

Opinion

Imagining the future self through
thought experimentsKentaro Miyamoto ,^{1,*} Matthew F.S. Rushworth,^{2,3} and Nicholas Shea^{4,5}

The ability of the mind to conceptualize what is not present is essential. It allows us to reason counterfactually about what might have happened had events unfolded differently or had another course of action been taken. It allows us to think about what might happen – to perform 'Gedankenexperimente' (thought experiments) – before we act. However, the cognitive and neural mechanisms mediating this ability are poorly understood. We suggest that the frontopolar cortex (FPC) keeps track of and evaluates alternative choices (what we might have done), whereas the anterior lateral prefrontal cortex (alPFC) compares simulations of possible future scenarios (what we might do) and evaluates their reward values. Together, these brain regions support the construction of suppositional scenarios.

Thought experiments in humans and non-human primates

Human insight and imagination depend, in part, on the ability to perform what Ernst Mach called 'Gedankenexperimente' (thought experiments), based on thinking about what consequences might follow from a course of action that has not yet been pursued or a counterfactual course of action that differs from the one actually taken [1,2]. The ability to think about possibilities even before they are realized allows us to conceive of novel paths to take and to consider why we might take them [3]. It is not only important when we are engaged in especially abstract thought processes – it is equally imperative in normal life. If we are to interact with the world in the most effective manner, we must be able to imagine those aspects of the world that are not currently being experienced but which we might encounter in the future [4]. By simulating future scenarios now, we are more likely to negotiate them successfully when we encounter them in the future. We construct these 'suppositional scenarios', that represent alternative options and future possibilities, in the service of learning new things about the world and working out what to do [5]. Our mind's ability to conceptualize and reason about what is not present is vital.

When we need to think prospectively and imagine the future, the ability to predict the outcome of sequential actions [6] is important for identifying the optimal course of action to take. In such situations, a multiplying range of states branch out as consideration moves each step further into the future. For example, initially taking either one choice or another may open up different sets of possible future choices (Box 1). In such situations, people do not only evaluate the future states that they might enter but, in addition, they estimate how quickly the expected values of these states might increase or decrease as the future sequence of decisions unfolds [6]. For effective learning of new skills, covert mental rehearsal of motor actions, rather than actually making overt physical actions, is sometimes beneficial [7]. However, the cognitive and neural mechanisms mediating these abilities are so far poorly understood.

The evaluation of real, overt choices is achieved by monitoring actual outcomes – successes and failures, or reward and error feedback. However, when simulated scenarios are evaluated, this

Highlights

Performing thought experiments – thinking about what consequences might follow from a course of action that has not yet been pursued – is an important element of mental life in both human and non-human primates.

Recent studies have found that both frontopolar cortex (FPC) and anterior lateral prefrontal cortex (alPFC) contribute to introspective evaluation of beliefs about events even when they are not directly observed.

FPC and alPFC are essential for tracking and evaluating counterfactual choices (what one might have done) and possible future choices (what one might do), respectively.

We propose a theoretical model in which the interplay between FPC and alPFC enables thought experiments.

¹Laboratory for Imagination and Executive Functions, RIKEN Center for Brain Science, Wako, Japan

²Department of Experimental Psychology, University of Oxford, Oxford, UK

³Wellcome Centre for Integrative Neuroimaging, University of Oxford, Oxford, UK

⁴Institute of Philosophy, School of Advanced Study, University of London, London, UK

⁵Faculty of Philosophy, University of Oxford, Oxford, UK

*Correspondence:
kentaro.miyamoto.wg@riken.jp
(K. Miyamoto).

Box 1. Prospective decision-making between exploring potential new alternatives and exploiting a currently available option

Before a person can decide to leave their current job to pursue new opportunities, it is important for them to consider many aspects of the job market such as the value of the alternative jobs that might come up, the variation in those values, and the maximum amount of time they can spend out of employment looking for a job. For example, it might be sensible to pursue new job opportunities if the average value of alternatives is high. However, care must be taken if variance in opportunities is also high because some of the job offers the job-seeker might encounter might be of much lower value than the mean. Moreover, variance matters little if the job-seeker can afford to carry on searching for a long time for the best opportunities in the current environment even after receiving a bad offer. It is, however, a problem if the job-seeker has only a narrow time horizon in which to find a new job (e.g., because of an impending rent deadline); this might mean that they cannot wait to encounter the best opportunities and they might need to settle for a worse offer. It is argued that, although a choice may have an immediate or myopic value, consideration of such additional factors means that choices also have longer-term prospective values linked to the potential future decisions that might arise once the current choice is taken [6]. Sometimes, people even avoid pursuing a sequential course of decisions if it is necessary to continually reassess prospective value at each step because of the rate at which prospective value changes as each new decision is taken. The degree to which new evaluations are undertaken and influence behavior can vary considerably across people in ways that may be linked to poor mental health [68]. Activity in IPFC and dorsal anterior cingulate cortex (dACC) reflects prospective value but, in addition, activity in area 9 reflects the changes in prospective value that will ensue when particular decisions are taken. It has been suggested that there may be two complementary exploratory systems: a directed exploration system that monitors goals that are alternatives to the current goal, and which has been linked especially to lateral FPC, and an undirected exploration system which has been linked to medial FPC. The directed exploration system is important when updating estimates of the values of alternative goals and deciding whether to switch to pursuing them [20,43,49]. By contrast, the medial system, by exerting more or less control over behavior, allows behavior to be more or less exploratory but in a more random and less directed manner. The interplay between FPC and aIPFC, especially in humans, for imagination and simulation of future counterfactual goals is complementary to these processes.

must be managed internally via comparison of probability and confidence estimates for the imagined outcomes associated with each scenario [8,9]. Related processes assist the socially mediated adoption of innovations; people do not simply copy the actions of others when they adopt a new innovation, they share simulations of the scenarios in which the innovation will be useful and they can do this even when they have different backgrounds or experiences [10]. So too for scientific discovery. Scientific reasoning is not merely a matter of logical inference. The ability to imagine, anticipate, visualize, and re-experience from memory is arguably crucial [11].

Even for non-human primates such as macaque monkeys, the ability to imagine events that are not actually occurring at a given time is important. Many monkeys range across extensive habitats in which food sources, such as fruiting trees, are sparsely distributed. However, because they are diurnal and have good vision, they can evaluate the foraging opportunities likely to be available in a different location, and compare them to current options before actually physically moving to a new location [12–15]. Similarly, in the social sphere, considering who to interact with and how, in a social group, could draw on the ability to imagine possible scenarios [16,17].

Although it is well established that mental imagery and perception of physical sensory inputs share at least partially overlapping neural bases [18,19], the neural mechanisms for simulating and evaluating the courses of action that one might take, or the scenarios one might explore, are poorly understood. There is evidence that the FPC and aIPFC enable introspective evaluation of beliefs about choices or events that have not actually occurred [20–24]. We argue that FPC and aIPFC play crucial but different roles in the process of constructing suppositional scenarios and generating a thought experiment.

Consider the day-to-day choices a person makes to earn a living, and how they think about the longer term. An online social influencer must decide what to post each day. Do they write about a current political controversy or do they post a cute cat picture? Having made a choice, to compare the results (likes, advertising revenue, social intrusion) against other options they might

have chosen, some system needs to keep track of the unchosen options and their likely reward values. By contrast, when thinking about what they would do if they were to stop being a social media influencer, the person needs to imagine what it would be like, for example, to open a new restaurant or to work as a taxi driver. We suggest that FPC and aIFPC play different but complementary roles in tracking and evaluating alternative options and prospective situations, respectively. Their interplay is crucial for the ability to conduct a thought experiment in which, through considering suppositional scenarios, a subject can evaluate situations they might encounter in the future.

FPC: keeping track of the road not taken

FPC (area 10) sits at the rostral end of the granular PFC. It is thought to be present only in primates, including humans and macaques [25]. There are, however, some differences in its organization in humans and macaques. Human FPC consists of more than one cytoarchitectonic region – different subdivisions of area 10 [14,26,27] – and these are associated with different anatomical networks of inputs and outputs [28,29]. It is not clear whether all the same subdivisions exist in the macaque. In the macaque, FPC is the origin of three long efferent association pathways to the anterior temporal pole, superior temporal gyrus, and retrosplenial cortex which run via the uncinate fascicle, the extreme capsule, and the cingulate fascicle, respectively [30]. These regions are involved in multisensory association [31], conceptual–semantic memory formation [32], and autobiographical memory retrieval [33], respectively.

A pioneering neuroimaging study using voxel-based morphometry in humans [34] demonstrated that, across participants, the gray matter volume of the lateral FPC is correlated with metacognitive efficiency in a perceptual decision-making task. The metacognitive processes that participants are engaged in during this task are similar to those needed to perform a Gedankenexperiment. To make a metacognitive report about whether they have taken the correct choice in the perceptual decision, participants think not only about the evidence for the choice they took but also the evidence for the alternative choice that they might have taken (Box 2 discusses the effect on confidence reports of the evidential strength of unchosen options).

Electrophysiological recordings from macaques have suggested that cell activity in the lateral FPC is not only coupled with behavior when behavior is not governed solely by external cues but also when a course of action must be held in memory [35,36]. In the study, the monkeys needed to retain the choice they had taken recently in working memory and decide whether to repeat it or switch to an alternative depending on which visual instruction they saw. Decision-selective activity is observed in FPC cells, but only during the feedback period when the animal learned whether the strategy that it had adopted on the current trial had been successful. The fact that animals are less vulnerable to task-irrelevant interruptions after FPC lesions [37] suggests that FPC does not determine the influence that external cues have on behavior, but instead that it is concerned with evaluating whether an internally generated alternative choice – perhaps a non-task choice – would be better to take than a task-instructed choice.

Box 2. Evaluation of confidence for chosen and unchosen evidence in a high-dimensional decision space

Experiments finding a 'positive evidence bias' suggest that confidence reports place more weight on the evidence in favor of the chosen option than on the evidential strength in favor of the alternative options [69,70]. Recent work suggests that this is not only a heuristic bias but reflects optimization for performance [71]. It may be due to weighing the relative evidence of a range of options in a high-dimensional decision space. Previous work relied on idealized, low-dimensional modeling frameworks such as signal detection theory or Bayesian inference. These frameworks may need revising to understand how decision confidence should be estimated when there is high-dimensional, naturalistic information. A deep neural network model optimized to assess decision confidence directly given high-dimensional inputs suggests that it is adaptive for decisions and confidence to depend on a common decision variable [71].

We recently developed a method to measure metacognition in macaques during memory retrieval using a post-decision wagering paradigm [38]. Monkeys are able to accurately predict their memory retrieval performance before receiving feedback. That they form accurate confidence estimates about their ability to perform a cognitive process even in the absence of feedback suggests that monkeys have insight into their internal cognitive states during decision-making. In this task, correct performance consisted of correctly reporting that a previously experienced stimulus had been seen before, or correctly reporting that a non-experienced stimulus was novel. We found that different areas are involved in metacognitive judgments about previously experienced and non-experienced stimuli. Whole-brain functional neuroimaging revealed that activity in FPC in response to previously unseen stimuli predicts, not recognition performance, but the accuracy of metacognitive evaluation, where accurate metacognition consists in preferentially betting high when correctly judging that the stimulus had indeed not been seen previously. It was more active when a correct rejection of a novel item was accompanied by high confidence or an incorrect judgment that a novel item had been seen before (i.e., a false alarm) was accompanied by low confidence [21] (Figure 1). Notably, the activation is focally situated on the dorsolateral surface of the frontal pole (area 10d) ([39,40] for anatomical demarcation within the FPC of macaque monkeys), and overlaps with the area in which neuronal activity in response to self-generated decisions was recorded [36]. FPC activation thus predicts accurate metacognition about judgments made about previously unseen stimuli. We argue that this is because these judgments require counterfactual thinking and recognition that the scenario under consideration is novel and does not match the actual past experience.

By contrast, we found that activity more posteriorly on the dorsolateral prefrontal surface (area 9l) predicts the accuracy of metacognitive evaluation of judgments made about previously experienced stimuli. Having judged that the current stimulus matches one of the previously presented stimuli, there is no need to go on to consider whether it would also match one of the other available options. This form of metacognition therefore does not call for comparison against counterfactual options in the way that metacognitive performance in relation to novel stimuli does. There may be anatomical specificity for metacognition in relation to different processes

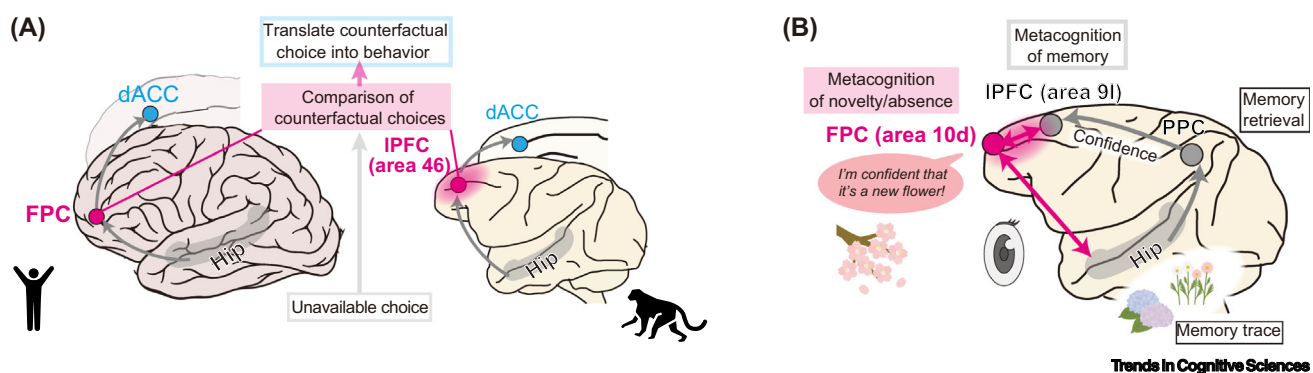


Figure 1. Frontopolar cortex (FPC) in metacognition and counterfactual decision-making. (A) Activity in human FPC (left) [23] reflects the value of switching to an alternative choice that is not currently being taken as opposed to the choice being taken. Activity in the macaque brain similarly reflects the relative value of an unchosen option as opposed to the chosen option, but now the region in which counterfactual choice values are represented is a broad swath of prefrontal cortex (PFC) including lateral PFC (IPFC; area 46) rather than a more specific FPC region as in humans [24]. Counterfactual choice-related activity in these prefrontal regions is related to the information held in other brain regions such as the hippocampus (Hip), in which activity covaries with the values of unavailable choices, and the dorsal anterior cingulate cortex (dACC), in which counterfactual choice information is translated into actual behavior. (B) Macaque FPC (area 10d) and IPFC (area 9l) contribute to confidence judgments about novel information and memorized information, respectively [21,38]. Functional connectivity between FPC and Hip during novelty recognition is important for confidence judgments about whether a stimulus is new (not previously stored in memory). By contrast, functional connectivity between IPFC and parietal cortex [38,72] during the reading out of memory confidence is important for metacognitive evaluation of the memory of a stimulus that has been previously experienced.

(memory or perception, etc.); in other words, there may be domain-specific metacognition [41,42]. Reversible silencing by microinjection of a GABA-A receptor agonist (muscimol) into the FPC (area 10d) impaired metacognitive performance for novel stimuli without impairing metacognitive performance for previously experienced stimuli or first-order task performance. These observations suggest that FPC plays a causally essential role when keeping track of and evaluating alternative options in the context of the current task. To return to the example of our social media influencer, the crucial cognitive process concerns keeping track of other topics they might have posted about today and evaluating them in the context of what they chose to do instead.

The role of FPC in maintaining information about alternative options has been confirmed during learning-based decision-making in both humans and macaques. In both cases, FPC codes the reward value of counterfactual choices – choices that the person or monkey is not taking on the current trial of the experiment but which they could have taken [23,24]. In contrast to the activity recorded in the human study, which is very focal and anterior in FPC [23], the activity recorded in monkeys is widely spread across the lateral PFC [24]. This may reflect the absence of functional counterparts of some human frontopolar subregions in macaque monkeys (as inferred from functional connectivity patterns) [29]. By contrast, other brain areas, such as ventromedial PFC, code the reward value of the choice currently being taken [23,24]. In a similar vein, it has also been proposed that FPC interposes an alternative behavioral plan instead of the current plan [43] and that it drives exploration by tracking the relative uncertainty present in the choice environment [44–48]. Flexible learning of new task-rules is impaired by FPC lesions [49]. One interpretation of this observation is that tracking of different potential choices, which is necessary for learning, is compromised by FPC lesions.

These findings suggest that, at least in humans, FPC contributes to the imagination of counterfactual scenarios in relation to oneself. In reference to historical discussion in philosophy and psychology, imagination is a suppositional mental state that allows the subject to consider situations other than the here and now [50]. When this imaginative process focuses on states that occurred in the past, it is highly related to episodic-like memory retrieval, an ability that animals may also possess [51]. Thus, it is possible that macaque monkeys also subjectively evaluate scenarios that are not currently present via FPC. Efferent connections from FPC to temporal and retrosplenial areas for mental imagery could reflect the role of FPC in switching focus among possible scenarios to generate a new strategy.

aIPFC: imagination for future possibilities relevant to oneself

So far we have been considering the evaluation of 'the road untaken': the imagination of counterfactual possibilities that run parallel to the actual option selected. Thought experiments also involve the imagination of possibilities that may be available in the future. Our social media influencer can think about what it would be like to do various other jobs if their online influence were to dry up. We all construct suppositional scenarios not only about what we might have done but also about what we might do in the future.

We recently developed a psychological paradigm to measure how people estimate the probability of their future success and compare it to estimates of the probabilities of external events happening in the environment [22] (Figure 2). Specifically, in the first stage of each trial – the metacognitive stage – participants were asked to choose between two motion-detection tasks each comprising a random-dot kinematogram (RDK) stimulus. In one task, success depended on how well participants could judge the average direction of motion of the dots, which could be easy or difficult depending on their coherence; in the other task, the direction was clear but reward was

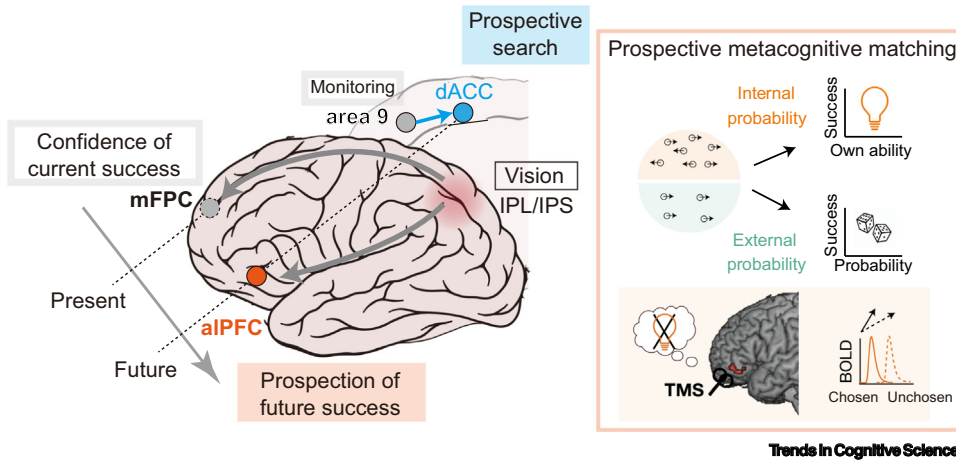


Figure 2. Role of anterior lateral prefrontal cortex (alPFC) for imagination for future possibilities that are relevant to oneself. Human alPFC contributes to prospective evaluation of the likelihood of future success through a process of metacognitive comparison of reward probabilities in relation to the environment (external probability) and one's own likely future performance (internal probability) [22]. Disruption of activity in alPFC by transcranial magnetic stimulation (TMS) changed metacognitive decisions about perceptual performance without changing perceptual performance itself. By contrast, the medial frontopolar cortex (mFPC) evaluates the decision-maker's confidence in their current performance (also [34]) as read out from perception and attention-related brain areas (e.g., inferior parietal lobule, IPL; and intraparietal sulcus, IPS). Activity in dorsal anterior cingulate cortex (dACC) reflects prospective value, but additionally activity in area 9 reflects the changes in prospective value that will ensue during the course of a series of actions when particular decisions are taken part way through the sequence (see also Box 1). Images are reproduced, with permission, from [22]. Abbreviation: BOLD, blood oxygen level dependent imaging signal.

only delivered probabilistically, with reward being more probable when there were more dots. Participants subsequently performed the task on a stimulus of the same type as the one they had selected at the first stage. They received a financial payout based on winning points at the end of the experiment. Thus, in the first stage of each trial, participants should choose between the two tasks on the basis of their predictions of reward probability for each task, weighing task difficulty in one task (internal probability) against probability of reward in the other (external probability).

Our fMRI experiments revealed activity in alPFC that increases in correlation with the internal probability both when the participants choose the internal probability task and when they reject it and choose the external probability task. This activity was in a region ventral to the anterior horizontal ramus at the boundary between areas 45 and 47 (alPFC; area 47/45). The time-course of the activity pattern in alPFC was predictive of whether the internal probability option would be chosen or not – the internal probability option was chosen when the alPFC signal rose more rapidly – suggesting a subjective evidence accumulation process estimating the internal probability. Disruption of neural activity in alPFC by transcranial magnetic stimulation (TMS; continuous theta-burst stimulation, cTBS) changed the way the decision at the start of the trial – the metacognitive decision between perceptual tasks varying in external probability or internal probability – was made. However, TMS had no impact on how subsequent decisions were made during the actual perceptual decision-making task. A related observation is that prospective evaluation skills are also altered by application of transcranial alternating current stimulation (tACS) over the same area [52]. In conjunction, these findings suggest that alPFC plays a key causal role in predicting future possibilities related to the actions that one might take oneself.

alPFC (area 47/45) is evolutionarily novel and well developed, especially in humans compared to non-human primates [53]. Cross-species comparisons of location in the whole-brain network

based on resting-state functional connectivity patterns failed to identify brain regions in macaque monkeys that were strongly homologous to the anterior part of area 47/12 [28]. However, other adjacent posterior ventrolateral frontal regions, such as area 45, are more similar across species. Another region near the orbital surface, area 47/12o, is also more clearly similar in humans and macaques. Area 47/12o has a role in credit assignment – linking an outcome to a choice based on past experience of whether a rewarding outcome is received in conjunction with the choice [54,55] – exactly as it does in humans [56–58]. We suggest that the anterior part of 47/12 may carry out a different but related process – estimating the outcome of an action but doing so based on an internal simulation rather than based on past experience of external events. In human fMRI studies, activity in this region is reported to encode subjective value under risk and ambiguity [59]. It may arbitrate the degree of control exerted over behavior by model-based and model-free systems as a function of the reliability of their respective predictions [60]. These observations suggest a role for alPFC in configuring a task model to predict future events relevant to oneself based on introspection about how one will be able to act – a crucial element of mental simulation. In relation to the hippocampus, a neural substrate for memory replay and offline simulation, stronger interaction with the inferior lateral PFC around alPFC has been linked to increasing demand for simulation planning [61]. A similar ventrolateral prefrontal area has been implicated in separating imagination from reality and in assessing the contribution made by bottom-up sensory signals from the occipital and temporal cortices to mental imagery [62]. Such findings also support the proposal that alPFC generates imagined scenarios during the prospective simulation and evaluation of choices that might be available in the future.

Interplay between FPC and alPFC to enable thought experiments

So far we have suggested that activity in the FPC (area 10) and alPFC (area 47/45) tracks information that might not be used for immediate decision-making but which may be important for preparing decision-makers – both human and non-human primates – for the decisions that they are likely to take in the future [22,24]. The activity patterns generated in these areas are metacognitive in nature and are, we argue, part of the neural foundation of exploratory behavior and prospective decision-making – the ability to imagine the decisions that a future self may need to take [6,48]. Similarly, metacognitive processes [8,38] are involved in the construction of novel and abstract conceptual schemes [63,64].

We suggest that the ability to imagine counterfactual scenarios and the actions that we might take in these scenarios is mediated by interactions between FPC, representing current unchosen options, and alPFC, representing prospectively available options (Figure 3). As a consequence, when new opportunities arise, we are able to make the most effective decisions because we have already been preparing what to do. The social media influencer can consider both the alternative topics that she could post about today and the alternative futures that different jobs would open up for her. The interplay between FPC and alPFC is proposed to contribute to switching between exploring new strategies and exploiting current strategies for ongoing decision-making [43]. In addition, our recent findings suggest that, although FPC may track and evaluate alternative options (what we might have done), alPFC may track and evaluate prospective future situations and opportunities (what we might do) [22]. Thus, interaction between these areas enables us to choose an optimal strategy, considering both currently available options and those that could be available in the future, before taking any actual physical action. As explained above, we found that alPFC compares the internal model (i.e., internal probability) of a potential choice with the external environment to generate prospective estimates of what the individual decision-maker, the self, might achieve in two potential future scenarios. That is, alPFC serves the function of evaluating prospective situations in relation to oneself. The strong functional connectivity between FPC and alPFC [29] suggests that currently available alternatives and prospective

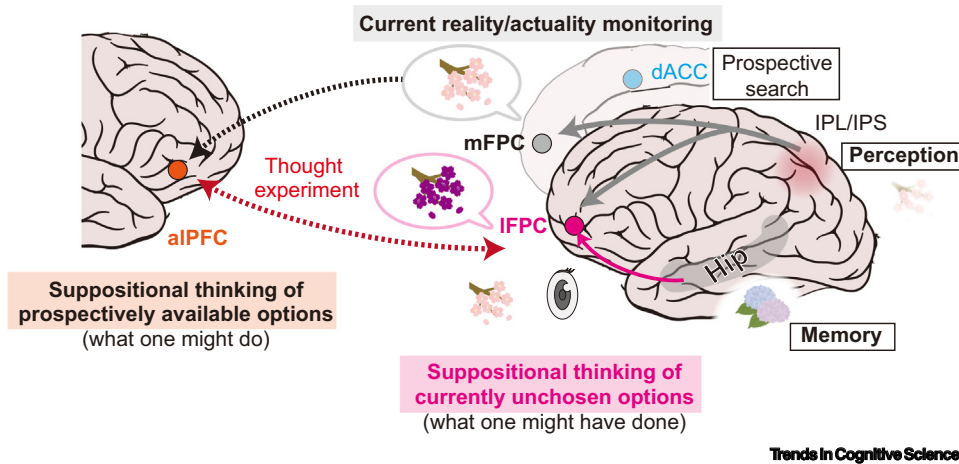


Figure 3. Proposed model of interplay between frontopolar cortex (FPC) and anterior lateral prefrontal cortex (alPFC) to enable thought experiments. We propose a model in which interaction between FPC and alPFC is the basis for imagining future possibilities for oneself (such as future actions that one might take) given one's assessment of what one might have done (lateral FPC, IFPC) and what one might do (alPFC). As a result, when opportunities arise in the future, the most effective decisions can be taken because they will have been considered in advance. In addition to simulation of counterfactual scenarios via IFPC and prospective metacognitive evaluation of choices that might be taken via alPFC, more medial FPC (mFPC) may monitor current confidence in the choices that are being taken. For example, when we see a flower, we might sometimes imagine an alternative – a suppositional flower at odds with the true reality of what is in front of us. Thinking about how the flower might have appeared in other circumstances (had there been more sun or more rain) might depend on reference to memory of other flowers in other situations and in IFPC. Thinking about how the flower might develop in the future (as its growth proceeds) might depend on alPFC. Abbreviations: dACC, dorsal anterior cingulate cortex; Hip, hippocampus; IPL, inferior parietal lobule; IPS, intraparietal sulcus.

scenarios can be evaluated jointly. Taken together, these functions give us the capacity to perform thought experiments.

The suppositional scenarios that are constructed and evaluated in prospective decision-making could be conscious or unconscious. However, there are reasons to expect that aspects of the process we have identified are reflected in subjective experience. Neural activity in FPC and alPFC is associated with the subjective experiences on which metacognitive introspection operates [65–67]. Neuronal correlates of explicit perceptual reports and specific perceptual content are consistently found in the anterior PFC [66,67]. These suggest that the interplay between FPC and alPFC especially in humans could be associated with the subjective experience of imagining future scenarios. The unusually strong preoccupation with imagining future scenarios that is present in humans may be a consequence of particularly prominent interactions between these two regions or the greater elaboration of the future-directed function of alPFC in humans. However, neuroimaging observations in macaques [21,24,38] suggest that the ability to imagine future possibilities may have first originated in other primates.

Concluding remarks

alPFC is suggested to play a role in the prospective evaluation of one's own potential future actions and their outcomes, which is an essential element of carrying out a 'thought experiment'. However, the absence of a completely similar region in other primates, such as monkeys, as suggested by the absence of a region with the same functional connectivity pattern, makes elucidation of its role difficult (see [Outstanding questions](#)). FPC is suggested to provide simulations of alternative scenarios, involving the self, in which the thought experiments might be conducted. This process may be linked to a circumscribed aspect of FPC in humans, a species in which

Outstanding questions

What neural mechanism enables the cognitive ability to mentally simulate alternative scenarios other than those that are immediately present?

How do differences in the development and organization of FPC across species relate to differences in the ability to simulate with suppositional scenarios? The FPC appears to be unique to primates and it is well developed in humans where it comprises several component parts. PFC areas around FPC also contribute to maintaining counterfactual information in some non-human primates, whereas human FPC appears to be more specialized for this role.

What is the evolutionary origin of the human talent to imagine future possibilities? Recent human studies suggested that the alPFC contributes to prospective evaluation of one's own levels of skill and one's own potential future actions. However, it is difficult, by comparing functional connectivity patterns, to identify a completely similar region in other primates such as monkeys. Does this mean that monkeys lack the ability to imagine potential future courses of action and to evaluate their likelihood of success? If not, then what neural mechanisms mediate these abilities in monkeys?

How do alPFC and FPC interact to achieve thought experiments? How do they generate possible and counterfactual scenarios and evaluate the plausibility of these scenarios before they have actually happened?

How do the neural mechanisms that are proposed to mediate thought experiments relate to the subjective experience of imagining future scenarios?

FPC is well developed, but other primates such as monkeys may employ a broader swath of PFC, including FPC but perhaps also in adjacent areas, to perform the same function. Comparisons of the neural mechanisms mediating these abilities in monkeys and humans may reveal how the skill of mental simulation, which is such a feature of human mental life, first evolved in primates. Investigation of the links and interactions between aIPFC and FPC, which enable thought experiments, is an important part of a new prospectus for the study of abstract and conceptual reasoning and the planning and generation of future actions.

Acknowledgments

This research was supported in part by Ministry of Education, Culture, Sports, Science, and Technology (MEXT)/Japan Society for the Promotion of Science (JSPS) KAKENHI grants JP22K18665 and JP22H00092, and by the Japan Agency for Medical Research and Development (AMED) under grant JP22wm0525034 to K.M., Wellcome Trust (221794/Z/20/Z) and Medical Research Council (MRC; MR/P024955/1) grants to M.F.S.R., and by the European Research Council (ERC) in the EU Horizon 2020 Research and Innovation Programme under grant agreement 681422 (MetCogCon) to N.S.

Declaration of interests

No interests are declared.

References

- Mach, E. (1906) *Erkenntnis und Irrtum: Skizzen zur Psychologie der Forschung*, Johann Ambrosius Barth
- Gendler, T.S. (2004) Thought experiments rethought – and re-perceived. *Philos. Sci.* 71, 1152–1163
- Grush, R. (2004) The emulation theory of representation: motor control, imagery, and perception. *Behav. Brain Sci.* 27, 377–396
- Aronowitz, S. and Lombrozo, T. (2020) Learning through simulation. *Philos. Imprint* 20, 1–18
- Shea, N. (2023) Concepts as plug & play devices. *Philos. Trans. Roy. Soc. Lond. B Biol. Sci.* 378, 202110353
- Kolling, N. *et al.* (2018) Prospection, perseverance, and insight in sequential behavior. *Neuron* 99, 1069–1082
- Vyas, S. *et al.* (2018) Neural population dynamics underlying motor learning transfer. *Neuron* 97, 1177–1186
- Shea, N. and Frith, C.D. (2019) The global workspace needs metacognition. *Trends Cogn. Sci.* 23, 560–571
- Miyamoto, K. *et al.* (2022) Conversion of concept-specific decision confidence into integrative introspection in primates. *Cell Rep.* 38, 110581
- Heyes, C. *et al.* (2020) Knowing ourselves together: the cultural origins of metacognition. *Trends Cogn. Sci.* 24, 349–362
- Nersessian, N.J. (2017) Cognitive science, mental modeling, and thought experiments. In *The Routledge Companion to Thought Experiment* (Stuart, M.T. *et al.*, eds), pp. 306–326, Routledge
- Ménard, N. *et al.* (2013) Effect of habitat quality on the ecological behaviour of a temperate-living primate: time-budget adjustments. *Primates* 54, 217–228
- Noser, R. and Byrne, R.W. (2007) Mental maps in chacma baboons (*Papio ursinus*): using inter-group encounters as a natural experiment. *Anim. Cogn.* 10, 331–340
- Klein-Flügge, M.C. *et al.* (2022) Medial and orbital frontal cortex in decision-making and flexible behavior. *Neuron* 110, 2743–2770
- Hunt, L.T. *et al.* (2021) Formalizing planning and information search in naturalistic decision-making. *Nat. Neurosci.* 24, 1051–1064
- Yizhar, O. and Levy, D.R. (2021) The social dilemma: prefrontal control of mammalian sociability. *Curr. Opin. Neurobiol.* 68, 67–75
- Milewski, T.M. *et al.* (2022) Behavioural and physiological plasticity in social hierarchies. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* 377, 20200443
- Kreiman, G. *et al.* (2000) Imagery neurons in the human brain. *Nature* 408, 357–361
- Kosslyn, S.M. *et al.* (2001) Neural foundations of imagery. *Nat. Rev. Neurosci.* 2, 635–642
- Mansouri, F.A. *et al.* (2017) Managing competing goals – a key role for the frontopolar cortex. *Nat. Rev. Neurosci.* 18, 645–657
- Miyamoto, K. *et al.* (2018) Reversible silencing of the frontopolar cortex selectively impairs metacognitive judgment on non-experience in primates. *Neuron* 97, 980–989
- Miyamoto, K. *et al.* (2021) Identification and disruption of a neural mechanism for accumulating prospective metacognitive information prior to decision-making. *Neuron* 109, 1396–1408
- Boorman, E.D. *et al.* (2009) How green is the grass on the other side? Frontopolar cortex and the evidence in favor of alternative courses of action. *Neuron* 62, 733–743
- Fouragnan, E.F. *et al.* (2019) The macaque anterior cingulate cortex translates counterfactual choice value into actual behavioral change. *Nat. Neurosci.* 22, 797–808
- Preuss, T.M. and Wise, S.P. (2022) Evolution of prefrontal cortex. *Neuropsychopharmacology* 47, 3–19
- Gilbert, S.J. *et al.* (2006) Functional specialization within rostral prefrontal cortex (area 10): a meta-analysis. *J. Cogn. Neurosci.* 18, 932–948
- Petrides, M. and Pandya, D.N. (1997) Comparative architectonic analysis of the human and the macaque frontal cortex. In *Handbook of Neuropsychology* (Boller, F. and Grafman, J., eds), pp. 17–58, Elsevier
- Neubert, F.X. *et al.* (2015) 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.* 112, E2695–E2704
- Neubert, F.X. *et al.* (2014) Comparison of human ventral frontal cortex areas for cognitive control and language with areas in monkey frontal cortex. *Neuron* 81, 700–713
- Petrides, M. and Pandya, D.N. (2007) Efferent association pathways from the rostral prefrontal cortex in the macaque monkey. *J. Neurosci.* 27, 11573–11586
- Blanke, O. (2012) Multisensory brain mechanisms of bodily self-consciousness. *Nat. Rev. Neurosci.* 13, 556–571
- Peelen, M.V. and Caramazza, A. (2012) Conceptual object representations in human anterior temporal cortex. *J. Neurosci.* 32, 15728–15736
- Vann, S.D. *et al.* (2009) What does the retrosplenial cortex do? *Nat. Rev. Neurosci.* 10, 792–802
- Fleming, S.M. *et al.* (2010) Relating introspective accuracy to individual differences in brain structure. *Science* 329, 1541–1543
- Mione, V. *et al.* (2019) Neural correlations underlying Self-generated decision in the frontal pole cortex during a cued strategy task. *Neuroscience* 404, 519–528
- Tsujimoto, S. *et al.* (2010) Evaluating self-generated decisions in frontal pole cortex of monkeys. *Nat. Neurosci.* 13, 120–126
- Mansouri, F.A. *et al.* (2015) Behavioral consequences of selective damage to frontal pole and posterior cingulate cortices. *Proc. Natl. Acad. Sci. U. S. A.* 112, E3940–E3949

38. Miyamoto, K. *et al.* (2017) Causal neural network of metamemory for retrospection in primates. *Science* 355, 188–193
39. Carmichael, S.T. and Price, J.L. (1994) Architectonic subdivision of the orbital and medial prefrontal cortex in the macaque monkey. *J. Comp. Neurol.* 346, 366–402
40. Rapan, L. *et al.* (2022) Cytoarchitectonic, receptor distribution and functional connectivity analyses of the macaque frontal lobe. *bioRxiv* Published online June 29, 2022. <https://doi.org/10.1101/2022.06.27.497757>
41. Fleming, S.M. *et al.* (2014) Domain-specific impairment in metacognitive accuracy following anterior prefrontal lesions. *Brain* 137, 2811–2822
42. Morales, J. *et al.* (2018) Domain-general and domain-specific patterns of activity supporting metacognition in human prefrontal cortex. *J. Neurosci.* 38, 3534–3546
43. Koechlin, E. and Hyafil, A. (2007) Anterior prefrontal function and the limits of human decision-making. *Science* 318, 594–598
44. Hogeveen, J. *et al.* (2022) The neurocomputational bases of explore-exploit decision-making. *Neuron* 110, 1869–1879
45. Cockburn, J. *et al.* (2022) Novelty and uncertainty regulate the balance between exploration and exploitation through distinct mechanisms in the human brain. *Neuron* 110, 2691–2702
46. Zajkowski, W.K. *et al.* (2017) A causal role for right frontopolar cortex in directed, but not random, exploration. *eLife* 6, 27430
47. Badre, D. *et al.* (2012) Rostrolateral prefrontal cortex and individual differences in uncertainty-driven exploration. *Neuron* 73, 595–607
48. Trudel, N. *et al.* (2021) Polarity of uncertainty representation during exploration and exploitation in ventromedial prefrontal cortex. *Nat. Hum. Behav.* 5, 83–98
49. Boschini, E.A. *et al.* (2015) Essential functions of primate frontopolar cortex in cognition. *Proc. Natl. Acad. Sci. U. S. A.* 112, E1020–E1027
50. Pearson, J. (2019) The human imagination: the cognitive neuroscience of visual mental imagery. *Nat. Rev. Neurosci.* 20, 624–634
51. Crystal, J.D. (2010) Episodic-like memory in animals. *Behav. Brain Res.* 215, 235–243
52. Soutschek, A. *et al.* (2021) Frontopolar theta oscillations link metacognition with prospective decision making. *Nat. Commun.* 12, 3943
53. Hayashi, T. *et al.* (2021) The nonhuman primate neuroimaging and neuroanatomy project. *Neuroimage* 229, 117726
54. Chau, B.K. *et al.* (2015) Contrasting roles for orbitofrontal cortex and amygdala in credit assignment and learning in macaques. *Neuron* 87, 1106–1111
55. Folloni, D. *et al.* (2021) Ultrasound modulation of macaque prefrontal cortex selectively alters credit assignment-related activity and behavior. *Sci. Adv.* 7, eabg7700
56. Noonan, M.P. *et al.* (2011) Distinct roles of three frontal cortical areas in reward-guided behavior. *J. Neurosci.* 31, 14399–14412
57. Boorman, E.D. *et al.* (2016) Two anatomically and computationally distinct learning signals predict changes to stimulus–outcome associations in hippocampus. *Neuron* 89, 1343–1354
58. Jocham, G. *et al.* (2016) Reward-guided learning with and without causal attribution. *Neuron* 90, 177–190
59. Levy, I. *et al.* (2010) Neural representation of subjective value under risk and ambiguity. *J. Neurophysiol.* 103, 1036–1047
60. Lee, S.W. *et al.* (2014) Neural computations underlying arbitration between model-based and model-free learning. *Neuron* 81, 687–699
61. Javadi, A.H. *et al.* (2017) Hippocampal and prefrontal processing of network topology to simulate the future. *Nat. Commun.* 8, 14652
62. Dijkstra, N. *et al.* (2022) Perceptual reality monitoring: neural mechanisms dissociating imagination from reality. *Neurosci. Biobehav. Rev.* 135, 104557
63. Bongioanni, A. *et al.* (2021) Activation and disruption of a neural mechanism for novel choice in monkeys. *Nature* 591, 270–274
64. Shea, N. (2018) Metacognition and abstract concepts. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 373, 20170133
65. Dehaene, S. *et al.* (2017) What is consciousness, and could machines have it? *Science* 358, 486–492
66. Odegaard, B. *et al.* (2017) Should a few null findings falsify prefrontal theories of conscious perception? *J. Neurosci.* 37, 9593–9602
67. Boly, M. *et al.* (2017) Are the neural correlates of consciousness in the front or in the back of the cerebral cortex? Clinical and neuroimaging evidence. *J. Neurosci.* 37, 9603–9613
68. Scholl, J. *et al.* (2022) The effect of apathy and compulsivity on planning and stopping in sequential decision-making. *PLoS Biol.* 20, e3001566
69. Zylberberg, A. *et al.* (2012) The construction of confidence in a perceptual decision. *Front. Integr. Neurosci.* 6, 79
70. Peters, M.A.K. *et al.* (2017) Perceptual confidence neglects decision-incongruent evidence in the brain. *Nat. Hum. Behav.* 1, 0139
71. Webb, T.W. *et al.* (2021) Performance-optimized neural networks as an explanatory framework for decision confidence. *bioRxiv* Published online November 2, 2022. <https://doi.org/10.1101/2021.09.28.462081>
72. Miyamoto, K. *et al.* (2013) Functional differentiation of memory retrieval network in macaque posterior parietal cortex. *Neuron* 77, 787–799