

Dissociable mechanisms of information sampling in prefrontal cortex and the dopaminergic system

Paula Kaanders^{1,2}, Keno Juechems^{1,3}, Jill O'Reilly^{1,2} and Laurence Hunt^{1,3}



Recently, neuroscientists have become increasingly interested in studying the interactions between information sampling and choice and the mechanisms underlying these. In machine learning, introducing intrinsic rewards for exploration has been found to greatly improve artificial agents' performance on 'hard exploration' problems. There is evidence that humans are intrinsically driven to sample both information that has no direct impact on reward outcome as well as information that reduces uncertainty on upcoming decisions. Recent findings from studies using a range of information sampling tasks suggest a functional dissociation between more posterior and anterior regions of prefrontal cortex (PFC). Specifically, pre-supplementary motor area (pre-SMA) and dorsal anterior cingulate cortex (dACC) are involved in decisions to sample more information to guide upcoming decisions, whereas the more anterior ventromedial prefrontal cortex (vmPFC), encodes the value of upcoming information. We argue that to effectively study information sampling in humans, the behavioral tasks we use must better reflect the large state space available to humans in real life. This, however, is challenging due to the complex model of the world humans have access to when choosing where to sample next.

Addresses

¹ Wellcome Centre for Integrative Neuroimaging, University of Oxford, Oxford OX3 9DA, United Kingdom

² Department of Experimental Psychology, University of Oxford, Oxford OX2 6GG, United Kingdom

³ Department of Psychiatry, University of Oxford, Oxford OX3 7JX, United Kingdom

Corresponding authors:

Kaanders, Paula (paula.kaanders@psy.ox.ac.uk),
Hunt, Laurence (laurence.hunt@psych.ox.ac.uk)

Current Opinion in Behavioral Sciences 2021, 41:63–70

This review comes from a themed issue on **Cognition and perception**
- ***value-based decision-making***

Edited by **Bernard Balleine** and **Laura Bradfield**

<https://doi.org/10.1016/j.cobeha.2021.04.005>

2352-1546/© 2021 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Our decision-making is determined by the information we gather before choosing. While much scientific work

has been done to investigate how humans decide what information to sample at what time, we do not yet fully understand how these sampling choices guide future decision-making, and in turn, how past decisions affect subsequent information sampling [1]. Here we review recent work aiming to better understand this: (i) through the development of reinforcement learning algorithms that display similar exploratory behavior to that seen in humans; (ii) the study of human participants' sampling behavior using experimental designs that dissociate information sampling from reward-maximizing choices; and (iii) the study of the neural mechanisms underlying these behaviors.

Information sampling by artificial agents

'Hard exploration' problems in machine learning are environments in which rewards are sparse and can only be obtained by performing a specific sequence of actions [2]. Traditional reinforcement learning agents struggle to obtain any reward in this type of environment, as they fail to identify and learn the long action sequences necessary to discover these sparse rewards. Introducing *intrinsic* rewards—rewards for gaining information about the environment, that augment the externally delivered reward signal—has improved performance of these agents on such problems, to the point where they outperform humans on tasks with sparse rewards, such as the Atari game 'Montezuma's Revenge' [3,4,5,6–12,13].

Are intrinsic rewards learned?

This idea of including an intrinsic reward for exploring suggests that there may be some innate preference for information search. It has been shown that reinforcement learning agents can learn intrinsic rewards through experience, but that these can also be passed down to new agents via their reward function in an evolutionary manner [13,14]. Furthermore, many findings within developmental psychology suggest younger human learners show more exploratory behavior than older learners, who are typically more goal-directed [15]. For this 'explore-first/exploit-later' strategy to be present, some innate motivation in young agents to seek new information is to be expected. Indeed, infants have been shown to prefer stimuli of medium complexity [16]. This behavior is in line with Loewenstein's information-gap hypothesis, which argues that curiosity is at its highest when the agent already possesses a moderate amount of knowledge [17].

Types of information sampling behavior

In neuroscience, ‘exploration’ is often defined as foregoing the current best option in order to sample information about an alternative, in multi-alternative bandit or foraging tasks [1*,18**]. However, to maximize reward effectively an agent must also learn the structure of its environment through exploration [19]. This may be a main reason why children are highly exploratory and take delight in pursuing seemingly arbitrary goals they have yet to learn the full structure of their surroundings [14]. This type of ‘exploration’ is often referred to as ‘curiosity’, but is not frequently studied experimentally in human participants [1*,20].

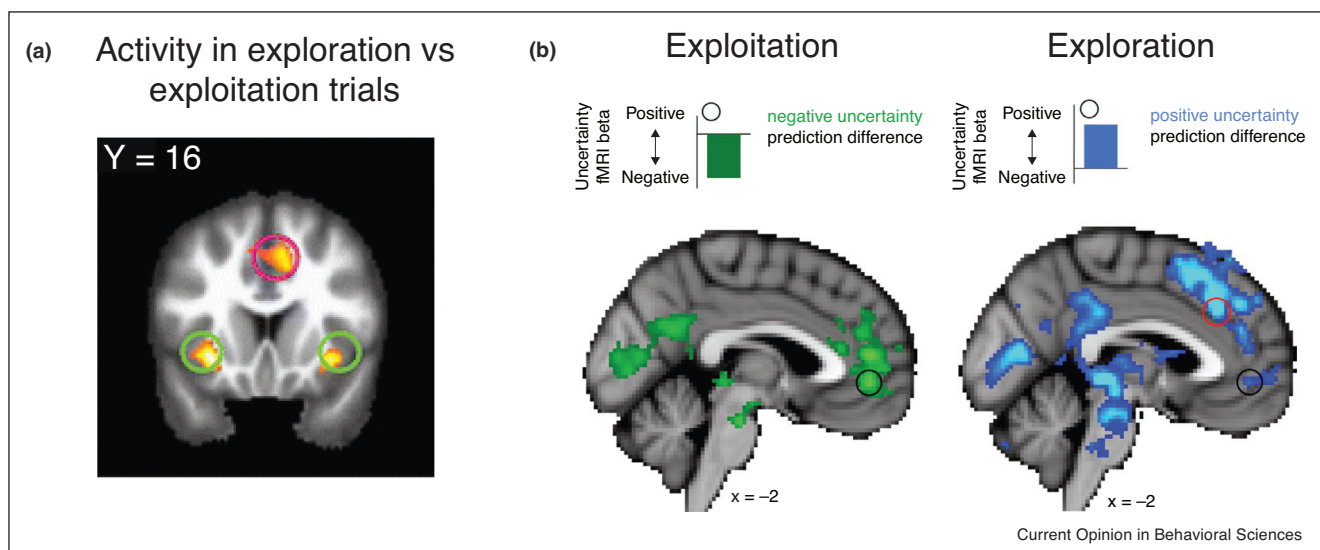
Recent discoveries in machine learning on hard exploration problems beg the question whether humans sample information in a similar way that is to say, by an intrinsic motivation to seek out novelty. Recent studies have indeed shown that both non-human primates and human participants are willing to pay for information to reveal information that does not affect the outcome [21*,22], building on an established literature documenting preferences for early ‘temporal resolution of uncertainty’ in such problems [23]. By contrast, information sampling could also be guided in a goal-driven manner in order to optimize future decisions [24]. Again, the goal here can either be to identify new sources of reward or to improve the agent’s model of the environment, but most behavioral tasks used to investigate this type of information search focus on the former [1*].

The distinction between sampling of information that is directly useful to guide upcoming decisions and of information that is not, is commonly referred to as instrumental versus non-instrumental information seeking. Recent behavioral data suggests that a combination of both non-instrumental and instrumental factors drive human information sampling [25,26,27*], but as of yet it is unclear what determines how each factor is weighted during sequential decisions to sample information. It is possible that what is deemed as non-instrumental information sampling in decision-making tasks, is in fact a form of exploration to reduce uncertainty about the structure of the environment rather than maximizing reward.

Exploitation versus exploration

One of the best-described types of information sampling behavior is that shown in explore-exploit tasks [18**,28,29]. In such studies, prefrontal cortex (PFC) activity has been found to predict exploratory choices of uncertain options (Figure 1; [18**,29,30,31**,32*,33**]). More specifically, Trudel *et al.* [33**] found that prefrontal subregions play distinct roles in resolving the exploitation–exploration dilemma: while dorsal anterior cingulate cortex (dACC) encoded uncertainty only when the agent was in an exploratory phase, ventromedial prefrontal cortex (vmPFC) encoded uncertainty throughout the task, but the polarity of its activity changed as a function of whether the agent was in an exploitative or explorative phase (Figure 1b).

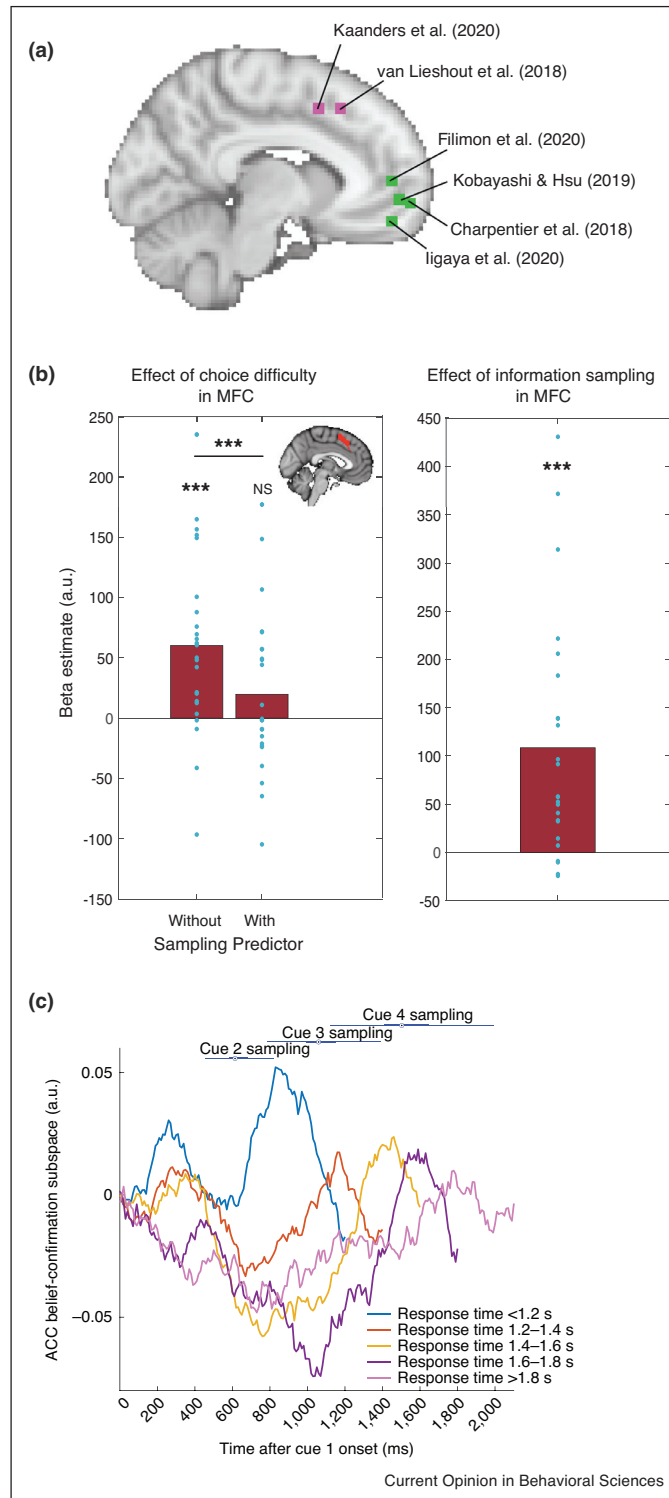
Figure 1



Activity in anterior cingulate cortex (ACC) is related to decisions to sample more information in exploitation-exploration tasks.

(a) In a task where participants could choose between ‘observing’ and ‘betting’ on each trial, trials in which participants chose to ‘observe’, larger activity in dACC (red circle) and insula (green circles) was observed in an fMRI study [18*]. (b) In another fMRI study, dACC only encoded uncertainty when participants were being exploratory (red circle, right panel), while vmPFC encoded uncertainty throughout the task (black circle), but uncertainty-related activity changed polarity depending on whether the participant was in an exploitative (left panel) or exploratory phase (right panel) [33**].

Figure 2



Prefrontal activity is associated with information sampling in tasks where sampling is explicitly separated from reward-maximizing choices. **(a)** Locations of peak activity associated with information sampling in pre-SMA/dACC (purple) and vmPFC (green) in a number of recent human fMRI studies, collapsed across the x-axis [34**,36**,39**,40**,42**,43*]. **(b)** An effect of inverse value difference on BOLD signal in medial frontal cortex (MFC) is significantly reduced when including information sampling as a co-regressor in an economic choice task [34*]. **(c)** Macaque single-cell activity in dACC is related to how much newly revealed information confirms a previously formed belief about which of two options to choose to maximize reward in an economic choice task [35].

A challenge in these tasks is that it is hard to dissociate decisions to sample information, or ‘exploration’, from decisions to maximize expected reward, or ‘exploitation’. Blanchard and Gershman [18**] addressed this by separating actions the agent can take into ‘observation’ and ‘betting’ actions and found larger dACC activity on observation trials (Figure 1a). This still presents information sampling and decision-making as a trade-off, though, while in reality this trade-off may not exist as agents continuously sample information to guide upcoming choices.

The value of information versus choosing to sample

Other recent studies have separated information sampling from choice by giving subjects the explicit opportunity to actively sample information before making a choice [34**,35,36**,37**]. In these tasks, prefrontal activity has again been found to be related to information sampling [34**,35,36**,37**,38*] (Figure 2b-c). Looking more carefully at the specific regions related to information sampling, there again appears to be a dissociation between activity in two regions: a more posterior and dorsal region, including pre-supplementary motor area (pre-SMA) and dACC, and ventromedial prefrontal cortex (vmPFC; [34**,36**,39**,40**,41,42**,43**], Figure 2a). At rest, these two regions show strong functional coupling with each other, but pre-SMA/dACC are also strongly coupled with sensorimotor regions, frontal pole, dorsal prefrontal cortex, and insula, among others [41]. On the other hand, vmPFC shows strong coupling with temporal cortex, posterior cingulate, precuneus and crucially, ventral striatum, which receives the strongest dopaminergic input from reward-sensitive parts of the dopamine midbrain [41]. In contrast, dACC is more strongly connected to dorsal striatum, which mostly receives inputs from motoric parts of the dopamine midbrain [44].

This dissociation in functional coupling can be linked to the different features of the specific behavioral tasks used to find neural activity in these regions related to information sampling. Generally, it seems to be the case that pre-SMA/dACC activity is predictive of subsequent actions by the agent to gather more information to guide an upcoming choice [34**], while vmPFC encodes the value of upcoming information [36**,40**,42**,43**]. For example, Iigaya *et al.* [40**] found an anticipatory value signal in vmPFC, which is the modelled value associated with the information that a reward will be received later on (Figure 3a-b). Similarly, Charpentier *et al.* [43**] found an adjacent region is activated in response to the opportunity to receive information about future outcomes (Figure 3c).

More evidence for such a functional dissociation between these two regions in information sampling comes from the long literature on novelty encoding in the dopaminergic

midbrain [46]. Recent findings elaborate on this by showing that ventral striatum, which is strongly coupled with vmPFC, itself also encodes the value of information [36**,39**,42**,43**,47**]. In other words, while more posterior regions, such as pre-SMA and dACC have been shown to drive sequential choices to sample information to guide future choice, vmPFC and the dopaminergic system mostly seem to encode the value of that information but do not necessarily themselves drive behavior. As these regions are strongly coupled with each other as well, it is likely that pre-SMA/dACC uses value representations in vmPFC to drive sampling. This is in line with broader accounts of the roles of these regions in decision-making in general: vmPFC is often considered to encode the value of stimuli [48], while pre-SMA/dACC have frequently been described as driving behavioral shifts [49*,50,51].

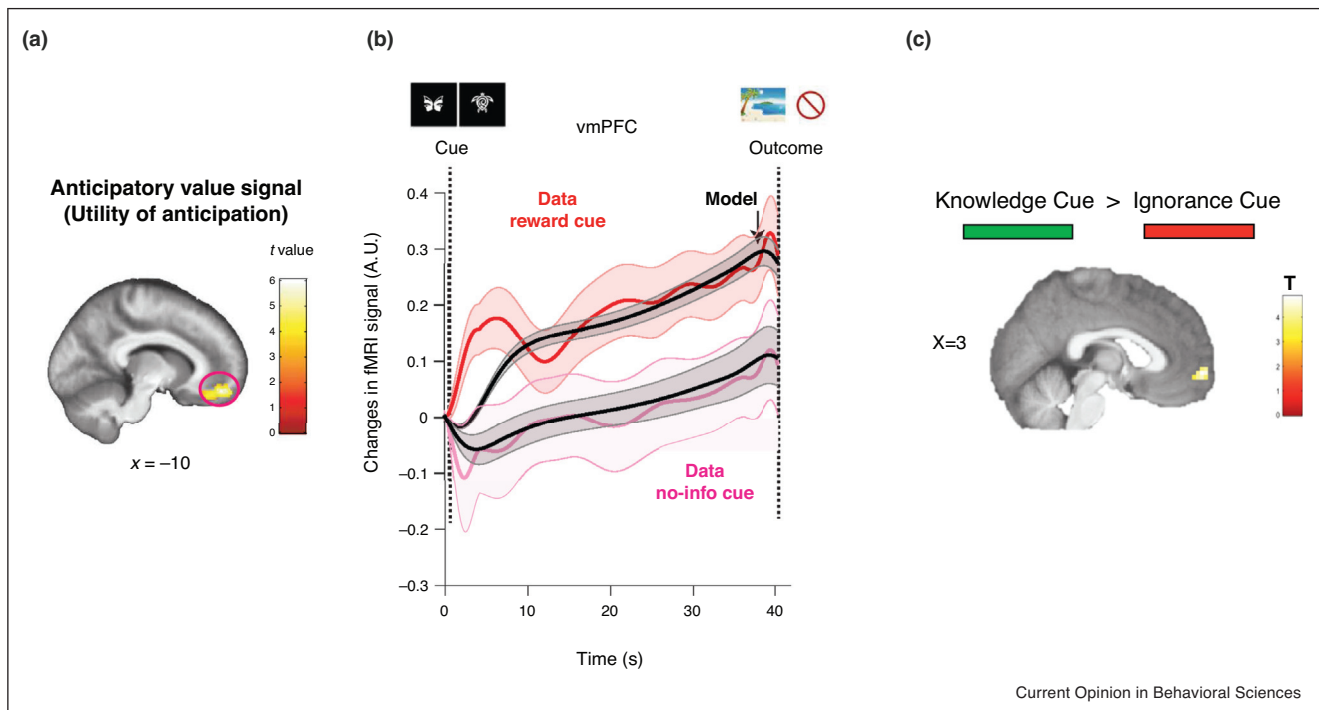
Activity in the intraparietal sulcus (IPS) has also been found in some paradigms to be related to sampling, to predict the moment of information gain, or outcome uncertainty [34**,39**,40**,52*,53]. Both IPS and dACC are part of a network often identified in decision-making tasks, responding to outcomes different from the agent’s predictions that require belief updating and behavioral shifts [49*,50,54–58]. More specifically, IPS seems particularly sensitive to surprising changes in the environment, while dACC activity is more related to the actual updating of beliefs that leads to behavioral change [49*]. This suggests this network might use estimates of uncertainty to drive information sampling for decision-making.

New approaches and challenges to studying information sampling for decision-making

Behavioral tasks used to investigate information sampling in decision-making are typically highly simplified, often with only two available choice options, and the amount of information that can be sampled to inform choice therefore also tends to be small [1*]. This approach has been crucial in characterizing decision-making mechanisms from behavioral and neural data [59,60], but it is less useful to begin to disentangle the many potential drivers of information sampling, their neural representations and how they affect choice [1*]. While we have argued that there is a functional dissociation in frontal cortex between signals that encode information value and those that directly drive sampling, our understanding of the different factors and their neural representations that lead an information source to be considered more or less valuable is poor. To improve it, the tasks we use need to better reflect the large amount of goals, choice options and information available in the world.

This is easier said than done. Firstly, human agents do not just sample information from their environment, they also use their prior knowledge of the world to make decisions or learn new tasks [61]. Without a full account of this prior

Figure 3



Activity in vmPFC is associated with the value of information. **(a)** BOLD activity in vmPFC is correlated with the anticipatory value of knowing about upcoming reward ahead of time [40**]. **(b)** The temporal evolution of this BOLD signal matches the anticipatory utility signal predicted by a reinforcement learning model that includes a preference for advance reward information [40**,45]. **(c)** vmPFC is more active during the delivery of informative cues than uninformative cues [43**].

knowledge, it is difficult to identify what exactly drives information sampling when faced with a new task. This is also an important distinction between humans and RL agents: the latter typically do not enter a new task with such prior knowledge. Dubey *et al.* [62**] found that gradually removing this prior knowledge from human agents (by modifying the visual information in a video game environment), drastically reduced performance, suggesting prior knowledge is a very important factor in effectively driving sampling to solve new problems. Another way in which human participants obtain such prior knowledge is through the instructions given by the experimenter before starting the task. Human participants perform much better at a novel task when given prior information in the form of instructions or by watching another agent play the game, which could be one of the main reasons why human agents learn new tasks much faster than RL agents [63].

The ability to construct and maintain a complex model of the world is likely crucial to the effective driving of information sampling for learning and solving new complex problems, and could be the main reason human agents struggle less with ‘hard exploration’ problems than their RL counterparts. Ideas derived from RL algorithms

have previously been successfully applied to study human behavior and neural data [64]. The knowledge that the performance of RL agents on ‘hard exploration’ problems is dramatically improved by including intrinsic motivation to seek out new information, could be used in a similar fashion to better understand human information sampling. This is unlikely to be a very good model, though, as human agents probably also use their model of the world to guide sampling of the information likely to be most valuable in the task, which is hard to replicate in an RL algorithm. As such, the presence of such a model in human agents might severely limit how much we can learn about human information sampling from studying RL agents. However, it is not inconceivable that some elements of a model of the world might effectively be introduced into AI algorithms, such as intuitive physics and psychology [65], which could lead to behavior that approximates that of humans.

Progress along these lines will likely come from two sides: First, human studies need to be suitably complex in the stimulus or action domain (e.g. by using many possible stimulus dimensions, or using video game-like tasks with many possible actions). This will allow researchers to harness recent advances in machine learning algorithms

that specifically target good compressions of information into a useful state space to enable efficient learning [66]. In doing so, researchers can use tasks and stimuli that are novel to participants. Second, researchers will likely benefit from the recent advances in ‘meta-learning’ [67]. Here, more complex tasks allow researchers to explicitly derive normative (or at least performative) exploration algorithms directly from data without the need to define them by hand. For instance, Zheng *et al.* [13**] encoded knowledge about likely task structures in the intrinsic reward signal, showing that this enabled the agent to behave efficiently in novel environments. Such work will allow researchers to interrogate learned curiosity signals and use these insights to design human experiments or to compare human performance against a largely assumption-free, yet performative baseline. Taken together, these tasks would yield predictions about both ‘what’ to sample (through the learned state space) and ‘whether’ to sample (through the learned exploration or intrinsic reward function) and would recover some of the tractability often lost when introducing more complexity into task design.

Conclusions

In this review, we have highlighted some recent findings from a range of behavioral tasks studying different types of information sampling. In these studies, prefrontal activity is often related to information sampling, but there appears to be a functional dissociation between more posterior regions, such as pre-SMA and dACC, and the more anterior vmPFC. We argue that pre-SMA/dACC drives the agent to sample more information before committing to a decision, while vmPFC activity encodes the value of upcoming information, but does not directly affect decisions to sample. The functional connectivity profiles of these regions, with pre-SMA/dACC being strongly connected to sensorimotor regions and vmPFC to the reward-sensitive dopaminergic system, support this hypothesis. What remains unclear is what the different drivers are of information search and how they are represented in the brain. In a number of the findings described here, the information sought by agents was non-instrumental to the decision at hand, suggesting information search is often driven by goals other than those set by the experimenter. Finally, we propose that to better understand how these information representations arise, we must develop behavioral tasks that better reflect the real decision-making problems humans face. RL may help here, but its use may be limited as RL agents do not possess the complex representation of the world we live in that we use to make decisions every day.

Conflict of interest statement

Nothing declared.

Acknowledgements

P.K. is supported by a Wellcome Trust studentship. K.J. is supported by a St John’s College Junior Research Fellowship. J.O.R. is supported by a Career Development Award from the Medical Research Council (MR/L019639/1). L.T.H. is supported by a Sir Henry Dale Fellowship from Wellcome and the Royal Society (208789/Z/17/Z).

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- of special interest
 - of outstanding interest
1. Gottlieb J, Oudeyer P-Y: **Towards a neuroscience of active sampling and curiosity.** *Nat Rev Neurosci* 2018, **19**:758-770
Excellent review of the information sampling literature so far, in which the authors encourage the systematic and careful investigation of curiosity-driven exploration in large environments.
 2. Bellemare M, Srinivasan S, Ostrovski G, Schaul T, Saxton D, Munos R: **Unifying Count-based Exploration and Intrinsic Motivation.** *NIPS*; 2016.
 3. Ng AY, Harada D, Russell S: **Policy invariance under reward transformations: theory and application to reward shaping.** *Proceedings of the 16th International Conference on Machine Learning* 1999.
 4. Brafman RI, Tenenbholz M: **R-MAX - a general polynomial time algorithm for near-optimal reinforcement learning.** *IJCAI International Joint Conference on Artificial Intelligence* 2001.
 5. Ecoffet A, Huizinga J, Lehman J, Stanley KO, Clune J: **Go-explore: a new approach for hard-exploration problems.** *arXiv* 2019. preprint arXiv: 1901.10995
Games that RL agents used to perform badly at are performed well by Go-Explore, an algorithm that uses intrinsic rewards combined with the use of memory of promising states that have been visited before in order to explore from them. This was also the first algorithm to score any points on Pitfall, an Atari game considered to be a ‘hard exploration’ problem.
 6. Alet F, Schneider MF, Lozano-Perez T, Kaelbling LP: **Meta-learning curiosity algorithms.** *arXiv* 2020. preprint arXiv: 2003.05325.
 7. Badia AP, Sprechmann P, Vitvitskyi A, Guo D, Piot B, Kapturowski S, Tieleman O, Arjovsky M, Pritzel A, Bolt A *et al.*: **Never give up: learning directed exploration strategies.** *arXiv* 2020. preprint arXiv: 2002.06038.
 8. Burda Y, Edwards H, Pathak D, Storkey A, Darrell T, Efros AA: **Large-scale study of curiosity-driven learning.** *arXiv* 2018. preprint arXiv: 1808.04355.
 9. Ermolov A, Sebe N: **Latent world models for intrinsically motivated exploration.** *arXiv* 2020. preprint arXiv: 2010.02302.
 10. Raileanu R, Rocktäschel T: **RIDE: rewarding impact-driven exploration for procedurally-generated environments.** *arXiv* 2020. preprint arXiv: 2002.12292.
 11. Sekar R, Rybkin O, Daniilidis K, Abbeel P, Hafner D, Pathak D: **Planning to explore via self-supervised world models.** *arXiv* 2020. preprint arXiv: 2005.05960.
 12. Twomey KE, Westermann G: **Curiosity-based learning in infants: a neurocomputational approach.** *Dev Sci* 2018, **21**: e12629.
 13. Zheng Z, Oh J, Hessel M, Xu Z, Kroiss M, van Hasselt H, Silver D, Singh S: **What can learned intrinsic rewards capture?** *arXiv* 2020. preprint arXiv: 1912.05500
This paper shows that reinforcement learning agents can learn intrinsic rewards from data and pass them onto new lifetimes via a reward function.
 14. Singh S, Lewis RL, Barto AG: **Intrinsically motivated reinforcement learning: an evolutionary perspective.** *IEEE Trans Auton Ment Dev* 2010, **2**:70-82.
 15. Gopnik A: **Childhood as a solution to explore-exploit tensions.** *Phil Trans R Soc B* 2020, **375**: 20190502.

16. Kidd C, Piantadosi ST, Aslin RN: **The Goldilocks effects: human infants allocate attention to visual sequences that are neither too simple nor too complex.** *PLoS One* 2012, **7**:e36399.
17. Loewenstein G: **The psychology of curiosity: a review and reinterpretation.** *Psychol Bull* 1994, **116**:75-98.
18. Blanchard TC, Gershman SJ: **Pure correlates of exploration and exploitation in the human brain.** *Cogn Affect Behav Neurosci* 2018, **18**:117-126
- Here, information sampling and reward-maximizing choice are separated by introducing two separate actions human participants can choose between: 'observation' and 'betting'. BOLD signal in dACC and insula are related to choices by the participant to observe.
19. Schwartenbeck P, Passecker J, Hauser TU, FitzGerald TH, Kronbichler M, Friston KJ: **Computational mechanisms of curiosity and goal-directed exploration.** *eLife* 2019, **8**:e41703.
20. Kim K, Sano M, De Freitas J, Haber N, Yamins D: **Active world model learning with progress curiosity.** *arXiv* 2020. preprint arXiv: 2007.07853.
21. Rodriguez Cabrero JAM, Zhu J-Q, Ludvig EA: **Costly curiosity: people pay a price to resolve an uncertain gamble early.** *Behav Process* 2019, **160**:20-25
- This paper shows that people are willing to pay for information about upcoming rewards to be revealed earlier, an example of non-instrumental information sampling.
22. Wang MZ, Hayden BY: **Monkeys are curious about counterfactual outcomes.** *Cognition* 2019, **189**:1-10.
23. Kreps DM, Porteus EL: **Temporal resolution of uncertainty and dynamic choice theory.** *Econometrica* 1978, **46**:185.
24. Foley NC, Kelly SP, Mhatre H, Lopes M, Gottlieb J: **Parietal neurons encode expected gains in instrumental information.** *Proc Natl Acad Sci U S A* 2017, **114**:E3315-E3323.
25. Golman R, Loewenstein G: **Information gaps: a theory of preferences regarding the presence and absence of information.** *Decision* 2018, **5**:143-164.
26. Golman R, Loewenstein G, Molnar A, Saccardo S: *The Demand for, and Avoidance of Information.* 2020.
27. Kobayashi K, Ravaioli S, Baranès A, Woodford M, Gottlieb J: **Diverse motives for human curiosity.** *Nat Hum Behav* 2019, **3**:587-595
- This behavioural study with a large sample size ($n = 257$) shows that information sampling in a lottery task is driven by multiple factors, including the need to reduce uncertainty, option value and anticipatory utility. The extent to which each factor affected information sampling was seen to vary considerably across participants.
28. Cohen JD, McClure SM, Yu AJ: **Should I stay or should I go? How the human brain manages the trade-off between exploitation and exploration.** *Phil Trans R Soc B* 2007, **362**:933-942.
29. Kolling N, Behrens TEJ, Mars RB, Rushworth MFS: **Neural mechanisms of foraging.** *Science* 2012, **336**:95-98.
30. Daw ND, O'Doherty JP, Dayan P, Seymour B, Dolan RJ: **Cortical substrates for exploratory decisions in humans.** *Nature* 2006, **441**:876-879.
31. Domenech P, Rheims S, Koehlin E: **Neural mechanisms resolving exploitation- exploration dilemmas in the medial prefrontal cortex.** *Science* 2020, **369**:11
- In a human intracranial electrophysiology study, vmPFC activity was found to track the reliability of the current default action plan in anticipation of upcoming feedback, while dmPFC activity reflected the continuation or rejection of that action plan as a result of feedback. The authors propose a two-stage predictive coding mechanism whereby vmPFC signals the potential importance of upcoming information, while dmPFC drives behaviour in response to outcome information.
32. Tomov MS, Truong VQ, Hundia RA, Gershman SJ: **Dissociable neural correlates of uncertainty underlie different exploration strategies.** *Nat Commun* 2020, **11**:2371
- While this paper focuses on the effects of relative and total uncertainty of choice options on frontopolar and lateral PFC activity, an effect of relative uncertainty is also seen in a medial frontal region in this human fMRI study.
33. Trudel N, Scholl J, Klein-Flügge MC, Fouragnan E, Tankelevitch L, Wittmann MK, Rushworth MFS: **Polarity of uncertainty representation during exploration and exploitation in ventromedial prefrontal cortex.** *Nat Hum Behav* 2021, **5**:83-98 <http://dx.doi.org/10.1038/s41562-020-0929-3>
- Here, a dissociation is found between vmPFC and dACC, where vmPFC activity is linked to uncertainty prediction difference (the difference in uncertainty of the estimated values of the chosen and unchosen options) during throughout the task, while dACC activity is related to uncertainty only when the agent is in an exploratory phase.
34. Kaanders P, Nili H, O'Reilly JX, Hunt LT: **Medial frontal cortex activity predicts information sampling in economic choice.** *bioRxiv* 2020 <http://dx.doi.org/10.1101/2020.11.24.395814>. preprint bioRxiv: 2020.11.24.395814
- Using fMRI in human participants, we found an inverse value difference (value difference between the unchosen and chosen options) signal in medial frontal cortex (MFC) at the time of cue presentation, which was significantly reduced when information sampling was included as a regressor into this model. Instead, a main effect of information sampling (predicting whether participants would choose to sample more information before committing to one of the choice alternatives) was found in the same region. We propose that accounts describing MFC solely as an evidence accumulator in decision-making may be incomplete, as this region also drives sequential decisions to sample information before committing to a choice.
35. Hunt LT, Malalasekera WMN, de Berker AO, Miranda B, Farmer SF, Behrens TEJ, Kennerley SW: **Triple dissociation of attention and decision computations across prefrontal cortex.** *Nat Neurosci* 2018, **21**:1471-1481.
36. Kobayashi K, Hsu M: **Common neural code for reward and information value.** *Proc Natl Acad Sci U S A* 2019, **116**:13061-13066
- Information sampling in human participants is shown to be driven by both instrumental and non-instrumental motives, which put together the authors refer to as 'subjective value of information' (SVOI). SVOI is found to share a common neural code with reward value in striatum and vmPFC, while SVOI-related activation was also found in vmPFC.
37. Wang MZ, Hayden BY: **Curiosity is associated with enhanced tonic firing in dorsal anterior cingulate cortex.** *bioRxiv* 2020 <http://dx.doi.org/10.1101/2020.05.25.115139>. preprint bioRxiv: 2020.05.25.115139
- In this study, monkeys are shown to be willing to forgo a juice reward to receive counterfactual information about outcomes of choice options they did not pick and learn from this information.
38. White JK, Bromberg-Martin ES, Heilbronner SR, Zhang K, Pai J, Haber SN, Monosov IE: **A neural network for information seeking.** *Nat Commun* 2019, **10**:5168
- Activity in both dACC and basal ganglia neurons predict upcoming information gain in macaque single-cell recordings.
39. van Lieshout LLF, Vandenbroucke ARE, Müller NCJ, Cools R, de Lange FP: **Induction and relief of curiosity elicit parietal and frontal activity.** *J Neurosci* 2018, **38**:2579-2588
- In a human fMRI study, uncertainty-related parietal cortex activity was found at the time of curiosity induction (presentation of an uncertain stimulus), and activations in insula, OFC and parietal cortex were seen at the time of information revelation. Furthermore, residual curiosity-related activations (not related to outcome uncertainty) were found in pre-SMA.
40. Iigaya K, Hauser TU, Kurth-Nelson Z, O'Doherty JP, Dayan P, Dolan RJ: **The value of what's to come: neural mechanisms coupling prediction error and the utility of anticipation.** *Sci Adv* 2020, **6**:eaba3828
- BOLD activity in vmPFC is related to anticipatory utility, the value associated with knowing about upcoming rewards in advance, from an RL model in this human fMRI study.
41. Neubert F-X, Mars RB, Sallet J, Rushworth MFS: **Connectivity reveals relationship of brain areas for reward-guided learning and decision making in human and monkey frontal cortex.** *Proc Natl Acad Sci U S A* 2015, **112**:E2695-E2704.
42. Filimon F, Nelson JD, Sejnowski T.J, Sereno MI, Cottrell GW: **The ventral striatum dissociates information expectation, reward anticipation, and reward receipt.** *Proc Natl Acad Sci U S A* 2020, **117**:15200-15208
- This human fMRI study dissociates information expectation, information revelation and outcome revelation in a visual categorisation task, showing that information expectation is related to activations in ventral striatum

and STS, information revelation activates vmPFC and posterior cingulate, and outcome revelation activates frontal pole, posterior cingulate and OFC.

43. Charpentier CJ, Bromberg-Martin ES, Sharot T: **Valuation of knowledge and ignorance in mesolimbic reward circuitry.** *Proc Natl Acad Sci U S A* 2018, **115**:E7255-E7264

OFC and the mesolimbic reward system were found to activate in response to the opportunity to receive information, with the mesolimbic reward system being sensitive to the valence of outcomes in a human fMRI study using a lottery task.

44. Haber SN, Behrens TEJ: **The neural network underlying incentive-based learning: implications for interpreting circuit disruptions in psychiatric disorders.** *Neuron* 2014, **83**:1019-1039.
45. Iigaya K, Story GW, Kurth-Nelson Z, Dolan RJ, Dayan P: **The modulation of savouring by prediction error and its effects on choice.** *eLife* 2016, **5**:e13747.
46. Kakade S, Dayan P: **Dopamine: generalization and bonuses.** *Neural Netw* 2002, **15**:549-559.
47. Lau JKL, Ozono H, Kuratomi K, Komiya A, Murayama K: **Shared striatal activity in decisions to satisfy curiosity and hunger at the risk of electric shocks.** *Nat Hum Behav* 2020, **4**:531-543

Human participants showed willingness to enter a gamble to discover how a magic trick worked. Striatal activity was seen in fMRI measurements when participants entered such a gamble, and also when they entered a gamble for food rewards. This also shows that information is itself considered valuable by human agents.

48. Padoa-Schioppa C, Assad JA: **Neurons in the orbitofrontal cortex encode economic value.** *Nature* 2006, **441**:223-226.
49. O'Reilly JX, Schuffelgen U, Cuell SF, Behrens TEJ, Mars RB, Rushworth MFS: **Dissociable effects of surprise and model update in parietal and anterior cingulate cortex.** *Proc Natl Acad Sci U S A* 2013, **110**:E3660-E3669

Parietal cortex is found to be sensitive to how surprising an observation is, while dACC is active when an observation is used to update a belief in a saccadic planning task with human fMRI.

50. Kolling N, Wittmann MK, Behrens TEJ, Boorman ED, Mars RB, Rushworth MFS: **Value, search, persistence and model updating in anterior cingulate cortex.** *Nat Neurosci* 2016, **19**:1280-1285.
51. Hare TA, Schultz W, Camerer CF, O'Doherty JP, Rangel A: **Transformation of stimulus value signals into motor commands during simple choice.** *Proc Natl Acad Sci U S A* 2011, **108**:18120-18125.
52. Horan M, Daddaoua N, Gottlieb J: **Parietal neurons encode information sampling based on decision uncertainty.** *Nat Neurosci* 2019, **22**:1327-1335

Macaque single-cell recordings reveal that activity in LIP neurons represents the information gained by the upcoming saccade in a perceptual choice task.

53. Duan H, Fernández G, van Dongen E, Kohn N: **The effect of intrinsic and extrinsic motivation on memory formation: insight from behavioral and imaging study.** *Brain Struct Funct* 2020, **225**:1561-1574.
54. Molenberghs P, Mesulam MM, Peeters R, Vandenberghe RRC: **Remapping attentional priorities: differential contribution of superior parietal lobule and intraparietal sulcus.** *Cereb Cortex* 2007, **17**:2703-2712.
55. Isoda M, Hikosaka O: **Switching from automatic to controlled action by monkey medial frontal cortex.** *Nat Neurosci* 2007, **10**:240-248.
56. Karlsson MP, Tervo DGR, Karpova AY: **Network resets in medial prefrontal cortex mark the onset of behavioral uncertainty.** *Science* 2012, **338**:135-139.
57. Schuck NW, Gaschler R, Wenke D, Heinzle J, Frensch PA, Haynes J-D, Reverberi C: **Medial prefrontal cortex predicts internally driven strategy shifts.** *Neuron* 2015, **86**:331-340.
58. Akam T, Rodrigues-Vaz I, Marcelo I, Zhang X, Pereira M, Oliveira RF, Dayan P, Costa RM: **The anterior cingulate cortex predicts future states to mediate model-based action selection.** *Neuron* 2021, **109**:149-163.e7.
59. Gold JI, Shadlen MN: **The neural basis of decision making.** *Annu Rev Neurosci* 2007, **30**:535-574.
60. Glimcher PW, Fehr E: *Neuroeconomics: Decision Making and the Brain.* Elsevier/AP; 2014.
61. Spelke ES, Kinzler KD: **Core knowledge.** *Dev Sci* 2007, **10**:89-96.
62. Dubej R, Agrawal P, Pathak D, Griffiths TL, Efros AA: **Investigating human priors for playing video games.** *arXiv* 2018. preprint arXiv: 1802.10217
- Human agents are shown to be much worse at performing an Atari game when their prior knowledge is gradually removed by modifying the visual environment of the game. This shows that prior knowledge is likely a very important factor in learning new tasks.
63. Tsivavidis PA, Pouncy T, Xu JL, Tenenbaum JB, Gershman SJ: **Human learning in Atari.** *AAAI Spring Symposium Series, Science of Intelligence: Computational Principles of Natural and Artificial Intelligence, Technical Report SS-17-07.* 2017.
64. Niv Y: **Reinforcement learning in the brain.** *J Math Psychol* 2009, **53**:139-154.
65. Lake BM, Ullman TD, Tenenbaum JB, Gershman SJ: **Building machines that learn and think like people.** *Behav Brain Sci* 2017, **40**:e253.
66. Nair A, Pong V, Dalal M, Bahl S, Lin S, Levine S: **Visual reinforcement learning with imagined goals.** *arXiv* 2018. preprint arXiv: 1807.04742.
67. Wang J, Kurth-Nelson Z, Kumaran D, Tirumala D, Soyer H, Leibo JZ, Hassabis D, Botvinick M: **Prefrontal cortex as a meta-reinforcement learning system.** *Nat Neurosci* 2018, **21**:860-868.