AMERICAN PSYCHOLOGICAL ASSOCIATION

Decision

ISSN: 2325-9965

2023, Vol. 10, No. 3, 247–267
https://doi.org/10.1037/dec0000179

Comparing Anticipation and Uncertainty-Penalty Accounts of Noninstrumental Information Seeking

Shi Xian Liew, Jake R. Embrey, Danielle J. Navarro, and Ben R. Newell
School of Psychology, University of New South Wales

Proposed psychological mechanisms generating noninstrumental information seeking in humans can be broadly categorized into two competing accounts: the maximization of anticipating rewards versus an aversion to uncertainty. We compare three separate formalizations of these theories on their ability to track the dependency of information-seeking behavior on increasing levels of cue-outcome delay as well as their sensitivity to outcome valence. Across three experiments using a variety of different stimuli, we observe a flat to monotonically increasing pattern of delay dependency and minimal evidence of sensitivity to outcome valence—patterns which are better predicted, qualitatively and quantitatively, by an uncertainty aversion information model.

Keywords: information seeking, temporal discounting, reinforcement learning, anticipation, computational modeling

Supplemental materials: https://doi.org/10.1037/dec0000179.supp

The human tendency to seek out information is well documented (Gottlieb et al., 2014; Kidd & Hayden, 2015; Navarro et al., 2016). This is perhaps unsurprising under situations where information can be used to guide future decisions. However, this tendency to want-to-know also manifests when information is apparently noninstrumental, that is, when it is irrelevant for future actions (Charpentier et al., 2018; Iigaya et al., 2016; Sharot & Sunstein, 2020; Vasconcelos et al., 2015; Zhu et al., 2017). Such behavior is observed even when the information incurs a cost (Bennett et al., 2016, 2021; Cabrero et al., 2019; Eliaz & Schotter, 2010; Pierson & Goodman, 2014). To illustrate the basic idea, imagine awaiting the draw of a raffle ticket from a lottery: the ticket is drawn, you can see it's not the color of the ticket you bought (so you have not won) but the temptation to know the winning number remains strong.

Several theories explaining noninstrumental information-seeking behavior adopt a reinforcement learning approach (e.g., Beierholm & Dayan,

This article was published Online First April 4, 2022.

Shi Xian Liew https://orcid.org/0000-0003-0432-1795
Danielle J. Navarro https://orcid.org/0000-0001-7648-6578
Ben R. Newell https://orcid.org/0000-0003-1898-205X

The data sets analyzed during the present study are available in the GitHub repository, https://github.com/shixianliew/infoseekDelayDist.

The model scripts used during the present study are available in the GitHub repository, https://github.com/shixianliew/infoseekDelayDist.

The authors declare no competing interests.

All experiments followed all ethical guidelines and were reviewed by University of New South Wales Human Research Advisory Panel C: Behavioural Sciences with approval number 3205. Informed consent was obtained from all participants prior to participation.

Shi Xian Liew played lead role in writing of original draft and writing of review and editing, supporting role in methodology and visualization, and equal role in conceptualization, formal analysis, and investigation. Jake R. Embrey played lead role in methodology and supporting role in formal analysis, investigation, and writing of review and editing. Danielle J. Navarro played lead role in visualization, supporting role in investigation, project administration, supervision and writing of original draft, and equal role in funding acquisition and writing of review and editing. Ben R. Newell played lead role in project administration and supervision and equal role in conceptualization, funding acquisition, investigation, writing of original draft and writing of review and editing.

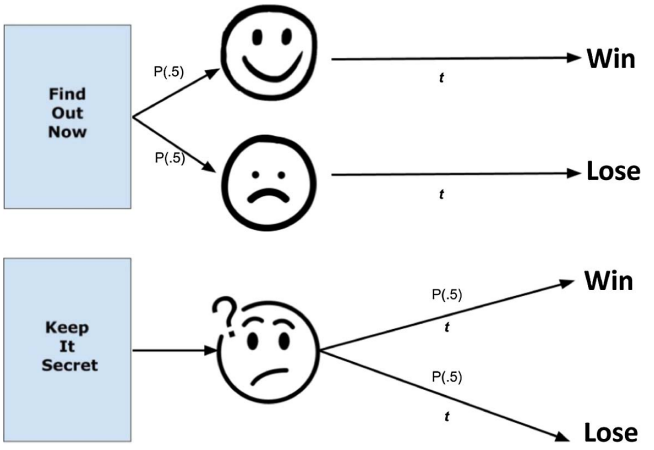
Correspondence concerning this article should be addressed to Shi Xian Liew, School of Psychology, University of New South Wales, Sydney, NSW 2052, Australia. Email: shixianliew@gmail.com

2010; Bromberg-Martin & Hikosaka, 2011). This class of models offers two important predictions: first, they suggest that people can avoid noninstrumental information about potentially aversive future outcomes (Iigaya et al., 2016; Kobayashi et al., 2019; Story et al., 2013; Zhu et al., 2017), and second, they typically involve a temporal discounting mechanism that assumes preference for information decreases as the delay between cue and outcome increases (Loewenstein, 1987).

The first prediction on the avoidance of information on aversive outcomes can be understood as a strong interpretation of valence sensitivity where the polarity of outcome valence (i.e., whether it is negative or positive) is directly mapped to the direction of information preference (i.e., to seek or to avoid). This can be understood within the general framework of information seeking described by Sharot and Sunstein (2020), who suggest that the knowledge of aversive outcomes can induce negative affect resulting in the avoidance of such information. Discussion involving general (usually instrumental) information-avoidant behavior is not new (e.g., see Golman et al., 2017; Hertwig & Engel, 2016), however the reinforcement learning models that we implement and compare here are among the few to explicitly consider the avoidance of *noninstrumental* information.

Empirical evidence demonstrating explicit noninstrumental information-avoidance behavior is weak, and to our knowledge only evidence for weak valence sensitivity exists (i.e., where information preference strength may be influenced by valence but are not necessarily directly mapped onto valence polarity). Specifically, Zhu et al. (2017) investigated this effect using an noninstrumental information-seeking paradigm known as the *secrets* task. In this task, observers are presented with a risky event and are given the choice to receive advance noninstrumental information in the form of completely predictive cues (find out now [FON]) or to receive an ambiguous cue (keep it secret [KIS]). After the cue is revealed, the observer is made to wait for a period of time before the outcome is received (e.g., see Figure 1 for general structure of task). When comparing outcomes involving positively valenced stimuli (erotic images) with negatively valenced stimuli (aversive gory images), Zhu et al. (2017) observed a significant decrease in information-seeking behavior with the negatively

Figure 1
General Experimental Design of The Secrets Task



Note. *t* indicates the delay between presentation of the cue and the reward. Specific rewards varied across experiments. Experiment 1 (win block): participants could win 100 points or otherwise lose nothing. Experiment 1 (loss block): participants could win nothing or otherwise lose 100 points. Experiment 2 (chocolate): participants could win an M&M chocolate or lose nothing. Experiment 2 (sound): participants could win nothing or otherwise receive an aversive noise. Experiment 3: participants could win an M&M chocolate or otherwise receive an aversive noise. See the online article for the color version of this figure.

valenced outcomes, demonstrating weak valence sensitivity. However, since the average response with aversive outcomes was not different from chance, there was no clear observation of information-avoidant behavior (i.e., strong valence sensitivity).

In a related study, using a modified secrets task in which probabilities of outcomes were varied but delays were held constant, the first experiment by Charpentier et al. (2018) found that when the probability of winning increased, so too did information-seeking behavior, with the opposite pattern observed for losses. This result thus provides further evidence indicating an effect of outcome valence on the strength of information seeking but again there was no evidence for information avoidance. Even when faced with a high probability of a loss, average choice proportions still showed a majority for information seeking.

The strongest evidence for information avoidance may be found in the second experiment of Charpentier et al. (2018). They conducted a stock market task where participants had to indicate their willingness to pay for information or to remain ignorant of the value of their investment portfolio under changing market conditions. They found a significant decrease in information preference (measured as contrasts in willingness to pay) when presented with decreasing markets compared with increasing markets. However, it is unstated whether people were significantly willing to pay more for ignorance (on average). Again, while we see evidence of weak valence sensitivity, there is no explicit sign of strong valence sensitivity in the form of information avoidance.

The second prediction of the general reinforcement learning approach involves the mechanism of temporal discounting and is founded on the idea that rewards themselves lose value the longer we have to wait to obtain them, consequently resulting in a decrease in the value of information on those rewards (Loewenstein, 1987). Iigaya et al. (2020, 2016) explicitly investigated this effect by systematically manipulating the cue-outcome delay within the secrets task. They hypothesized that between short to medium delays, information preferences should increase as anticipation for the (positive) outcome increases. As delay extends beyond a moderate amount, they predict that the temporal discounting of the outcome value eventually overpowers the increase in anticipation and results in the attenuation of information preferences toward a level of indifference. After testing a range of delay values between 1 and 40 s, they observed a monotonic increase in information preference without any clear decrease. Although no evidence of temporal discounting was observed, they speculated that such behavior may be observed at extended cue-outcome delay lengths beyond 40 s.

In contrast to the reinforcement learning class of models are theories that focus on temporal resolution of uncertainty (e.g., Bennett et al., 2016; Epstein & Zin, 1989; Kreps & Porteus, 1979). Specifically, they describe information-seeking behavior as the preference for resolving uncertainty at an earlier (rather than later) point in time. These models are agnostic to the valence of the outcome—they predict a preference for uncertainty reduction irrespective of whether a future outcome is potentially positive or negative. Additionally, they do not typically incorporate any temporal discounting mechanism.

One impediment to comparing many of the existing studies is that the specifics of the tasks used vary, thus making it more difficult to draw general conclusions regarding the impact of different variables. For example, Zhu et al. (2017) varied valence in the secrets task but did not manipulate delay; Iigaya et al. (2016, 2020) did the opposite, and Charpentier et al. (2018) used a subtly (but importantly)[1] different task and used monetary outcomes in comparison to the primary reinforcers used by Iigaya et al. (2016, 2020) and Zhu et al. (2017). Primary reinforcers refer to rewards that are available for immediate consumption or indulgence and have been argued to be important in eliciting delay-dependent choice effects (Crockett et al., 2013; Iigaya et al., 2016) and contrasts with secondary reinforcers, which are rewards provided as a proxy for primary reinforcers (e.g., money, a secondary reinforcer, being used as a medium to buy food, a primary reinforcer). However, the impact of varying reinforcer type in the same task is yet to be examined, thus leaving open questions about the importance of the reinforcer for eliciting noninstrumental information seeking.

Here, we take a comprehensive approach by manipulating delays—including delays that are

[1] In Charpentier et al. (2018) the presentation of cues themselves were probabilistic, and gamble outcomes of the noninformative cue were also hidden.

twice as long as those in previous work—the valence of outcomes, and the type of outcomes. Specifically, we use monetary outcomes (gains and losses) in Experiment 1, and then primary reinforcers of chocolate (gain) and aversive noise (loss) in Experiments 2 and 3. We chose these outcome types over images because although the images used in previous work can be considered primary reinforcers, it is difficult to measure the extent to which participants actively engaged with the stimuli at the time of image presentation. In contrast, our primary reinforcers directly stimulated sensory modalities outside the visual esthetics provided by images.

Beyond this empirical contribution, we also compare the performance of recent models of anticipation-based reinforcement learning models against a competing model that focuses on temporal resolution of uncertainty known as the uncertainty penalty (UP) model (Bennett et al., 2016). We chose these models for their qualitatively different predictions within our paradigm.[2] Our two reinforcement learning models are the reward prediction error-anticipation (RPE-A) model (Iigaya et al., 2016) and the anticipated prediction error (APE) model (Zhu et al., 2017). Both models assume the two key predictions outlined earlier, that is, predictions of information avoidance as well as temporal discounting of rewards. Both models also incorporate some interpretation of anticipated values. However, they differ in terms of how these anticipated values affect behavior. At a broad level, RPE-A assumes that people accumulate anticipation of positive future events while dreading future negative events. As the delay between cue and outcome increases, the anticipation (or dread) also increases, resulting in greater information-seeking behavior. However, due to the temporal discounting mechanism, at extended delay lengths the discount in reward is expected to overpower increases in anticipation, ultimately showing a nonmonotonic influence of delay on information preference (see Figure 2a).

In contrast, APE does not assume a delay-dependent savoring mechanism, consequently it does not predict an increase in information preference over increases in delay. Rather, APE assumes that people are sensitive to the prediction errors (i.e., differences between predicted and expected reward) themselves, with a preference for larger (positive) prediction errors (typically from winning outcomes) and an aversion to negative prediction errors (typically from losing outcomes). As delay lengths increase, the prediction errors themselves decrease, consequently revealing monotonically decreasing information-seeking behavior (see Figure 2b). Additionally, both models offer different explanations of information avoidance. While RPE-A directly maps the direction of information preferences to outcome valence, APE accounts for information avoidance as the result of overweighting losing outcomes relative to winning outcomes.

Founded on a more general framework of uncertainty resolution (Epstein & Zin, 1989; Kreps & Porteus, 1978), UP describes information-seeking behavior as the preference for resolving uncertainty at an earlier (rather than later) point in time. This model is agnostic to the valence of the outcome—it predicts a preference for uncertainty reduction irrespective of whether a future outcome is potentially positive or negative. Additionally, it does not typically incorporate any temporal discounting mechanism. Qualitatively, this model predicts a monotonically increasing information preference as the cue-outcome delay increases (Figure 2c).

Ultimately, these models allow for varying levels of cue-outcome delay dependency and directions of information preferences using different explanatory mechanisms. Our overall goal is to see which of these models offers the best account of information-seeking behavior in our experiments. The paper proceeds as follows. Experiment 1, Experiment 2, and Experiment 3 sections report the empirical results from Experiments 1–3 respectively; Models of Information Seeking section presents the model comparison and Model Fitting section discusses the implications of our findings.

Experiment 1

In their original study, Iigaya et al. (2016) did not find strong temporal discounting effects on information-seeking behavior at their longest

[2] We note that other information-seeking models exist (e.g., Beierholm & Dayan, 2010), however these models appear to mimic our current set of models within the context of varying delay and valence. Since we have selectively opted for models that perform qualitatively different from each other, we do not include models that qualitatively show the same behavior. We address this further in the discussion and Supplemental Materials.

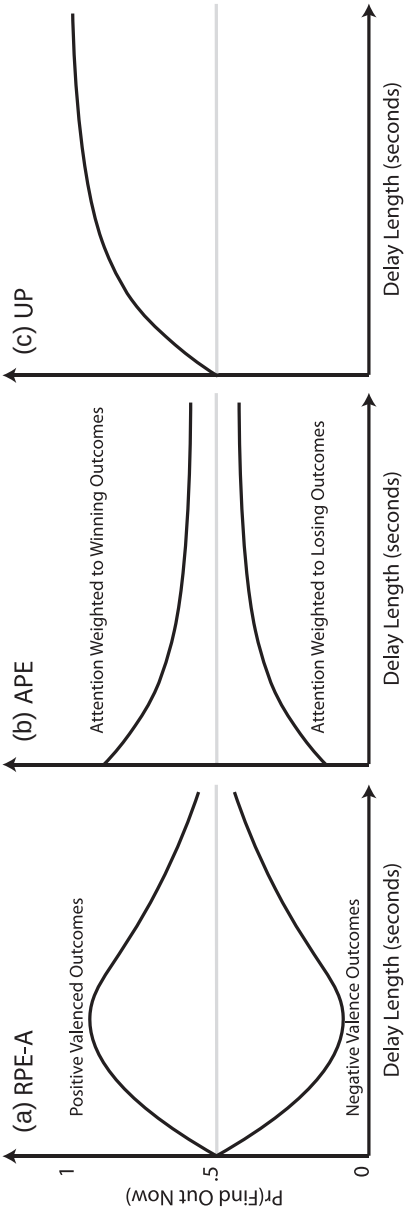


Figure 2
Qualitative Predictions of RPE-A, APE, and UP

Note. RPE-A = reward prediction error-anticipation; APE = anticipated prediction error; UP = uncertainty penalty.

delay duration (40 s), speculating that such effects may only manifest at even longer delay durations. We designed our first experiment with this objective, manipulating delay lengths up to twice as long as Iigaya et al. (2016; i.e., 80 s in a secrets task using secondary reinforcers where participants could separately win or lose points.

Method

Participants

Following the sample sizes of previous studies (e.g., Charpentier et al., 2018), we recruited 40 UNSW psychology undergraduate students (M_{age} = 20.13 years, 25 females, 15 males) in exchange for course credit, as well as a cash amount depending on the points awarded during the task (M = 3.50 AUD).

Materials

We implemented the experiment in jsPsych (De Leeuw, 2015) within the Google Chrome browser on a desktop computer.

Design and Procedure

Each participant completed two blocks of trials in a randomized order. All participants started the experiment with 3,000 points to prevent the possibility of obtaining a negative balance. In the "win block," the positive outcome of each trial was receiving a random sample from a uniform distribution between 90 and 110 points, and the negative outcome was receiving 0 points. In the "loss block," the positive outcome was receiving 0 points and the negative outcome was losing a random sample from a uniform distribution between 90 and 110 points. Participants were informed that the probability of obtaining a positive outcome or a negative outcome were equal and both were independent from the participants' choice of option.

On each trial, participants were presented with the cue-outcome delay duration as well as the option to either FON or KIS. If participants chose the FON option, they immediately received either a smiley face (cueing a positive outcome) or a sad face (cueing a negative outcome) with equal probability. If participants chose the KIS option, they were presented with a confused face (a noninformative cue) instead. Cues remained on

screen for as long as the delay duration required, followed by delivery of the outcome. There were seven delay lengths of 1, 2, 5, 10, 20, 40, and 80 s randomly interleaved throughout the session for each participant, with 10 trials for each delay length from 1 to 20 s, and five trials for delay lengths of 40 and 80 s, resulting in 60 trials in total per participant. The generic structure of the experiment is presented in Figure 1. After completion, participants were debriefed on the experiment's aims and reimbursed according to the amount of money they earned: 1,000 points = $1 AUD.

Results and Discussion

The mean probability of choosing FON across increasing delay durations is presented in Figure 3a for the win blocks, and Figure 3b for the loss blocks. Visual inspection of the results suggests an overall preference for the FON option that is constant across delay and unaffected by valence. Averaged across all trials, this preference was significantly greater than chance for the win ($M = 0.63$, $t(39) = 3.67$, 95% CI[0.56, 0.69], $p < .001$, $d = 0.59$) and loss blocks ($M = 0.63$, $t(39) = 3.72$, 95% CI[0.56, 0.69], $p < .001$, $d = 0.60$).

We used generalized linear mixed models (GLMM) to investigate the effect of delay and valence on information-seeking behavior. We compared the baseline model using only random intercepts for individual participants to models with random slopes (as a function of delay and valence) as well as delay length and valence as fixed effects. We found no evidence for an effect of delay or valence (Table 1).[3]

Contrary to Zhu et al. (2017) we found no effect of outcome valence on information preference. In addition, unlike Iigaya et al. (2016) we found no delay-dependent effect on information-seeking behavior. These patterns suggest that primary reinforcers are important for eliciting delay-dependent changes in the preference for noninstrumental information.

Experiment 2

In light of the absence of delay-dependent effects with the secondary reinforcers of Experiment 1, Experiment 2 used primary reinforcers; namely M&M chocolates as the positively valenced reward, and a high pitched aversive tone as the negative stimulus (microphone feedback, ranked 2nd-most aversive stimulus out of 34 other stimuli; Cox, 2008). These two conditions were designed to be analogous to the win and loss block conditions from Experiment 1, but in this experiment were run on different groups of participants (i.e., valence of reward was between-subjects rather than within).

Methods

Participants

The sound condition in Experiment 2 had 51 undergraduate psychology student participants (M_{age} = 19.49, 38 females, 13 males) who were granted course credit for their participation. The chocolate condition in Experiment 2 had 49 participants ($M_{age} = 21.82$, 20 females, 29 males). Participants were a mix of undergraduate students who received course credit for participation and paid participants who received $15 AUD. Participants were required to fast for 2 hr prior to the experiment and self-reported that they liked M&M chocolates. We also ensured participants had not previously completed Experiment 1 or other similar studies.

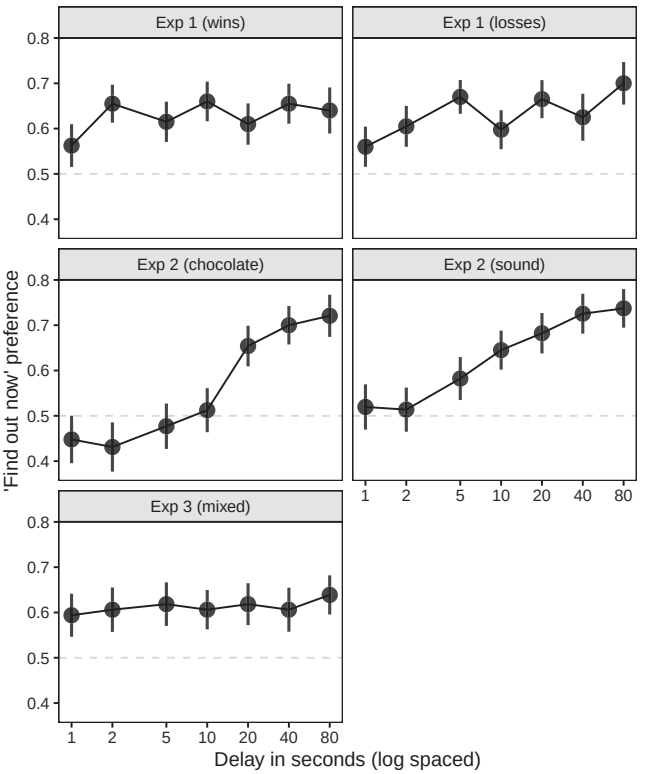
Design and Procedure

The chocolate condition was similar to the win block in Experiment 1. Participants who received a positive outcome were presented with a single M&M chocolate delivered via an automated dispenser, otherwise a negative outcome delivered nothing. In the sound condition (comparable to the loss block in Experiment 1), a negative outcome presented an aversive microphone feedback noise that played for 10 s via headphones. A positive outcome in the sound condition did not result in any specific event, however participants had to wait for 10 s before they were able to proceed to the next trial. As with Experiment 1, the probability of positive and negative outcomes was equal.

In order to increase the robustness of estimates, we raised the number of trials at the 40 and 80 s delays to 10 for each delay length resulting in 70 trials in total across the experiment. The design was otherwise identical to Experiment 1.

[3] While it falls outside the immediate scope of this study, we also analyzed the correlation between FON choices and trial number. Details of this analysis can be found in the Supplemental Materials.

Figure 3
Mean FON Proportions Across Increasing Delay (Log-Scaled) For Each Experiment Condition



Note. Error bars indicate 1 standard error of the mean. FON = find out now.

Results and Discussion

Mean information preferences across increasing levels of delay are presented in Figure 3c for the chocolate condition and Figure 3d the sound condition. On average, across all levels of delay, the chocolate condition did not yield significant information-seeking behavior ($M = 0.56$, $t(47) = 1.89$, 95% CI[.50, .63], $p = .06$). The sound condition indicated significant preference for advance information ($M = 0.63$, $t(50) = 4.07$, 95% CI[.57, .69], $p < .001$, $d = 0.58$).

We analyzed the average preference for the FON option using a GLMM analysis analogous to that of Experiment 1 (see Table 2). In the chocolate condition, we found that the addition of random slopes significantly improved the fit of the model, suggesting information preference across time varied significantly between participants. The addition of delay also significantly

Table 1
GLMM Statistics for Experiment 1

Linear model	Parameters	BIC	χ^2	*df*	*P*
Baseline	2	5646.9	—	—	—
Random slopes	4	5539.3	149.9838	5	<.001
Random slopes with delay and valence	6	5553.9	2.24	1	.13

Note. GLMM = generalized linear mixed models; BIC = Bayesian Information Criterion.

Table 2
GLMM Statistics for Experiment 2

Condition	Linear model	Parameters	BIC	χ^2	*df*	*p*
Chocolate	Baseline	2	3959.5	—	—	—
	Random slopes	4	3269.1	694.43	2	<.001
	Random slopes with delay	5	3261.4	9.67	1	.002
Sound	Baseline	2	4056.3	—	—	—
	Random slopes	4	3609.2	463.53	2	<.001
	Random slopes with delay	5	3607.7	9.61	1	.002

Note. GLMM = generalized linear mixed models; BIC = Bayesian Information Criterion.

improved the model's fit, indicating that delay length had an effect on information preference. On average we observed that the preference for the FON option increased from short to long delays. Specifically, at 1 and 2 s delays, people were slightly information avoidant (preferences of .45 and .43, respectively), with preferences for FON increasing to .70 and .72 at the 40 and 80 s delays, respectively.

Similar to the chocolate condition, the addition of random slopes and delay improved the model's fit (both $p < .01$) in the sound condition. The pattern of results was also similar to the chocolate condition: at 1 and 2 s delays, mean preference was .52 and .51, respectively, with these preferences increasing to .73 and .74 at 40 and 80 s, respectively.

Unlike the secondary reinforcers used in Experiment 1, both tasks using primary reinforcers found that delay has an effect on information preference. These results are consistent with those of (Iigaya et al., 2016; Sharot & Sunstein, 2020). Despite doubling the length of the longest delay (40 s–80 s) we did not, however, see the nonmonotonic pattern—decreasing preference for information as delay increases—speculated by Iigaya et al. (2016). In addition, we did not observe any clear information-avoidant behavior in any condition; a finding which is at odds with the predicted pattern of avoidance, or at least lower preference for non-instrumental information about negative outcomes (Charpentier et al., 2018; Sharot & Sunstein, 2020; Zhu et al., 2017).

Experiment 3

Experiment 2 was successful in eliciting delay-dependent information behavior, however the absence of information avoidance or a reduction in information seeking when facing negative primary reinforcers is puzzling given the predictions of some accounts (Iigaya et al., 2016; Sharot & Sunstein, 2020). In Experiment 3 we examined the possibility that information seeking depends not only on the absolute valence of the outcome, but also its valence relative to other potential outcomes. Specifically, it is possible that the negative outcomes in Experiments 1 and 2 were not *relatively negative* enough compared to the other neutral outcomes (i.e., winning 0 points, or not receiving the aversive sound). By replacing the neutral outcomes with a positively valenced reward, we increase the relative negativity of the negative outcomes as well as the relative positivity of positive outcomes. If information-avoidant behavior is contingent on both absolute and relative valences of outcomes, we predicted that this manipulation (i.e., using outcome stimuli that are different in relative and absolute valence) would lead to a pattern of information avoidance. To investigate this possibility, in Experiment 3 we removed the neutral outcome and used primary reinforcers for all possible outcomes. Specifically, M&M chocolates were rewarded as the positive outcome and the aversive microphone sound was delivered as the negative outcome. This mixed paradigm is similar to the mixed condition in Zhu et al. (2017) which used erotic and aversive images for the positive and negative rewards, respectively (but held delay constant at 20 s for all trials).

Method

Participants

Fifty-one undergraduate students from the UNSW psychology cohort (M_{age} = 19.27 years, 30 females, 21 males) were recruited via the SONA platform in exchange for course credit.

Design and Procedure

The design and procedure were similar to Experiment 2, with a single block consisting of 70 trials (10 per delay) where the positive outcome was an M&M chocolate and the negative outcome was the aversive microphone feedback noise.

Results and Discussion

Overall, a significant preference for advance information was observed ($M = 0.61$, $t(48) = 3.27$, 95% CI[.55, .68], $p < .002$, $d = 0.47$), but somewhat surprisingly the average choice proportion of FON did not vary considerably across the different delay lengths, indicating no clear effect of delay-dependence (Figure 3e).

Applying the same GLMM analysis to this data indicated that preference for information varied significantly between participants, where the addition of random slopes significantly improved the fit of a baseline model (Table 3). However, the addition of delay did not significantly improve the model's fit.

Despite using primary reinforcers, and offering a choice between a positive and a negative outcome on each trial, we saw no evidence of information avoidance. If anything, overall preference for information appeared to be highest (on average) of all three experiments. Contrary to the suggestions of previous studies (e.g., Iigaya et al., 2016; Sharot & Sunstein, 2020) these findings suggest that the avoidance of information is not dependent on the relative valence of negative outcomes compared to positive outcomes.

In addition, the results of Experiment 3 yielded no evidence of delay-dependent information preference. Not only was there no attenuation in information preference at longer delays as predicted by earlier literature Iigaya et al. (2016), there was no observation of any increase in shorter delays. While a general effect of delay dependence was not the central prediction in this experiment, its absence here is a point worth raising for future investigation, which we further address in the General Discussion.

Models of Information Seeking

As prefaced in Introduction section, we consider three formal models that provide different accounts of noninstrumental information seeking. Two of the models give an explanation based on anticipatory utility, in which the decision-maker is presumed to derive additional value from the anticipation of the arrival of the rewarding outcome. According to such an account, people seek (or avoid) noninstrumental information about a future outcome because it allows us to savor (or dread) the event to come. The third model relies on an entirely different principle, namely that uncertainty is inherently aversive, and people seek information in order to remove uncertainty about an outcome, quite irrespective of the utility of the outcome itself. In this section we describe all three models. Notation schemes vary across papers, and to minimize confusion we apply a consistent notation scheme to describe events in the experiment, as outlined in Figure 4. If an agent stores some internal representation of the value of state s it is denoted as $V(s)$. The reward experienced when the trial progresses from some state s to a later state s' is denoted as $R(s, s')$. In cases where this reward does not depend on the initial state s, this is denoted as $R(\cdot, s')$. Unless stated otherwise, the rewards and values are subject to temporal discounting and the notation refers to the temporally discounted reward/value at the time the agent makes the choice. For reinforcement learning models, prediction errors—between expected and experienced rewards associated with a state change—are denoted as $E(s, s')$. Additionally, when a value, reward or prediction error includes an anticipatory component, the relevant functions are denoted using $\tilde{V}$, $\tilde{R}$, or $\tilde{E}$. To maintain consistency with previous work, the free parameters in each model are expressed in the notation from the original papers.

Reward Prediction Error With Anticipation

The first model we consider is the reinforcement learning model introduced by Iigaya et al. (2016), which they referred to as a *reward prediction error model with anticipation* (RPE-A). The central assumption of this model is that observers savor the subjective utility from the anticipation of future events (Loewenstein, 1987; Story et al., 2013), and allows the savoring itself to serve as a reinforcer.

The formal structure of the model is as follows. On any given trial the agent assigns a predicted reward value $V(s)$ to the two choice states s_f and

Table 3
GLMM Statistics for Experiment 3

Linear model	Parameters	BIC	χ^2	*df*	*p*
Baseline	2	3859.8	—	—	—
Random slopes	4	3575.6	300.44	2	<.001
Random slopes with delay	5	3583.5	0.29	1	.59

Note. GLMM = generalized linear mixed models; BIC = Bayesian Information Criterion.

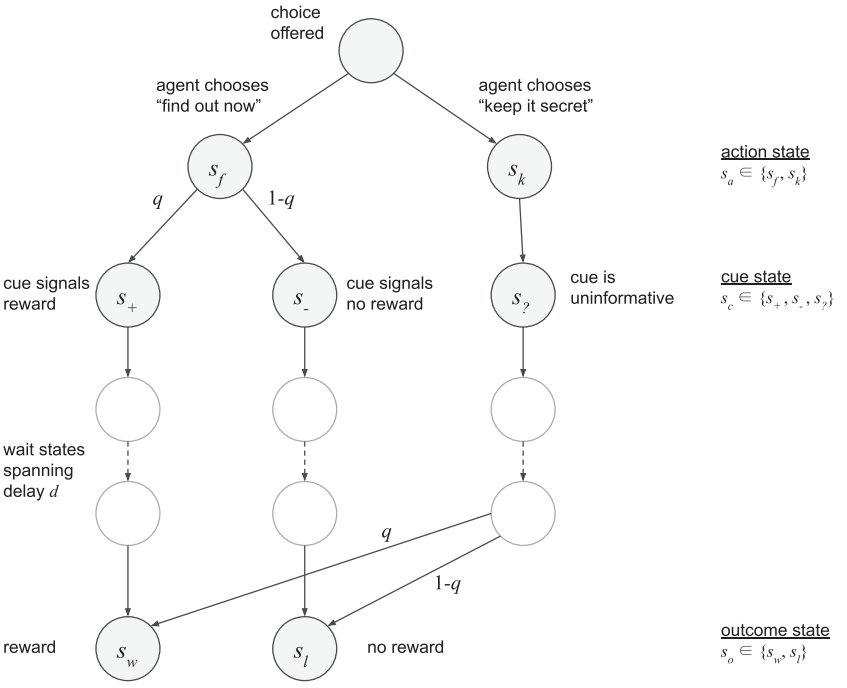
s_k, and makes decisions in accordance with a softmax rule

$$P(s_f) = \frac{1}{1 + e^{\beta(V(s_k) - V(s_f))}}, \quad (1)$$

where β is a free parameter in the model that governs the agent's willingness to choose a lower value option. After making the choice s_a that eventually leads to outcome s_o and yields an accumulated, temporally discounted reward $R(s_a, s_o)$ that is experienced by the agent, the value assigned to that action state s_a is updated according to the following rule:

$$V(s_a) \leftarrow V(s_a) + \alpha E(s_a, s_o), \quad (2)$$

Figure 4
Notation Used to Describe the Events Within a Single Trial in the Secrets Task



Note. When the agent makes a decision the trial enters an action state s_a, which is either s_f if the agent chooses to find out now, or s_k if they decide to keep it secret. This is then followed by a cue state s_c, in which the agent is presented a cue that may signal a reward (state s_+), signal no reward (state s_-), or be uninformative (state $s_?$). This is then followed by wait states that span some delay interval d, then finally an outcome state s_o is reached. Irrespective of which decision the agent made, this outcome state will be a win state s_w with probability q, or a loss state s_l with probability $1 - q$.

where the reward prediction error is the difference

$$E(s_a, s_o) = R(s_a, s_o) - V(s_a). \quad (3)$$

This expression describes a standard reward prediction error learning rule with α denoting a learning rate parameter: the agent had assigned value $V(s_a)$ to this state, and adjusts this slightly to more closely match the experienced reward $R(s_a, s_o)$.

Where the RPE-A model departs from more traditional reinforcement learning accounts, the total reward received has two distinct components: there is a primary reward $R(\cdot, s_o)$ corresponding to the temporally discounted reward from the outcome itself, but there is also an anticipatory reward $\tilde{R}(s_a, s_o)$ experienced during the wait states that last for the delay length d, because the agent savors (or dreads) the arrival of the outcome, and this also is subject to temporal discounting. Formally, the present value (at the time of choice) of total reward to be experienced by the agent is given by the weighted sum

$$R(s_a, s_o) = R(\cdot\,, s_o) + \eta\tilde{R}(s_a, s_o), \quad (4)$$

where we use $\tilde{R}$ to indicate that the experienced reward is anticipatory and

$$\eta = \eta_0 + c|E(s_a, s_c)|. \quad (5)$$

In this expression the "baseline anticipation" parameter η_0 and the "error boosting" parameter c are both free parameters of the model, and $E(s_a, s_c)$ denotes the reward prediction error experienced when the cue state is revealed. For example, if the agent knows that a keep it secret choice s_k is always followed by the ambiguous $s_?$ then there will be no prediction error when the cue is revealed, because $V(s_k) = V(s_?)$. However, when the agent decides to find out now some prediction error will be encountered because $V(s_f) < V(s_+)$ and $V(s_f) > V(s\_)$. As such the magnitude of the prediction error will be greater when the agent chooses to find out now, $|E(s_f, s_c)| > |E(s_k, s_c)|$ simply by virtue of the fact that information is revealed at the cue stage when a find out now decision is made. This "boosting" mechanism is central to the RPE-A model. It asserts that the savoring is more intense when it follows a surprising cue than when it follows a predictable cue.

To complete the model, we require expressions for the reward prediction error $E(s_a, s_c)$ that follows the cue, as well as primary reward value $R(\cdot, s_o)$ and the anticipatory reward value $\tilde{R}(s_a, s_o)$. As noted earlier these two quantities need to be assessed at the time of the choice, so temporal discounting mechanism is applied. Thus, if $R_\dagger(\cdot, s_o)$ denotes the primary reward value of the outcome state s_o at the time it is delivered, the relevant discounted reward after a delay of length d is given by

$$R(\cdot\,, s_o) = R_\dagger(\cdot\,, s_o)e^{-\gamma d}, \quad (6)$$

and as noted by Iigaya et al. (2016), the present value at time of choice for the accumulated anticipation can be calculated analytically by integrating the instantaneous time-discounted anticipatory reward across the duration of the delay d, thereby marginalizing the wait states. If the momentary strength of anticipatory savoring rises over time with rate ν, the integral yields

$$\tilde{R}(s_a, s_o) = (\eta_0 + c|E(s_a, s_c)|)\frac{R_\dagger(\cdot\,, s_o)}{\nu - \gamma} \times (e^{-\gamma d} - e^{-\nu d}), \quad (7)$$

(see original paper for details). Finally, regarding the cue prediction error itself Iigaya et al. (2016) derived an analytic expression for $E(s_a, s_c)$ which holds whenever the agent has learned the true choice-cue association (i.e., the probability q of seeing a particular cue following the choice),

$$E(s_a, s_c) = \frac{(1-q)(\eta_0\tilde{R}(s_a, s_o) + R(\cdot\,, s_o)}{1 - (1-q)c\tilde{R}(s_a, s_o)}. \quad (8)$$

Note that the expressions for the anticipatory reward $\tilde{R}(s_a, s_o)$ and the cue prediction error $E(s_a, s_c)$ depend on one another, but as noted by Iigaya et al. (2016) there are generally stable solutions to this system of equations: our implementation used fixed point iteration to find the stable values.

The predictions of the RPE-A model is a consequence of three distinct mechanisms: the amount of momentary anticipation rises over time, the value of future rewards (both primary and anticipatory) is discounted as a function of delay, and that the amount of anticipation experience is larger when it follows a surprising information signal (the cue). The qualitative effect of

these mechanisms is illustrated in the left panel of Figure 2.

Anticipated Prediction Error

The second model we consider is a reinforcement learning model that assumes anticipation is the primary driver of information-seeking behavior but—unlike the previous model—it does not incorporate any notion of savoring a future reward. Rather, the theoretical claim is that the agent makes decisions not merely on the perceived value associated with an action, but also on the (attention weighted) anticipated learning signal that they will receive on the basis of that choice. Formally, the model assumes that the decision weight $\tilde{V}(s_a)$ associated with the choice action s_a is given by

$$\tilde{V}(s_a) = V(s_a) + \tilde{E}(s_a, s_c), \tag{9}$$

where $V(s_a)$ denotes the temporally discounted reward value associated with state s, and $\tilde{E}(s_a, s_c)$ denotes the anticipated value of the prediction error that will occur when the cue is revealed.

According to the APE model, the subjective values associated with each state in the task are learned by a standard temporal difference learning mechanism.

$$V(s_a) \leftarrow V(s_a) + \alpha E(s_a, s_o), \tag{10}$$

where α is the learning rate parameter and $E(s_a, s_o)$ is the reward prediction error. For the APE model in this task—where the outcomes are independent of the learner actions and no savoring is built into the subjective reward—the prediction error is

$$E(s_a, s_o) = R(\cdot, s_o) - V(s_a), \tag{11}$$

where $R(\cdot, s_o)$ is the temporally discounted value of the reward to be received at the end of the delay period. Like the RPE-A model the discounting function is presumed to be exponential in nature, but is parameterized differently

$$R(\cdot, s_o) = \gamma^d R_\dagger(\cdot, s_o), \tag{12}$$

where d is the delay, γ is a discount parameter, and $R_\dagger(\cdot, s_o)$ is the primary reward value of the outcome s_o.

Because the expected reward is independent of the agent's decision, the long run behavior of this learning rule ensures that the reward value $V(s_f)$ of the "find out now" state s_f is the same as the value $V(s_k)$ of the "keep it secret" state. The reward signal itself provides no reason for the agent to prefer one choice over the other. Any preference that appears in the model is entirely due to the anticipatory component of the decision weight, $\tilde{E}(s_a, s_c)$.

As noted earlier, the APE model does not incorporate any inherent notion of savoring, and instead adopts the perspective that the agent constructs a myopic model of the task environment that allows them to seek out future learning signals. Specifically, what the learner does is anticipate every possible prediction error that might be encountered one step into the future. Formally, the anticipated prediction error is given by

$$\tilde{E}(s_a, s_c) = \sum_{s_c} P(s_c|s_a) w(s_c) (\gamma^d V(s_c) - V(s_a)). \tag{13}$$

In this expression, $P(s_c|s_a)$ denotes the probability that cue state s_c will follow the action state s_a. For instance the probability that the positive cue state s_+ follows the find out now state s_f is equal to the reward probability q, and the probability that the neutral cue state s_0 follows the keep it secret state s_k is 1. The $\gamma^d V(s_c) - V(s_a)$ term denotes the (temporally discounted) reward prediction error that would be encountered if state s_c does in fact follow state s_a.[4] Finally, the value of $w(s_c)$ is an attention weight parameter whose value denotes the extent to which the learner's anticipation process is "focused" on the possibility of reaching state s_c.

Because the design of the secrets task is simple, it is possible to construct analytic expressions that correspond to the difference $\tilde{V}(s_k) - \tilde{V}(s_f)$ for the two choices, as described by Zhu et al. (2017), which makes implementation of the model considerably simpler (e.g., simplifying the model by excluding free parameters such as the learning rate α). For our simulations, we apply the following equation:

[4] More generally, if the delay between making the choice and observing the cue is anything other than 1, the γ term would be raised to an exponent that reflects that delay length.

$$\tilde{V}(s_k) - \tilde{V}(s_f) = (w(s_w) - w(s_l)) \times \frac{\gamma^d |R(\cdot, s_w) + R(\cdot, s_l)|}{2}, \quad (14)$$

where s_w and s_l refer to win and lose states, respectively.

As with the other models, the probability of choosing to find out now is determined by a softmax decision rule applied to the decision weights $\tilde{V}$

$$P(s_f) = \frac{1}{1 + e^{\beta(\tilde{V}(s_k) - \tilde{V}(s_f))}}. \quad (15)$$

An illustration of the typical profile of choice probabilities as a function of delay length d is shown in the middle panel of Figure 2b.

Uncertainty Penalty Model

The *uncertainty penalty* (UP) model differs from the anticipation-based models in two respects, one theoretical and the other technical. Theoretically, instead of anticipating future rewards (like RPE-A) or future reward prediction errors (like APE), the agent is presumed to have an inherent information preference: uncertainty is assumed to be aversive and agents should prefer to receive advance information that resolves uncertainty faster than if the information was withheld. More formally, the outcome uncertainty $U(s)$ associated with any state s is given by the entropy function

$$U(s) = -P(s_w|s)\log_2 P(s_w|s) - P(s_l|s)\log_2 P(s_l|s), \quad (16)$$

where $P(s_w|s)$ and $P(s_l|s)$ are the respective probabilities of (eventually) reaching the winning state s_w or the losing state s_l given that the agent is currently in state s. This function is maximized when the probabilities of winning and losing are equal, and minimized (i.e., zero) either the probability of winning or losing is equal to one.

At a technical level, the UP model differs slightly in how value functions are computed. The RPE-A and APE models use iterative update rules to learn the value of a state $V(s)$ based on prediction error signals. The UP model does not specify a specific learning rule and simply notes that the optimal value function must satisfy Bellman's equation, which can be solved using dynamic programming in the general case. For the present purposes, it suffices to note that once the choice is made the events of a secrets trial are independent of the agent. In this case, the value $V(s)$ of any state in the task is given by

$$V(s) = \sum_{s'} P(s'|s)\,(R_\dagger(s, s') + V(s')e^{-kU(s')}), \quad (17)$$

where $P(s'|s)$ denotes the probability that the next state will be s', $R_\dagger(s, s')$ is the primary reward received when the agent leaves state s to arrive at state s' (which is zero unless s' is the win state s_w), and $V(s')$ is the value of the next state. In this expression the free parameter k is a weighting assigned to the agent's uncertainty $U(s')$, but in this task it functions in a fashion that is broadly similar to the temporal discounting parameter used in the other models. As with previous models, we apply a softmax rule to calculate choice probabilities

$$P(s_f) = \frac{1}{1 + e^{\beta(V(s_k) - V(s_f))}}, \quad (18)$$

where β is a free parameter in the model.

We assume that UP adopts a straightforward representation of cue-outcome delay. By defining a number of discrete waiting states that correspond to the length of the delay, UP can capture delay duration inherently in the MDP structure (Figure 4).[5]

The qualitative behavior of the UP model is illustrated in the right panel of Figure 2. Because the UP model relies on the uncertainty in the outcome independent of whether that outcome is good or bad, it generates the same curves for both positive and negative rewards. These curves are positively sloped because the uncertainty penalty associated with the entropy is applied repeatedly during the delay: the longer one has to wait for uncertainty to be resolved, the more aversive that uncertainty becomes.

Ultimately, each model offers differing accounts of how cue-outcome delay affects

[5] We note that UP does not necessitate a discrete implementation of delay, however we chose to do so to follow the original design of the model where increasing amounts of information was provided to the model in discrete steps.

Table 4
Features of Qualitative Predictions of Information-Seeking Models in Valence Sensitivity and Increasing Levels of Delay Length

Feature	RPE-A	APE	UP
Delay effect	Positive skew	Monotonic decreasing	Monotonic increasing
Information avoidance	Yes	Yes	No

Note. RPE-A = reward prediction error-anticipation; APE = anticipated prediction error; UP = uncertainty penalty.

information-seeking behavior, as well as how information-avoidant behavior can be produced (if at all). The key features of the qualitative behavior of each model are summarized in Table 4.

Model Fitting

We fit the RPE-A, APE, and UP models to the average FON choice proportions in each experiment. For Experiments 1 and 2, we fitted the models separately to each gain (or chocolate) and loss (or sound) condition and present them here as distinct fits for clarity of illustration. We also conducted individual-level model analysis by fitting each model to each participant's data and measuring the proportion of each sample to which a model was best fit.

- For the RPE-A model, we simulated 100 sessions of 200 trials to obtain stable measures of average choice probability of the informative cue, with initial state values set to 0. Six parameters were freely estimated: the temporal discounting parameter γ, the rate of anticipation gain ν, the relative baseline weight of anticipation η_0, the weight of prediction error boost c, the learning rate α, and the softmax determinism parameter β.
- For the APE model, we varied three parameters: the temporal discounting parameter γ, the attention weight toward the winning cue $w_w = w(s_w)$, the attention weight toward the losing cue $w_l = w(s_l)$. We fix the softmax determinism parameter $\beta = 1$ to allow for the identification of w_w and w_l.
- For the UP model, two parameters were estimated: the weight placed on uncertainty penalty k, and the determinism parameter for the choice rule β.

For the RPE-A model, reward values for positively valenced outcomes are specified with a value of 1, neutral outcomes are specified as 0, and negatively valenced outcomes are specified as -1. The APE and UP models are not sensitive to outcome valence, consequently for these models we fix the value of the more positive outcome to 1 and the less positive outcome to 0.

Model Fitting Results

The optimized model parameters are presented in Tables 5–7. Generally, these results seem to accord with our intuition—given that most data sets here do not show any evidence of temporal discounting, we would expect the role of the γ parameter for each anticipation model to be minimal. Note that the RPE-A uses γ as a negative exponential term, consequently, values close to 0 indicate minimal influence of γ, and increasing values indicate stronger temporal discounting.

Table 5
Optimized Parameters for RPE-A for Each Data Set

Data set	γ	ν	η_0	c	α	β
Exp1: Points win	0.337	30.001	15.043	1.724	0.167	0.010
Exp1: Points loss	1.073	5.92×10^{16}	<0.001	1.014	<0.001	0.908
Exp2: Chocolate	<0.001	0.437	1.216	0.873	<0.001	0.034
Exp2: Sound	89.028	$>10^{100}$	<0.001	0.004	1.000	5.480
Exp3: Mixed	8.397	$>10^{100}$	<0.001	0.089	<0.001	2.562

Note. RPE-A = reward prediction error-anticipation.

Table 6
Optimized Parameters for APE for Each Data Set

Data set	γ	w_w	w_l
Exp1: Points win	1.000	50.837	49.815
Exp1: Points loss	1.000	2.117	1.081
Exp2: Chocolate	1.000	0.510	10.001
Exp2: Sound	1.000	1.059	<0.001
Exp3: Mixed	1.000	0.752	−0.164

Note. APE = anticipated prediction error.

For data sets where RPE-A makes reasonable predictions, these γ values appear close to zero. For data sets that involve negatively valenced stimuli, this is not the case. APE implements γ as a multiplicative term bounded between 0 and 1, with values closer to 0 indicating greater influence of temporal discounting. Across all data sets, the model is optimized at γ values very close to 1.

Model predictions are overlaid on behavioral data in Figure 5. A visual inspection of the model predictions indicates that the UP model appears to be producing the best fits across all data sets. While RPE-A can generate comparable fits in data sets from conditions involving only positively valenced outcomes, the model struggles to predict data from conditions involving negatively valenced outcomes.

As indicated earlier, APE assumes no growth in anticipation over time, consequently preventing it from generating patterns of increasing information preference with increasing delay length. With such data, the best predictions it can generate resemble flat lines at the approximate average level of information seeking across all delay lengths. Although this is a poor fit to behavioral data showing delay-dependent behavior (i.e., both conditions in Experiment 2), APE is a reasonably accurate model of data indicating delay-invariant information preference (Experiments 1 and 3).

Table 7
Optimized Parameters for UP for Each Data Set

Data set	k	β
Exp1: Points win	0.556	1.132
Exp1: Points loss	0.356	1.248
Exp2: Chocolate	0.017	2.754
Exp2: Sound	0.062	2.101
Exp3: Mixed	0.781	0.959

Note. UP = uncertainty penalty.

Quantitatively, a comparison of Bayesian Information Criterion (BIC) values (Table 8) reveals that the UP model is preferred across all data sets. Even in conditions where only positively valenced outcomes are involved, UP ultimately shows better performance due to its ability to account for the patterns in the data using fewer free parameters.

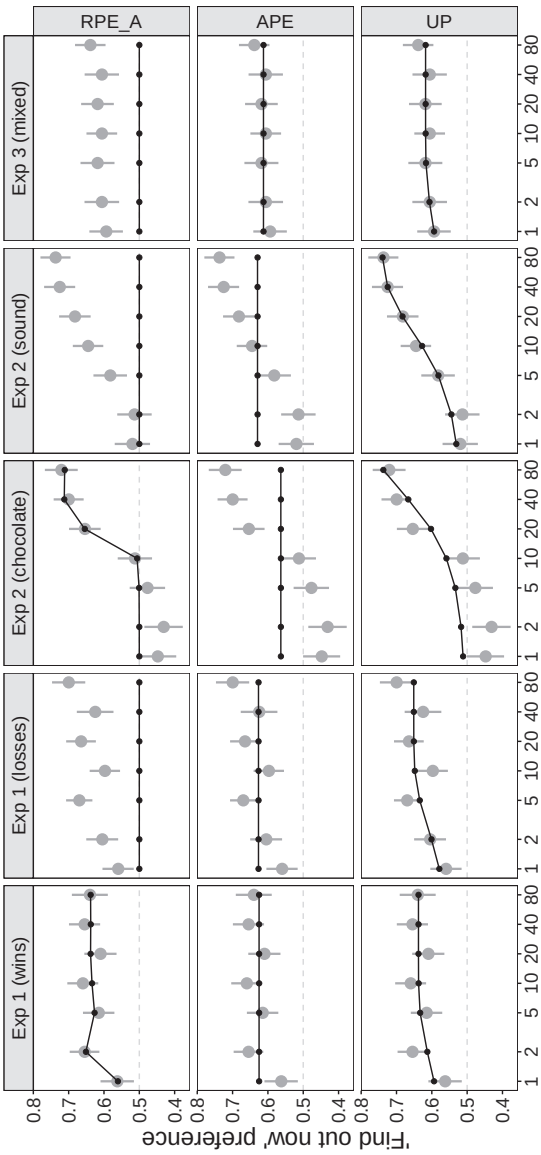
The strong performance of UP is also observed at the individual level where it is the best-fitting model for the majority of participants in each data set (Figure 6). Although APE's predictions of averaged FON choices were not particularly accurate, the model was able to strongly predict a minority of individual data. RPE-A was not found to best fit any individual data.

General Discussion

The observation of noninstrumental information seeking suggests information can be valued beyond how it can guide future decisions. While the literature indicates a general consensus that noninstrumental information seeking is robust under different situations, they deviate in their descriptions of how such behavior can vary. Specifically, contemporary models disagree on (a) how the length of cue-outcome delay can affect magnitude of information seeking, and (b) how, if at all, information-avoidance behavior might be generated. Our work sought to address these two questions at an empirical and theoretical level.

Unlike predictions offered by prior research (e.g., Iigaya et al., 2016; Spetch et al., 1990) we did not observe any effect of temporal discounting on information preferences at long delay lengths. None of our empirical data indicated any significant decrease in information-seeking behavior as delay length increased. Following speculation by Iigaya et al. (2016), it is possible that such effects would only be observable in additionally extreme lengths of delay. However, we remain skeptical of this position for three reasons. First, our current data do not indicate any negative trends with increasing delay lengths. Any extrapolation of current patterns indicate either delay invariance or a monotonically increasing preference for information. Second, our modeling work indicates that across all experiments, each model's best account of the data

Figure 5
Behavioral Data (Large Gray Circles, Vertical Lines Represent 1 Standard Error of the Mean) With Model Predictions of Mean FON Preferences (Black Points and Lines)



Note. FON = find out now.

Table 8
Model Comparison Statistics Across All Experiments

	BIC			BIC weights		
Data set	RPE-A	APE	UP	RPE-A	APE	UP
Exp1: Points win	3213.53	3198.85	**3187.0**	0.00	0.00	1.00
Exp1: Points loss	3373.81	3194.74	**3177.1**	0.00	0.00	1.00
Exp2: Chocolate	4484.81	4628.15	**4477.1**	0.02	0.00	0.98
Exp2: Sound	4998.15	4731.71	**4613.1**	0.00	0.00	1.00
Exp3: Mixed	4803.83	4604.16	**4595.0**	0.00	0.01	0.99

Note. BIC = Bayesian Information Criterion; RPE-A = reward prediction error-anticipation; APE = anticipated prediction error; UP = uncertainty penalty. Bold values indicate lowest BIC values across models.

involve specifications of their temporal discount parameter to a value indicating minimal influence. Model simulations with these specific parameter values cannot produce any significant pattern of temporal discounting at any arbitrarily long delay length. In essence, our best theoretical accounts indicate no evidence for temporal discounting in the current paradigm. Third, monotonic patterns may not be uncommon outside the immediate literature. For instance, Story et al. (2013) found that the aversiveness of a future painful outcome (e.g., a dentist's appointment) increased monotonically with delay (measured on a scale from days to months) with no sign of decrease at extended delays. If we can expect that information preferences are related to the aversiveness (or attractiveness) of an event, then monotonic patterns over delay length may not ultimately be surprising.

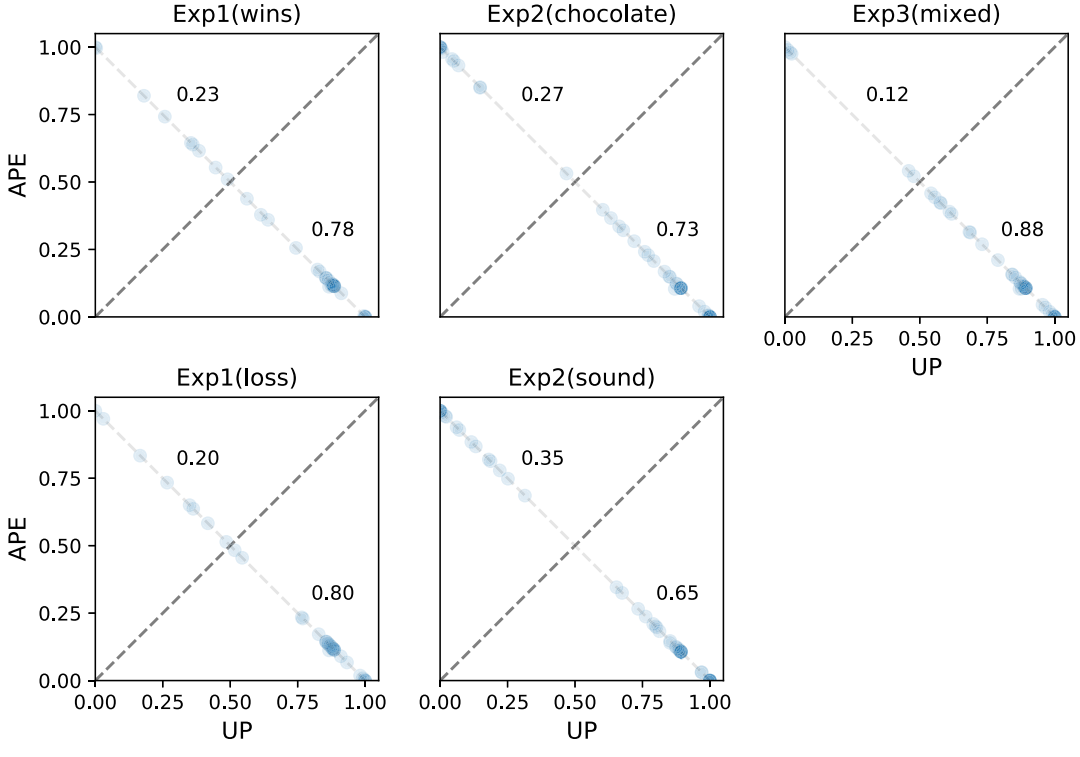
In Experiment 3 we found not only no evidence of temporal discounting, but also an absence of any clear delay dependency of information-seeking behavior. This absence is surprising because there is no clear mechanism to suggest how the lack of a neutral outcome can make seeking information less dependent on cue-outcome delays. More generally, the findings of Experiment 3 indicate that the neutral reward may be necessary in order to produce a monotonically increasing delay-effect—as seen in Iigaya et al. (2016) who also used neutral rewards. One possible explanation for this pattern is that with neutral rewards, people may have been more interested in obtaining information about gaining/losing nothing at long delays than short delays (possibly due to having to "endure" longer periods of time for nothing). However, current models do not immediately allow for a straightforward account of this interaction between neutral rewards and delay length, nor do they explain how this can differ between primary and secondary reinforcers. More generally, none of the current models offer any straightforward account of the invariant behavior observed in Experiment 3, indicating that future work should focus on both empirical and theoretical investigations of this finding.

Our research also provided no clear evidence of noninstrumental information-avoidant behavior in the current paradigm. Contrary to predictions by Iigaya et al. (2016), Zhu et al. (2017), the presence of losses as a possible outcome indicated primarily information-seeking behavior. The strongest hints of information avoidance was observed only in the chocolate condition of Experiment 2, and only specifically at very low delays (<= 2 s). Even then, the average results for that data set indicated information indifference as opposed to information avoidance. It is possible that robust noninstrumental information avoidance can be generated under different situations, however current models do not clearly indicate how this can occur.

To be clear, our results do not indicate that information-avoidant behavior is nonexistent, but rather, it is not the typical (averaged) behavior in our paradigm. Indeed, individual variation in our data indicates a small proportion of people showing information-avoidant behavior. More generally, there are many situations outside the laboratory that clearly indicate some level of information avoidance (e.g., see Golman et al., 2017; Sweeny et al., 2010), and while most appear to involve instrumental information, some arguably contain information that can be understood to be noninstrumental. For instance, Sicherman et al. (2016) observed that investors were more likely to avoid repeated (and therefore

Figure 6

BIC Weights for UP (x-Axis) and APE (y-Axis) Fit to Individual Data



Note. The positively sloped dashed line indicates equal weight between uncertainty penalty (UP) and anticipated prediction error (APE)—markers located on the lower-right of the plot indicate greater weight on UP and markers on the upper-left indicate greater weight on APE. Inset numerical values indicate the proportions of markers in these respective locations. BIC weights for reward prediction error-anticipation (RPE-A) are indicated by the distance of markers from the negatively sloped dashed line toward the origin (the lack of markers with any noticeable deviation from the line indicate minimal weight on RPE-A). See the online article for the color version of this figure. BIC = Bayesian Information Criterion.

noninstrumental) information on the value of their investment portfolio when the stock market was performing poorly (similar to the findings of Charpentier et al.'s (2018) laboratory task). Although we assert that information avoidance is not the dominant behavior with constructed noninstrumental information tasks, these assertions cannot easily be generalized outside the laboratory without further work. More generally, future investigations would benefit from a systematic approach to explaining the individual variation in information-seeking behavior.

Our modeling work ultimately indicates that when people seek noninstrumental information, they are seeking the resolution of uncertainty rather than savoring the anticipation of future outcomes. Unlike the latter, the former assumes no temporal discounting of rewards and no dependency of information avoidance on any specific mechanism, which is clearly observed in our data.

While both RPE-A and UP models are able to accurately reproduce patterns of information seeking seen in the data from conditions with only positively valenced outcomes, it is the relative simplicity of UP in these comparisons that makes this model particularly appealing. Without assuming multiple aspects to its core mechanism of information seeking (as is seen in RPE-A where anticipation is decomposed into a baseline, boost strength, and gain rate separately), UP elegantly describes information seeking using a single parameter (assuming minimal impact of β). Even though RPE-A can produce similar behavior with only positively valenced rewards, it does so with a structurally and conceptually more complex design.

Although it is conceptually similar to RPE-A, the APE model struggled to reproduce the data to the same extent as the other models. In particular, its inability to accumulate anticipation over increasing delay length prevented it from providing a reasonable fit to averaged data from Experiment 2.

It is possible that these complexities of RPE-A as defined by the original authors are not completely necessary—we could imagine that the mechanistic reinforcement-learning structure is reduced into a single construct (with fewer free parameters) that captures the relevant amount of anticipation associated with a given choice. This simplified anticipation-based model may end up performing just as well as, or even better than, the UP model. However, this can only happen when no negatively valenced rewards are available in the task. As long as the utility from anticipation is directly mapped onto outcome valence, it will not be possible for the model to predict the information-seeking patterns observed with negatively valenced rewards. We could then go even further and allow the single anticipation-based construct to be free from the valence of rewards—that is, the model can be agnostic to outcome valence just as the UP model is. However, doing so risks redefining the original RPE-A model into one that is nearly indistinguishable (at least, within our paradigm) from that of UP, at which point a model comparison between the two may no longer be diagnostic. Given that the goal of the current work is to compare models that are qualitatively distinct in their predictions, the consideration of other models (e.g., Beierholm & Dayan, 2010), or model variants, that can produce the same kind of behavior is ultimately of limited utility to our investigation. We discuss this issue further in the Supplemental Materials, in particular demonstrating how the Beierholm and Dayan (2010) model can produce the different qualitative predictions of RPE-A, APE, and UP.

While our results challenge prior speculative predictions, empirically they align closely with findings from previous studies (e.g., Charpentier et al., 2018; Iigaya et al., 2016; Zhu et al., 2017), none of which have found reliable attenuation of information seeking at long delays or consistent information-avoidant behavior. The diverse background of participants from these various studies and ours (from university students to general members of the public) suggest that the observed patterns of behavior are robust and not specific to any particular sample of adults. However, we note that these patterns have only been tested and observed within controlled laboratory studies, and consequently we have no evidence to suggest that such behavior necessarily generalizes outside a controlled environment. We have no reason to believe that our specific results depend on other characteristics of the participants, materials, or context.

The apparent simplicity of the observed data compared to current theories also raises a separate issue: under what other circumstances (beyond cue-outcome delay) can we expect noninstrumental information-seeking behavior to change? While one way to explore this further would be to investigate other empirical means of producing noninstrumental information avoidance, alternative avenues include examining how quickly people can learn the relative optimality of

choosing an informative versus noninformative cue, or whether the absolute magnitude of rewards affects information-seeking behavior. Suffice to say, the robustness of general noninstrumental information seeking is easily apparent, but the conditions under which (and relatedly, how) such behavior systematically differs remains an important avenue for further investigation.

References

Beierholm, U. R., & Dayan, P. (2010). Pavlovian-instrumental interaction in "observing behavior." *PLoS Computational Biology*, *6*(9), Article e1000903. https://doi.org/10.1371/journal.pcbi.1000903

Bennett, D., Bode, S., Brydevall, M., Warren, H., & Murawski, C. (2016). Intrinsic valuation of information in decision making under uncertainty. *PLoS Computational Biology*, *12*(7), Article e1005020. https://doi.org/10.1371/journal.pcbi.1005020

Bennett, D., Sutcliffe, K., Tan, N. P.-J., Smillie, L. D., & Bode, S. (2021). Anxious and obsessive–compulsive traits are independently associated with valuation of noninstrumental information. *Journal of Experimental Psychology: General*, *150*(4), 739–755. https://doi.org/10.1037/xge0000966

Bromberg-Martin, E. S., & Hikosaka, O. (2011). Lateral habenula neurons signal errors in the prediction of reward information. *Nature Neuroscience*, *14*(9), 1209–1216. https://doi.org/10.1038/nn.2902

Cabrero, J. M. R., Zhu, J.-Q., & Ludvig, E. A. (2019). Costly curiosity: People pay a price to resolve an uncertain gamble early. *Behavioural Processes*, *160*, 20–25. https://doi.org/10.1016/j.beproc.2018.12.015

Charpentier, C. J., Bromberg-Martin, E. S., & Sharot, T. (2018). Valuation of knowledge and ignorance in mesolimbic reward circuitry. *Proceedings of the National Academy of Sciences*, *115*(31), E7255–E7264. https://doi.org/10.1073/pnas.1800547115

Cox, T. J. (2008). Scraping sounds and disgusting noises. *Applied Acoustics*, *69*(12), 1195–1204. https://doi.org/10.1016/j.apacoust.2007.11.004

Crockett, M. J., Braams, B. R., Clark, L., Tobler, P. N., Robbins, T. W., & Kalenscher, T. (2013). Restricting temptations: Neural mechanisms of precommitment. *Neuron*, *79*(2), 391–401. https://doi.org/10.1016/j.neuron.2013.05.028

De Leeuw, J. R. (2015). jsPsych: A javascript library for creating behavioral experiments in a web browser. *Behavior Research Methods*, *47*(1), 1–12. https://doi.org/10.3758/s13428-014-0458-y

Eliaz, K., & Schotter, A. (2010). Paying for confidence: An experimental study of the demand for non-instrumental information. *Games and Economic Behavior*, *70*(2), 304–324. https://doi.org/10.1016/j.geb.2010.01.006

Epstein, L. G., & Zin, S. E. (1989). Substitution, risk aversion, and the temporal behavior of consumption and asset returns: A theoretical framework. *Econometrica*, *57*(4), 937–969. https://doi.org/10.2307/1913778

Golman, R., Hagmann, D., & Loewenstein, G. (2017). Information avoidance. *Journal of Economic Literature*, *55*(1), 96–135. https://doi.org/10.1257/jel.20151245

Gottlieb, J., Hayhoe, M., Hikosaka, O., & Rangel, A. (2014). Attention, reward, and information seeking. *Journal of Neuroscience*, *34*(46), 15497–15504. https://doi.org/10.1523/JNEUROSCI.3270-14.2014

Hertwig, R., & Engel, C. (2016). Homo ignorans: Deliberately choosing not to know. *Perspectives on Psychological Science*, *11*(3), 359–372. https://doi.org/10.1177/1745691616635594

Iigaya, K., Hauser, T. U., Kurth-Nelson, Z., O'Doherty, J. P., Dayan, P., & Dolan, R. J. (2020). The value of what's to come: Neural mechanisms coupling prediction error and the utility of anticipation. *Science Advances*, *6*(25), Article eaba3828. https://doi.org/10.1126/sciadv.aba3828

Iigaya, K., Story, G. W., Kurth-Nelson, Z., Dolan, R. J., & Dayan, P. (2016). The modulation of savouring by prediction error and its effects on choice. *eLife*, *5*, Article e13747. https://doi.org/10.7554/eLife.13747

Kidd, C., & Hayden, B. Y. (2015). The psychology and neuroscience of curiosity. *Neuron*, *88*(3), 449–460. https://doi.org/10.1016/j.neuron.2015.09.010

Kobayashi, K., Ravaioli, S., Baranes, A., Woodford, M., & Gottlieb, J. (2019). Diverse motives for human curiosity. *Nature Human Behaviour*, *3*(6), 587–595. https://doi.org/10.1038/s41562-019-0589-3

Kreps, D. M., & Porteus, E. L. (1978). Temporal resolution of uncertainty and dynamic choice theory. *Econometrica: Journal of the Econometric Society*, *46*(1), 185–200. https://doi.org/10.2307/1913656

Kreps, D. M., & Porteus, E. L. (1979). Dynamic choice theory and dynamic programming. *Econometrica: Journal of the Econometric Society*, *47*(1), 91–100. https://doi.org/10.2307/1912348

Loewenstein, G. (1987). Anticipation and the valuation of delayed consumption. *The Economic Journal*, *97* (387), 666–684. https://doi.org/10.2307/2232929

Navarro, D. J., Newell, B. R., & Schulze, C. (2016). Learning and choosing in an uncertain world: An investigation of the explore–exploit dilemma in static and dynamic environments. *Cognitive Psychology*, *85*, 43–77. https://doi.org/10.1016/j.cogpsych.2016.01.001

Pierson, E., & Goodman, N. (2014). Uncertainty and denial: A resource-rational model of the value of information. *PLOS ONE*, *9* (11), Article e113342. https://doi.org/10.1371/journal.pone.0113342

Sharot, T., & Sunstein, C. R. (2020). How people decide what they want to know. *Nature Human Behaviour*, *4*(1), 14–19. https://doi.org/10.1038/s41562-019-0793-1

Sicherman, N., Loewenstein, G., Seppi, D. J., & Utkus, S. P. (2016). Financial attention. *The Review of Financial Studies*, *29*(4), 863–897. https://doi.org/10.1093/rfs/hhv073

Spetch, M. L., Belke, T. W., Barnet, R. C., Dunn, R., & Pierce, W. D. (1990). Suboptimal choice in a percentage-reinforcement procedure: Effects of signal condition and terminal-link length. *Journal of the Experimental Analysis of Behavior*, *53*(2), 219–234. https://doi.org/10.1901/jeab.1990.53-219

Story, G. W., Vlaev, I., Seymour, B., Winston, J. S., Darzi, A., & Dolan, R. J. (2013). Dread and the disvalue of future pain. *PLoS Computational Biology*, *9* (11), Article e1003335. https://doi.org/10.1371/journal.pcbi.1003335

Sweeny, K., Melnyk, D., Miller, W., & Shepperd, J. A. (2010). Information avoidance: Who, what, when, and why. *Review of General Psychology*, *14* (4), 340–353. https://doi.org/10.1037/a0021288

Vasconcelos, M., Monteiro, T., & Kacelnik, A. (2015). Irrational choice and the value of information. *Scientific Reports*, *5*(1), 1–12. https://doi.org/10.1038/srep13874

Zhu, J.-Q., Xiang, W., & Ludvig, E. A. (2017). Information seeking as chasing anticipate prediction errors. In G. Gunzelmann, A. Howes, T. Tenbrink, & E. J. Davelaar (Eds.), *Proceedings of the 39th annual conference of the cognitive science society* (pp. 3658–3663). Cognitive Science Society.

Received June 1, 2021
Revision received February 8, 2022
Accepted February 9, 2022 ■