability of the observed proportion under the null hypothesis was calculated as the fraction of permutations with a higher match percentage than the actual (non-shuffled) match.

280

281 A difference between Tug of War models and the Sequential Choice Model (Shapiro et al.,
282 2008; Kacelnik et al., 2010) is that the former implies that latency to respond towards a given
283 option in choice trials should be longer than latency towards the same option in single option
284 trials, while the latter predicts the opposite. Further, the SCM predicts that the shortening of
285 latencies in choice trials (relative to forced trials) should be stronger for the non-preferred
286 than for the preferred option. This is caused by the assumption of parallel processing of each
287 option under the SCM model. As in each choice trial the option yielding a shorter latency to
288 act is expressed as the preferred alternative, the obtained distribution of latencies for each
289 option in choice trials has a richer representation of shorter latencies than the distribution for
290 the same option in the absence of competition. This bias should be greater the stronger the
291 censorship suffered by the original (no competition) latency, and this occurs for the least
292 preferred option. To test this, we determined the difference between single option and choice
293 trial latencies as a function of the level of preference for that option (complete, partially
294 preferred and partially non-preferred). This poses statistical difficulties because latency data
295 are right skewed and the sample size (number of trials) of each option in choice trials is
296 determined by the preferences of each subject. Therefore, we used a metric (the *latency*
297 *index*), that is robust in this respect.

298 For a given option, this latency index takes a value of 0 when all latencies for that option in
299 choice trials are shorter than the median latency for the same option in single option trials, 1
300 when all choice trials yield longer latencies than the median latency in forced trials, and 0.5
301 when choice and forced trials have the same median latency. The *latency index* was
302 computed as follows. For each session of each subject the following procedure was carried
303 out once (if all choice trials belonged to the same option) or twice (once for each option if
304 preferences were partial): 1) determine the median latency on forced trials (the *reference*) 2)
305 determined the number of choice trials of the option with a latency shorter than the *reference*
306 (*shorter* trials) and longer than the *reference* (*longer* trials). For each session with partial
307 preferences this resulted in two pairs of values (the number of *shorter* and *longer* trials for
308 the preferred and non-preferred options) while when preference was absolute (all choice trials
309 were allocated to the same option) one pair of values was obtained per session (the number of
310 *shorter* and *longer* trials for the option that was completely preferred). The total number of
311 *shorter* and *longer* trials for each of the three categories (completely preferred, partially
312 preferred, and partially non-preferred) was then calculated by summing across all sessions of

the same subject, and the index was computed by dividing the number of *longer* trials by the total number of trials (*shorter* plus *longer*). The SCM predicts values below 0.5, the Tug of War models predicts values above 0.5, and, of course, under the null hypothesis the index would approximate 0.5.

3. Results

3.1 Cue discrimination

The average cue discrimination performance during the five last sessions of pre-training is shown in **Figure 2**. The figure shows that the rats' head poking duration during cue presentations varied appropriately according to each cue's reward contingency. A one-way repeated-measures ANOVA with cue as a within-subject factor revealed a significant effect of cue: ($F_{3,45}=41.74$, $P<0.0001$). Differences between cues were confirmed by post-hoc pairwise comparisons, with significant differences between all cue pairs (Bonferroni-corrected, $P<0.001$), except for the *n1* vs *n2* pair, as should be expected given that these two cues signal the same reward probability. In the main experimental phase, the difference between the *i-minus* cue and all other cues was maintained (Bonferroni-corrected, $P<0.001$), but there was no longer a significant difference between the *i-plus* cue and the *n1* or *n2* cues ($P>0.05$, for details see Figure S1 in Supplementary Results). This confirms that the bias towards the *info* option could not be explained by a failure to discriminate the corresponding higher incidence of no-reward.

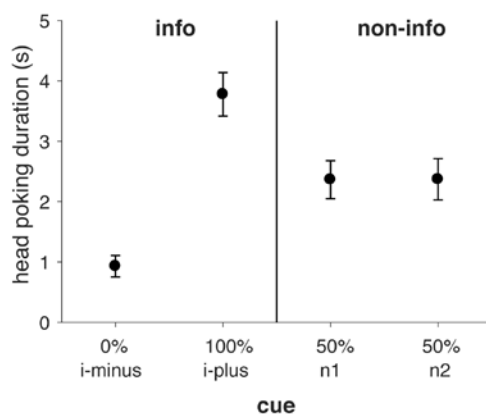


Figure 2. Cue discrimination

Mean \pm S.E.M. ($n=16$) duration of head poking into the food magazine during cue presentation during the five last sessions of the pre-training phase. The rats showed good discrimination between the cues associated with

the *info* (informative) and *non-info* (non-informative) options, and within the *info* signals when a reward was due and when it was not.

3.2 Option preferences (Choice trials)

Figure 3 shows that in choice trials the rats preferred *info* when its reward probability was the same or 20% lower than that in *non-info* (0.5 *info* vs 0.5 *non-info* and 0.4 *info* vs 0.5 *non-info*). However, when reward probability in *info* was 60% lower (0.2 *info* vs 0.5 *non-info*) rats preferred the *non-info* option (**Figure 3**). A two-way repeated-measures ANOVA with treatment and session as within-subject factors and preference as the dependent variable revealed a significant effect of treatment ($F_{2,30}=27.75$, $P<0.0001$), session ($F_{1,15}=268.19$, $P<0.0001$), and session by treatment interaction ($F_{2,30}=44.12$, $P<0.0001$). Pooled over the last five sessions, preference for *info* is significantly above 50% for treatments 0.5 *info* vs 0.5 *non-info* ($t_{15}=11.86$, $P<0.0001$) and 0.4 *info* vs 0.5 *non-info* ($t_{15}=9.18$, $P<0.0001$), and significantly below 50% for treatment 0.2 *info* vs 0.5 *non-info* ($t_{15}=3.87$, $P<0.002$). For individual subject data see Figure S2 in Supplementary Results.

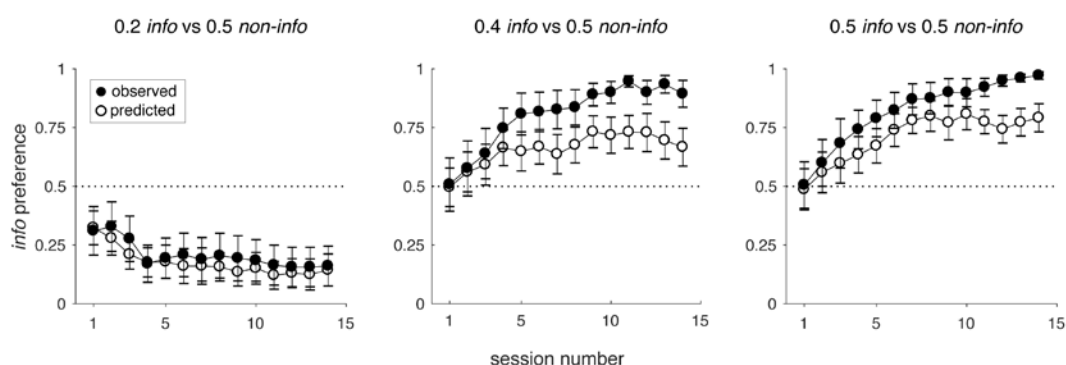


Figure 3. Observed and predicted preferences

Mean \pm S.E.M. ($n=16$) proportion of choices for the *info* (informative) option in each treatment. Each panel is a treatment with a different reward probability for the *info* option. Observed preferences in choice trials are

360 depicted in full symbols. Empty symbols show predicted preferences in choice trials derived from latencies in
361 forced (single-option) trials, according to an implementation of the Sequential Choice Model. The dotted line
362 represents indifference.

363

364

365 *3.3.1 Latencies (shortening)*

366 The Sequential Choice Model predicts differences in latencies to respond to each option in
367 forced (single-option) and choice trials (i.e. when the option was chosen), as well as
368 differences between preferred and non-preferred options (defined as being chosen above or
369 below 50% of choice trials, respectively). According to the model, when preferences are
370 partial latency to select an option on choice trials should be shorter than when the same
371 option is encountered on its own, due to cross-censoring of the distributions of latencies. This
372 shortening effect should be more pronounced for the option preferred less than 50% of the
373 trials, because its latency distribution is censored more strongly, than for the preferred
374 alternative. However, when an option is exclusively preferred, latencies in choices should
375 equal those in single-option trials, because when the alternative is never chosen there is no
376 censorship. To this effect we split latencies according to which was the preferred option for a
377 given subject and a given session, and whether the degree of preference was extreme
378 (absolute preference, hence no censorship), or partial (i.e. both options received at least some
379 choices). **Figure 4** shows the results. Latency to respond to each option was either
380 indistinguishable or shorter in choice trials relative to forced trials. When preference was
381 absolute, the chosen option showed no detectable difference between the two kinds of trials
382 ($t_{15}=1.57$ $P=0.13$). When preference was partial, both the preferred and non-preferred options
383 showed significantly shorter latencies in choice trials, and the shortening effect was stronger
384 for the non-preferred option ($t_{15}=2.4$, $P<0.03$ and $t_{15}=3.3$, $P<0.005$ for preferred and non-
385 preferred respectively).

386

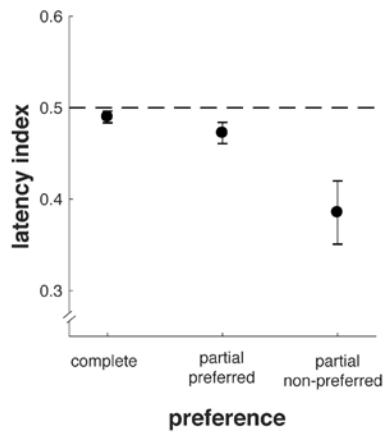


Figure 4. latencies (choice versus forced trials)

Mean \pm S.E.M. (n=16) latency index as a function of option preference. Values lower than 0.5 indicate that latency to respond to an option when the alternative is present (choice trials) is shortened with respect to when the same option is presented in isolation (forced trials). When preference was absolute, the index did not differ from 0.5, but when preference was partial, both preferred and non-preferred options showed significantly shorter latencies in choice trials. The shortening effect was stronger for the non-preferred option.

394

395 3.3.2 Latencies (choice predictions)

396 In addition to predicting the shortening of latencies in choice trials, the Sequential Choice
397 Model states that the difference in latencies between options in forced trials should predict
398 preference in choice trials. This analysis is shown in **Figure 3** (empty symbols). Predicted
399 preferences derived from latencies on forced (single-option) trials mirror observed
400 preferences on choice trials (Figure 3, empty versus solid symbols), albeit with predicted
401 preferences tending to be less extreme than observed ones. The procedure used to predict
402 preference from latencies is described in detail in the Methods section 2.7.3. A two-way
403 repeated-measures ANOVA with predicted preference as the dependent variable and
404 treatment and session as within-subject factors revealed a significant effect of treatment
405 ($F_{2,30}=47.38$, $P<0.0001$), session ($F_{1,15}=138.74$, $P<0.0001$), and treatment by session
406 interaction ($F_{2,30}=64.16$, $P<0.0001$) Pooled over the last 5 sessions, predicted preference for
407 *info* was significantly above 50% for treatments 0.5 *info* vs 0.5 *non-info* ($t_{15}=4.99$, $P<0.0002$)
408 and 0.4 *info* vs 0.5 *non-info* ($t_{15}=2.78$, $P<0.02$), and significantly below 50% for treatment 0.2
409 *info* vs 0.5 *non-info* ($t_{15}=5.67$, $P<0.0001$), just as for the experimental data. The predicted
410 preferred option for a session matched the actual preferred option in 88% of all sessions. This
411 match is significantly above chance (permutation test, $P<0.0001$).

412

413

414

415 **4. Discussion.**

416 We discuss first the specific mechanism of paradoxical choice with respect to outcome
417 information and then the more general problem of how subjective valuation of alternatives
418 leads to choice behavior.

419 *4.1 Paradoxical choice in rats*

420 Our results show that rats, like pigeons and starlings, prefer an informative over a non-
421 informative option, to the extent of showing systematic deviations from reward rate
422 maximization. Rats' preference for an informative option is not absolute, but depends on how
423 much loss it causes: rats preferred an informative option yielding a 50% or 40% reward
424 probability against a non-informative alternative offering 50% reward probability, but when
425 the informative option offered a 20% reward probability they switched preference towards
426 the higher yielding non-informative option. Birds (pigeons and starlings), in contrast, seem to
427 go as far as preferring an informative option yielding as little as 10% reward probability
428 against a 50% non-informative alternative (Vasconcelos et al., 2015; Smith et al., 2016).

429 A potential critic could argue that the difference in reward probability between 0.4 info and
430 0.5 non-info is not large enough for the rats to be able to detect it and thus their preference
431 for the informative option is not really paradoxical. With the present data we cannot know if
432 the animals can discern that the informative option has a lower probability of reward and are
433 still willing to incur the cost of selecting it, or rather if they fail to discriminate the two
434 reward probabilities. This ambiguity however does not make the behavior non-paradoxical. If
435 preference for an informative option leads an animal to prefer an option with a lower reward
436 probability, then the mechanism that generates this behavior is costly regardless of what the
437 animals "know". Our results are 'paradoxical' because the rats were not indifferent: they
438 systematically preferred an option causing a 20% loss of reward. The precise reasons for this
439 paradoxical preference do of course, admit further investigation.

440 Rats' preference for the informative option cannot be attributed to a putative inability to
441 discriminate between the terminal link cues, as the (Pavlovian) nose-poking behaviour
442 recorded in the presence of each cue during the pre-training phase (Figure 2) showed that
443 subjects could discriminate between cues according to their reward contingencies. This was

also the case for starlings, which, when given a choice between the terminal links showed orderly preference for the higher payoff alternative (Vasconcelos et al., 2015). Interestingly, in the rats, during the main experimental condition the cue for absence of reward continued to be associated with a suppressed Pavlovian response during the waiting time, but the three cues associated with a possibility of reward elicited similar durations of head-poking even though one signaled sure reward and the other two partial reinforcement with equal chance of reward or no reward. Although this is curious in its own right, it cannot explain the preference for *info* as that would presumably manifest itself as a lack of discrimination between the cue for the absence of reward and rest of the cues. Also, it does not impact our conclusions as their behaviour during the pre-training phase demonstrates that the rats can discriminate between the cues, even if they do not explicitly show it during the main experiment.

The fact that the phenomenon is present in birds and mammals argues against labelling it ‘sub-optimal choice’ (Zentall, 2011), as its taxonomic ubiquity suggests that the systematic failure to maximize reward in this specific laboratory situation probably reflects a psychological mechanism maintained and/or reinstated by natural selection under natural conditions. The challenge lies in identifying the critical differences between natural and experimental situations.

The comparison between rat and bird results requires some caution. Previous work with rats in similar protocols has had conflicting results, calling for a detailed analysis of the protocols used in each case. In the present experiments we used four auditory cues, one signaling sure reward, another one sure lack of reward, and 2 more associated with an equal chance of reward or lack of it. We also ensured that the animals did discriminate between the four contingencies (Figure 2). These tests serve to reject the possibility that differences between tests in birds and rats reflect differences in discrimination and not in valuation.

A previous study also found evidence of paradoxical choice in rats (Chow et al., 2017), but using a protocol that had an important difference relative to ours. Chow and co-workers used one cue (a lever) to signal partial reinforcement in the non-informative option and another cue (another lever) to signal sure reward in the informative option, signalling sure no-reward implicitly by a blackout. Such a setup could have biased results in favour of the informative option by not providing a specific stimulus signalling sure no-reward. Martínez et al. (2017) made a similar argument stating that Chow et al.’s 2-cue setup lacks a stimulus that can

function as a conditioned inhibitor. To support their argument, Martínez and co-workers (2017) performed an experiment that also used levers as stimuli, but with a 4-cue setup (one lever for each contingency, including sure lack of reward) and reported that they were unable to find preference for the informative option. Given that we found paradoxical choice with a 4-cue setup, and these cues were highly salient, our result cannot be attributed to the lack of a specific stimulus signalling sure no-reward. This still begs the question of why in Martínez et al. (2017) as well as other studies in rats (Chow et al., 2017; Trujano et al., 2016; Trujano and Orduña, 2015) paradoxical choice was not evident. One likely cause of failures to show the added value of information is that in some of these experiments the reward loss associated with choosing the informative option might have been beyond the range of what rats will pay. Martínez and co-workers (2017) gave rats a choice between an informative option offering a reward probability of 0.5 and a non-informative option offering 0.75, a loss of 33% of possible rewards. In our experiments rats reverted to preferring the higher reward but non-informative option when the loss associated with preference for the informative option increased from 20% to 60% of possible rewards. Therefore, Martínez et al.'s null finding is consistent with the results of one of our treatments. Nonetheless, other possible procedural differences call for detailed analysis. Several of the previous studies used lights as cues (Chow et al., 2017; Trujano et al., 2016; Trujano and Orduña, 2015) although the visual modality is not ideally suited to albino laboratory rats. Generalization is another potential issue. In the studies that used lights or levers as cues (Chow et al., 2017; Martínez et al., 2017; Trujano et al., 2016; Trujano and Orduña, 2015), what made the cues different from each other was their spatial position and this could have led to some degree of generalization between them.

In the present study, we used four auditory cues, one for each outcome, randomised the assignment of cues to outcomes, ensured that rats discriminated between these cues, and varied the magnitude of the loss in rewards associated with the informative option between no loss, a 20% loss, and a 60% loss. Our subjects preferred the higher probability, non-informative option for the more extreme condition, but showed a very strong and reliable preference for the informative option up to a loss of 20% (Figure 3). We surmise that the evidence for the presence of the phenomenon in rats is conclusive, and sensitive to the magnitude of the incurred loss.

The fact that rats' preference for an informative option is not strong enough to overcome large losses contrasts with the results found with pigeons and starlings (Smith et al., 2016; Vasconcelos et al., 2015), and may reflect a true species difference worth of further investigation. The argument put by Vasconcelos et al. (2015) to explain birds' extreme preference is that the *cue that signals a sure absence of a forthcoming reward does not contribute to the animal's computation of lost opportunity* (Vasconcelos et al., 2015; Smith et al., 2016), because in nature animals in such situations can re-direct behavior to other targets, so they would not incur the time cost of waiting for something that will with certainty not come. An associative learning account of the same phenomenon would posit that the cue for the absence of reward does not acquire inhibitory properties. This argument leads to expecting that preference for the informative option will be extreme regardless of its probability of reward, a prediction that has been supported by bird data (Vasconcelos et al., 2015; Smith et al., 2016) but does not apply to the picture that is emerging for rats.

4.2 Mechanism of choice: the Sequential Choice Model

The value a subject attributes to an option, and the process by which this valuation translates into choices between options are distinct and separable empirically, because subjective valuation is expressed also in no-choice situations by readiness to take an option. One particular model, the Sequential Choice Model (SCM, Kacelnik et al., 2010; Shapiro et al., 2008) implements the following rationale: 1) when a subject encounters signals that anticipate the possibility of a reward in nature, it normally does so in isolation (different kinds of opportunities (such as prey species) are met sequentially); 2) An experienced animal has expectations for the potential outcome of each kind of opportunity, and for how it compares with the average payoff experienced in the environment as a whole; 3) These expectations translate into noisy distributions of latency to respond to each signal, with signals leading to more favourable outcomes resulting in shorter average latencies; 4) If 2 (or more) signals are encountered simultaneously (a 'choice'), then the said distributions are sampled, and the one yielding the shorter latency elicits a response, resulting in a choice at behavioural level. In this scenario, the observed distribution of latencies to choose results from statistical cross-censorship between two distributions, and hence has some predictable properties. The most specific, and that which separates this model from more conventional models for cognitive processes underlying behavioural choices, is that when responses to a given signal occur in the context of a choice, the observed latencies will be on average the

same or shorter than when the same signal elicits a response in a sequential encounter. An option that is always preferred faces no censorship and its latencies should remain the same. If preferences are partial the preferred and non-preferred options cross-censor each other and therefore have shorter average latencies. The predicted shortening should be more evident for the signal that elicits longer latencies when alone and as a consequence is chosen less often (the non-preferred option), as only the left tail of its distribution will be expressed in choice situations. These predictions contrast with more intuitive and widespread hypotheses, which fail to relate choices to behaviour in sequential encounters and predict, with different degree of explicitness, that choices rely on specific cognitive mechanisms of comparison which deal with the attributes of signals' outcomes in the choice scenario. This leads to the usual expectation of a cognitive evaluation time applicable to models that have been called "Tug-of-War".

Some of the different predictions of these two kinds of model have been compared empirically in birds and the SCM's assumptions linking sequential situations with simultaneous choices have on the whole been supported (Freidin et al., 2009; Kacelnik et al., 2010; Shapiro et al., 2008; Vasconcelos et al., 2015, 2010), but what has hitherto remained most elusive is the demonstration of a latency shortening, affecting primarily the non-preferred alternative. In the present experiment, latencies to accept each option in forced (single-option) trials predicted preferences in choice trials (**Figure 3**), supporting the SCM's assumption that the subjective value of an option is inversely related to the latency to accept it in sequential (forced) trials. Further, **Figure 4** examines the results in terms of the relation between latencies in sequential and choice trials, showing that latency to accept an option in choice trials was equal or shorter than when presenting the option in isolation (forced single-option trials). When an option was preferred in exclusivity, its latency did not differ, and when preferences were partial, both the preferred and non-preferred options had shortened latencies, with the non-preferred option experiencing the stronger shortening between choice and sequential trials. These results constitute the most direct corroboration of the predictions of the SCM reported so far.

In summary: two hypotheses derived from behavioural research in birds have been corroborated using rats. First, rats, like previously tested birds, objectively sacrifice reward rate in exchange for a reduction in uncertainty, but this does not mean that uncertainty avoidance acts as primary reinforcement. The mechanism underlying this tradeoff is not yet

clear, but a strong candidate relates to the learning process by which signals are credited with the value of their outcomes, so that time costs for periods of lost opportunity due to waiting for sure no reward are somehow excluded, or down-weighted, respect to times waiting for at least some probability of reward. Second, the mechanism that leads from valuation to preference in choice situations seems well accounted for by a model that gives priority to sequential encounters without postulating any special mechanism adapted for choice situations. The fact that these two somewhat counterintuitive phenomena are present in such distant taxa suggests that both are adaptive in natural circumstances, rather than resulting from unavoidable cognitive constraints.

Acknowledgements:

AO was supported by a doctoral grant from the Mexican Council for Science and Technology (CONACYT) and a graduate scholarship from Pembroke College. We thank Mark Walton for technical advice and insightful discussions.

586 **References**

- 587 Bennett, D., Bode, S., Brydevall, M., Warren, H., Murawski, C., 2016. Intrinsic Valuation of
588 Information in Decision Making under Uncertainty. PLOS Comput. Biol. 12, e1005020.
589 <https://doi.org/10.1371/journal.pcbi.1005020>
- 590 Berger-Tal, O., Nathan, J., Meron, E., Saltz, D., 2014. The Exploration-Exploitation
591 Dilemma: A Multidisciplinary Framework. PLOS ONE 9, e95693.
592 <https://doi.org/10.1371/journal.pone.0095693>
- 593 Blanchard, T.C., Hayden, B.Y., Bromberg-Martin, E.S., 2015. Orbitofrontal Cortex Uses
594 Distinct Codes for Different Choice Attributes in Decisions Motivated by Curiosity. Neuron
595 85, 602–614. <https://doi.org/10.1016/j.neuron.2014.12.050>
- 596 Chow, J.J., Smith, A.P., Wilson, A.G., Zentall, T.R., Beckmann, J.S., 2017. Suboptimal
597 choice in rats: Incentive salience attribution promotes maladaptive decision-making. Behav.
598 Brain Res. 320, 244–254. <https://doi.org/10.1016/j.bbr.2016.12.013>
- 599 Freidin, E., Aw, J., Kacelnik, A., 2009. Sequential and simultaneous choices: Testing the diet
600 selection and sequential choice models. Behav. Processes, Comparative cognition in context:
601 A tribute to the contributions of Sara J. Shettleworth 80, 218–223.
602 <https://doi.org/10.1016/j.beproc.2008.12.001>
- 603 Golland, P., Liang, F., Mukherjee, S., Panchenko, D., 2005. Permutation Tests for
604 Classification, in: Learning Theory, Lecture Notes in Computer Science. Presented at the
605 International Conference on Computational Learning Theory, Springer, Berlin, Heidelberg,
606 pp. 501–515. https://doi.org/10.1007/11503415_34
- 607 Grafen, A., Hails, R., 2002. Modern statistics for the life sciences. Oxford, UK.: Oxford
608 University Press.
- 609 Heffner, H.E., Heffner, R.S., 1985. Hearing in Two Cricetid Rodents: Wood Rat (*neotoma*
610 *floridana*) and Grasshopper Mouse (*onychomys leucogaster*). J. Comp. Psychol. 99, 275–288.
- 611 Igaya, K., Story, G.W., Kurth-Nelson, Z., Dolan, R.J., Dayan, P., 2016. The modulation of
612 savouring by prediction error and its effects on choice. eLife 5, e13747.
613 <https://doi.org/10.7554/eLife.13747>
- 614 Kacelnik, A., Vasconcelos, M., Monteiro, T., Aw, J., 2010. Darwin’s “tug-of-war” vs.
615 starlings’ “horse-racing”: how adaptations for sequential encounters drive simultaneous
616 choice. Behav. Ecol. Sociobiol. 65, 547–558. <https://doi.org/10.1007/s00265-010-1101-2>
- 617 Krebs, J.R., Kacelnik, A., Taylor, P., 1978. Test of optimal sampling by foraging great tits.
618 Nature 275, 27–31. <https://doi.org/10.1038/275027a0>
- 619 Loomes, G., Sugden, R., 1982. Regret Theory: An Alternative Theory of Rational Choice
620 Under Uncertainty. Econ. J. 92, 805–824. <https://doi.org/10.2307/2232669>
- 621 Martínez, M., Alba, R., Rodríguez, W., Orduña, V., 2017. Incentive salience attribution is not
622 the sole determinant of suboptimal choice in rats: Conditioned inhibition matters. Behav.
623 Processes 142, 99–105. <https://doi.org/10.1016/j.beproc.2017.06.012>

624 McDevitt, M.A., Dunn, R.M., Spetch, M.L., Ludvig, E.A., 2016. When good news leads to
625 bad choices. *J. Exp. Anal. Behav.* 105, 23–40. <https://doi.org/10.1002/jeab.192>

626 Palminteri, S., Lefebvre, G., Kilford, E.J., Blakemore, S.-J., 2017. Confirmation bias in
627 human reinforcement learning: Evidence from counterfactual feedback processing. *PLOS*
628 *Comput. Biol.* 13, e1005684. <https://doi.org/10.1371/journal.pcbi.1005684>

629 Shapiro, M.S., Siller, S., Kacelnik, A., 2008. Simultaneous and Sequential Choice as a
630 Function of Reward Delay and Magnitude: Normative, Descriptive and Process-based
631 Models Tested in the European Starling (*sturnus vulgaris*). *J. Exp. Psychol. Anim. Behav.*
632 *Process.* 34, 75–93. <https://doi.org/10.1037/0097-7403.34.1.75>

633 Smith, A.P., Bailey, A.R., Chow, J.J., Beckmann, J.S., Zentall, T.R., 2016. Suboptimal
634 Choice in Pigeons: Stimulus Value Predicts Choice over Frequencies. *PLOS ONE* 11,
635 e0159336. <https://doi.org/10.1371/journal.pone.0159336>

636 Stagner, J.P., Laude, J.R., Zentall, T.R., 2011. Sub-optimal choice in pigeons does not
637 depend on avoidance of the stimulus associated with the absence of reinforcement. *Learn.*
638 *Motiv.* 42, 282–287. <https://doi.org/10.1016/j.lmot.2011.09.001>

639 Stagner, J.P., Zentall, T.R., 2010. Suboptimal choice behavior by pigeons. *Psychon. Bull.*
640 *Rev.* 17, 412–416. <https://doi.org/10.3758/PBR.17.3.412>

641 Stephens, D.W., Krebs, J.R., 1986. *Foraging Theory*. Princeton University Press.

642 Trujano, R.E., López, P., Rojas-Leguizamón, M., Orduña, V., 2016. Optimal behavior by rats
643 in a choice task is associated to a persistent conditioned inhibition effect. *Behav. Processes*
644 130, 65–70. <https://doi.org/10.1016/j.beproc.2016.07.005>

645 Trujano, R.E., Orduña, V., 2015. Rats are optimal in a choice task in which pigeons are not.
646 *Behav. Processes* 119, 22–27. <https://doi.org/10.1016/j.beproc.2015.07.010>

647 Vasconcelos, M., Monteiro, T., Aw, J., Kacelnik, A., 2010. Choice in multi-alternative
648 environments: A trial-by-trial implementation of the Sequential Choice Model. *Behav.*
649 *Processes, Proceedings of the 32nd Annual Meeting of the Society for the Quantitative*
650 *Analyses of Behavior* 84, 435–439. <https://doi.org/10.1016/j.beproc.2009.11.010>

651 Vasconcelos, M., Monteiro, T., Kacelnik, A., 2015. Irrational choice and the value of
652 information. *Sci. Rep.* 5, 13874. <https://doi.org/10.1038/srep13874>

653 Zentall, T.R., 2016. Resolving the paradox of suboptimal choice. *J. Exp. Psychol. Anim.*
654 *Learn. Cogn.* 42, 1–14. <https://doi.org/10.1037/xan0000085>

655 Zentall, T.R., 2011. Maladaptive “gambling” by Pigeons. *Behav. Processes, Society for the*
656 *Quantitative Analyses of Behavior Choice, Discounting, and the Future Value of Quantitative*
657 *Analyses 33rd Annual Meeting of the Society for the Quantitative Analyses of Behavior* 87,
658 50–56. <https://doi.org/10.1016/j.beproc.2010.12.017>

659 Zentall, T.R., Stagner, J., 2011. Maladaptive choice behaviour by pigeons: an animal
660 analogue and possible mechanism for gambling (sub-optimal human decision-making
661 behaviour). *Proc. R. Soc. Lond. B Biol. Sci.* 278, 1203–1208.
662 <https://doi.org/10.1098/rspb.2010.1607>

663 **Supplementary Methods**

664

trial number	1	2	3	4	5	6	7	8	9	
trial type	forced info	forced non-info	choice	forced info	forced non-info	choice	forced info	forced non-info	choice	
Latency (s)	3	2		2	3		2	4		
predicted	non-info			info			info			total= 2/3 info
observed			info			info			non-info	total= 2/3 info

665
666
667
668

Table 1. Schematic example of the prediction of preferences for a session

Supplementary Results

Figure S1 shows the head poking behavior during the different treatments of the main experiment. A two-way repeated-measures ANOVA with cue and treatment as within-subject factors revealed a significant effect of cue: ($F_{3,45}=80.91, P<0.0001$), treatment ($F_{2,30}=3.97, P<0.0001$) and their interaction ($F_{6,90}=2.78, P<0.02$). Post-hoc comparison between cues revealed a significant difference between the *i-minus* cue and all other cues, below is the table with the p-values (Bonferroni-corrected) for all pairwise comparisons.

		p-value
'i-minus'	'i-plus'	<0.0001
'i-minus'	'n1'	<0.0001
'i-minus'	'n2'	<0.0001
'i-plus'	'n1'	0.0875
'i-plus'	'n2'	0.2415
'n1'	'n2'	0.9234

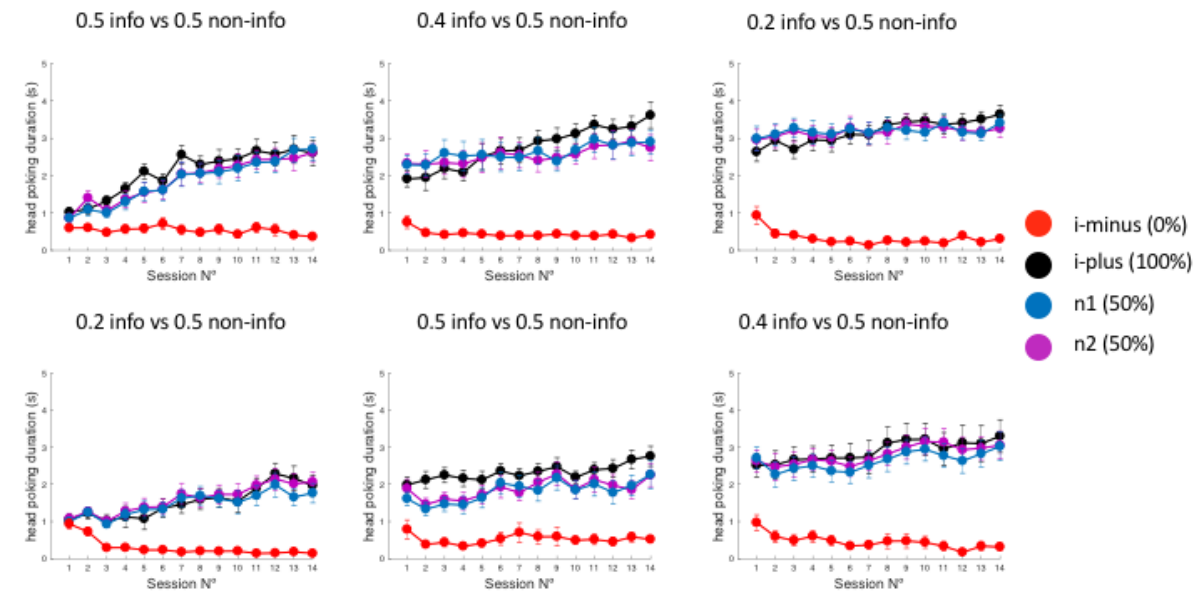
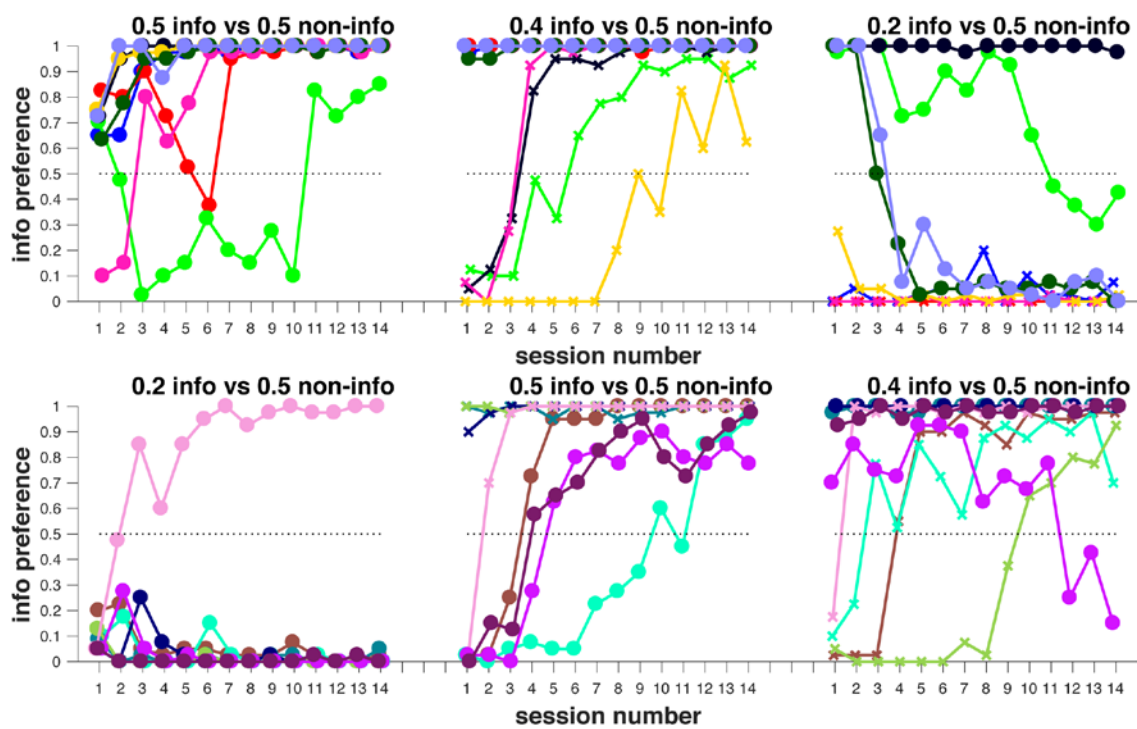


Figure S1. Mean \pm S.E.M. duration of head poking into the food magazine during cue presentation in each of the conditions of the main experimental phase

682



683

684 **Figure S2.**

685 Preference for the informative option in each of the treatments of the main experiment. Each color represents a
686 different subject. Between phases the spatial position of the options was reversed for half of the subjects, points
687 from subjects that experienced a reversal are marked with x's.

688

689