Please cite this article in press as: Phillips et al., How We Know What Not To Think, Trends in Cognitive Sciences (2019), https://doi.org/10.1016/j.tics.2019.09.007

Trends in Cognitive Sciences

CellPress REVIEWS

Opinion

How We Know What Not To Think

Jonathan Phillips,[1,*] Adam Morris,[2] and Fiery Cushman[2]

Humans often represent and reason about unrealized possible actions – the vast infinity of things that were not (or have not yet been) chosen. This capacity is central to the most impressive of human abilities: causal reasoning, planning, linguistic communication, moral judgment, etc. Nevertheless, how do we select possible actions that are worth considering from the infinity of unrealized actions that are better left ignored? We review research across the cognitive sciences, and find that the possible actions considered by default are those that are both likely to occur and generally valuable. We then offer a unified theory of why. We propose that (i) across diverse cognitive tasks, the possible actions we consider are biased towards those of general practical utility, and (ii) a plausible primary function for this mechanism resides in decision making.

Highlights

Representations of possible actions pervade human high-level cognition, and shape how we plan, attribute causal responsibility, comprehend language, and make moral judgments.

There are too many 'possible actions' for us to consider them all. Recent studies offer a strikingly convergent picture of how we call to mind a limited, useful set of possible actions to consider.

This process of 'sampling' alternative possibilities has a distinctive fingerprint: it focuses on possible actions that are valuable and probable. This fingerprint arises across many diverse tasks that rely on the representation of alternative possibilities.

We provide a novel theoretical proposal that helps to explain this convergence: by default, the possibilities that come to mind are those worth considering during decision making.

The Promise of, and Problem with, Possibilities

Humans can only experience the world as it is, but our most impressive powers of thought require us to imagine how it could be. To choose a city to move to, we imagine different things we might do there. When judging a person's actions, we ask if they could have done otherwise. To assign responsibility for a tragedy, we consider what might have been done to prevent it. When reading that a job candidate is 'punctual', we consider what his recommender might have emphasized instead. In each case, our thoughts are shaped by the alternative possibilities we consider.

This ability is as mysterious as it is pervasive. In theory, there are infinite alternative possibilities we can imagine, and we could therefore spend infinite time generating and considering them. In practice, however, we evaluate alternative possibilities so quickly and effortlessly that the evaluation often goes unnoticed. We must have some rapid and unconscious ability to generate a small set of nonactual possibilities that merit consideration, while excluding the infinity of useless others. How do we know what not to think?

We answer this question at the computational level, as Marr [1] defined the term. Synthesizing recent insights from diverse fields of study, we offer a model for how humans identify one type of possibility: possible actions. These are a crucial part of **modal cognition** (see Glossary) – our general ability to reason about all types of alternative possibilities, including actions as well as events, states of affairs, etc. When humans engage in tasks that require modal cognition (e.g., causal reasoning or responsibility attribution), they sample only a small number of specific possible actions. Moreover, across diverse tasks, people tend to sample possible actions sharing two properties: they are probable (i.e., they occur often) and valuable (i.e., they are usually good). We propose that people's sampling strategy has an adaptive origin: it helps people to efficiently make effective decisions in their own life. We call this the shared adaptive sampling model of modal cognition – it is shared across domains, adaptively organized, and allows us to sample effectively from the vast space of possible actions.

We conclude by exploring the proposed origin of this shared mechanism: decision making. Humans consistently face decisions involving enormous numbers of options (every potential job, next sentence, afternoon activity, etc.). Instead of exhaustively considering all options, people sample a small subset of possible actions for evaluation. Crucially, the possible actions they consider by default are skewed towards actions which are valuable and likely to be chosen. This feature makes perfect sense when trying to efficiently identify effective actions. Hence, we suggest that the shared sampling mechanism essential for modal cognition – one biased towards valuable and probable actions – was principally designed for decision making. In short, our most basic sense of which actions are possible reflects a first draft of what we would choose. Before laying out our proposal in more detail, we review previous work on reasoning about possibilities.

[1]Program in Cognitive Science, Dartmouth College, 201 Reed Hall, Hanover, NH 03755, USA

[2]Department of Psychology, Harvard University, 33 Kirkland Street, Cambridge, MA 02138, USA

*Correspondence: jonathan.s.phillips@dartmouth.edu

Please cite this article in press as: Phillips et al., How We Know What Not To Think, Trends in Cognitive Sciences (2019), https://doi.org/10.1016/j.tics.2019.09.007

Traditional Approaches to Reasoning about Possibilities

There is a strong tradition of research on possibilities across the cognitive sciences. Within philosophy and linguistics, it is common to model possibilities by defining a space of **possible worlds** and then partitioning it in various ways. For instance, we might partition possible worlds by whether they are consistent with what is known (defining what 'might be') or by whether they involve actions an agent is physically capable of performing (defining what she 'could do'). Once the relevant part of the partition is selected (e.g., the part corresponding to what 'might be'), traditional approaches often induce **orderings** of the possibilities within the relevant part of the partition (e.g., from most to least probable). The relevant dimension of ordering will depend on the task at hand [2–9]. It could involve ordering possibilities according to their probability, morality, accessibility, etc. The most highly ranked possibilities are typically those most relevant to the modal task faced. Because this tradition has typically not sought to model the psychological processes involved in modal cognition, these models have had relatively little to say, however, about which possibilities come consciously to mind from within the potentially vast space delimited by the partition.

Within psychology, modal thought has been studied as an essential part of many cognitive domains. Past research includes the role of modal cognition in 'mental models' that are used in reasoning [10–12], explicit counterfactual thought (reviewed in [13]), reflective assessments of which events are 'possible' (e.g., [14,15]), and, recently, the brain networks that support episodic simulation of nonactual future or past events (e.g., [16,17]). Like philosophers, psychologists assume we have some way to define the relevant types of possibilities we are interested in, which can again be modeled as inducing a partition followed by ordering of possibilities within the relevant part of the partition. By partitioning possibilities at the outset, we might, for example, typically not consider possibilities where the premises of an argument are violated when reasoning with mental models [10] or engaging in counterfactual reasoning [18], or follow a given syntactic structure when generating new names for pubs [19]. Further, by ordering possibilities and then focusing on the 'best' of that ordering, we might, for example, focus on morally better counterfactual alternatives when making moral judgments [20], or consider relatively probable alternative possibilities when making causal judgments [21].

Abstracting away from the details of any one particular approach, we can think of all of these models as proposing that we generate the 'alternative possibilities' fundamental to modal cognition by (i) delimiting a task-relevant partition within the vast space of conceivable possibilities, (ii) considering a smaller subset of particular possibilities within the relevant part of that partition, and then (ii) ordering or evaluating them in task-relevant ways to inform our final modal judgments (Figure 1A, Key Figure). Notably, there has been much research indicating how we may construct task-specific partitions and orderings, but there is relatively little work addressing how, before explicit evaluation, specific options are sampled from within the relevant part of the partition. This understudied part of the architecture is our focus.

Possible Actions: A Shared Adaptive Sampling Proposal

Unlike much of this previous work, our proposal focuses specifically on representations of possible actions, a particularly important type of possibility representation that plays a crucial role in high-level cognition. We build on the groundwork laid by traditional approaches in two key ways. First, our proposal gives crucial new structure to 'step 2' – the process by which particular possibilities within the relevant part of the partition are generated. Such spaces often include vast numbers of potentially relevant actions (e.g., 'possible ways to get to an airport', 'gifts one could buy for Christmas', 'ways to spend an afternoon', etc.), but ordinarily we only have the time and resources to explicitly evaluate a few. We propose that our minds adaptively sample a small set of possible actions from within the relevant space: those of probable practical utility are prioritized for consideration.

Second, although the traditional model assumes that the important work (partitioning and ordering) is task-specific (Figure 1B), our model proposes that a key intermediate step (sampling) is not. In principle, different adaptive sampling procedures might be used in different tasks (i.e., one for deciding what people should do, another for predicting their actions, and so on; Figure 1C); however, we

Glossary

Modal auxiliaries: in natural language, terms such as 'must', 'can', 'may', 'ought', etc. are called modal auxiliaries. They are used to make statements about nonactual possibilities and vary in primarily two ways: (i) the 'force' of the modal – whether they suggest that something is necessary (e.g., 'must') or merely possible (e.g., 'can'), and (ii) the 'flavor' of the modal – whether they concern what is good or right, and thus have a 'deontic flavor' (e.g., 'ought'), or what is known, and thus have an 'epistemic flavor' (e.g., 'might'), or some other flavor instead [6].
Modal cognition/modality: modal concepts allow us to represent and reason about sets of nonactual possibilities, for instance by grouping various 'possible worlds' in different ways. Modal statements often have to do with necessity or possibility: which things must be the case and which things could be, which are typically taken to correspond to the possibility and necessity operators in modal logic [115].
Ordering: possible worlds can be 'ordered'. For instance, some are considered to be 'closer' to the real world (e.g., worlds with no newspapers) and others are 'farther' (e.g., worlds with no Earth). Ordering is useful because it allows us to define sets based on points in the resulting order. For instance, some threshold on a probability ordering might separate 'relevant' worlds from irrelevant ones. Possibilities can be ordered in many different ways. We may, for example, order on 'ethicality': worlds with murder are worse than worlds without it, and various points along this ordering might define the set of worlds judged to be 'impermissible', 'permissible', or 'best'.
Possible worlds: philosophers and linguists often describe alternative possibilities in terms of 'possible worlds'. In the actual world – the one we are in – everything is a particular way. However, we can consider other ways the world might be. Each of these is a 'possible world' – for instance, the world in which everything is the same, except that you are reading this paper