CellPress REVIEWS

Opinion

People as Intuitive Scientists: Reconsidering Statistical Explanations of Decision Making

Aba Szollosi [1,*] and Ben R. Newell [1]

A persistent metaphor in decision-making research casts people as intuitive statisticians. Popular explanations based on this metaphor assume that the way in which people represent the environment is specified and fixed *a priori*. A major flaw in this account is that it is not clear how people know what aspects of an environment are important, how to interpret those aspects, and how to make decisions based on them. We suggest a theoretical reorientation away from assuming people's representations towards a focus on explaining how people themselves specify what is important to represent. This perspective casts decision makers as intuitive scientists able to flexibly construct, modify, and replace the representations of the decision problems they face.

Highlights

Theories of decision making often implicitly or otherwise assume that people's behavior is based on a highly specific repertoire of representations of the environment.

Such accounts rarely explain how the aspects of the environment the representations designate as important are specified by people.

We suggest that such specification requires people to be capable of detecting and correcting any potential errors in their representations.

We describe problems resulting from the lack of consideration of this aspect of cognition in theories of decision making and discuss potential solutions and future directions.

Mismatches of Scientists' Expectations and People's Behavior

People's interpretation of and responses to experiments are often not what scientists expect them to be. People identify patterns in randomly generated environments [1–4], detect nonexistent associations between unrelated events [5–7], and seek non-relevant information to guide their decisions [8–10]. Experimental findings demonstrating such discrepancies between scientists' expectations and people's behavior are numerous – how can such mismatches be resolved?

Multiple lines of enquiry in decision-making research have attempted to answer this question. A prominent one of these can be roughly characterized as based on rational mathematical models such as expected utility and probability theory [11–16]. Such theories suggest that mismatches are the results of systematic and unsystematic errors in the application of these models. A traditionally different line of research, based on a more explicit consideration of people's cognitive and natural environmental constraints, follows from the bounded rationality approach [17–21]. Under this view, mismatches result from people's use of heuristics: these tools can sometimes be too crude, people sometimes misapply, or are just unaware of how to use them correctly. Although these research traditions are often considered to be distinct and to imply qualitatively different ways in which mismatches can emerge, we will argue that both of them are based on the metaphor that people are **intuitive statisticians** (see Glossary). Human choice, in this sense, reduces to the application of statistical models to environments.

Advantageous decision making depends on the accuracy of people's (mental) **representations** [22] of how their environment behaves. The difficulty of developing representations of this sort is known as the **correspondence problem** [23–25]. In the current paper we argue that the fundamental flaw in theories that rely on the intuitive statistician metaphor is that the correspondence problem is solved in advance by the researcher. Drawing on the research of related fields [26–29], we outline a more fitting metaphor: one that casts people as **intuitive scientists**. In this view, people act like scientists to understand and control the environment not just by applying, but also by attempting to improve their representations. By considering how people themselves

[1]School of Psychology, University of New South Wales, Sydney, Australia

*Correspondence: aba.szollosi@gmail.com (A. Szollosi).

 https://doi.org/10.1016/j.tics.2020.09.005

CellPress REVIEWS

address the correspondence problem – that is, how they specify what to consider important in an environment, how to interpret those aspects, and how to use them in their decisions – this metaphor can help in providing a unified explanation for (i) the mismatches of people's expected and observed behaviors, (ii) people's capability to flexibly represent decision environments, and (iii) people's capability to improve their representations. Additionally, we will consider the consequences of this view for theories of rational choice and its implications for future experimental work on decision making.

Intuitive Statisticians and the Correspondence Problem

Perhaps best embodied in Peterson and Beach's metaphor [30], the idea that people are intuitive statisticians has had a lasting influence on theories of decision making. This metaphor unites rational and bounded rational research traditions through the (often implicit) assumption that, to address the correspondence problem, people rely on a fixed repertoire of possible representations of potential environments, which encapsulate all of the important aspects of those environments (Box 1). The assumed repertoire of representations is routinely based on a rational or bounded rational **benchmark**, which is essentially the researcher's own solution to the correspondence problem. Such representations can be likened to statistical models, because, in many applications of statistical models, what the variables refer to and how they are related to each other are highly specific and fixed *a priori*.

The major problem with this view is that defining what is important in any environment is not trivial [31–35]. For example, variables in statistical models need to refer to regularities in the environment, and there are an infinite number of ways a model can refer to these. Generating such connections between statistical models and the environment always requires a chain of reasoning; it needs to

Box 1. Benchmark-Based Solutions to the Correspondence Problem

Intuitive statistician theories base the way in which people represent the environment on researchers' benchmark solutions to the correspondence problem. Explanations of people's choices in now-classic base-rate neglect tasks [84,85] provide a good example of such theorizing. In the paradigmatic task, participants were asked to assess the probability of a person belonging to a particular group: they were given a personality description of a hypothetical person (consisting of e.g., 'married', 'conservative', and 'hobbies include mathematical puzzles'), and information about the base rate of group membership (e.g., 70 lawyers, 30 engineers) in the particular population from which the individual was sampled.

When solving the correspondence problem for the task, researchers often focus on the information they intended the participant to find important. In the rational tradition, a common starting point is to construct a 'rational' benchmark representation of the task based on this information. A typical representation would be to introduce the base rate information into a Bayesian equation along with a form of biasing influence. For example, an explanation in the Heuristics and Biases tradition [14] would assume that people designate the base rates as important, but their use of the simplifying representativeness heuristic makes them assign more weight to the particular group of which the personality description is representative. Note that such a benchmark provides a fixed specification of the aspects of the task the researcher wanted the participant to find interesting (the base rates and the personality descriptions), while ignoring every other aspect of the physical environment.

Although theories in the bounded rational tradition would start by specifying a more detailed benchmark that takes people's cognitive and environmental constraints into account, they take a similar approach in specifying what people find interesting. For example, in the fast-and-frugal heuristics approach [19], the benchmark would include assumptions of what kind of information format is easier for people to process – in this case that the base-rate information is easier to process in a frequency format than in a probability format [86]. Note that this is still a fixed formulation of the decision environment: the designated aspects of interest are now extended by the format of the presented information, but they are still highly specific.

These examples demonstrate that, although exactly what is designated as important (and the basis for that choice) differs between research traditions, they are alike in that they make this choice for people. That is, how the environment is represented by the participant is introduced as a fixed set of assumptions based on either the researcher's own solution to the correspondence problem or on the researcher's idea of how the participant solved the correspondence problem. Such specifications ignore the multitude of alternative ways in which the same environment can be represented [31,32,35,85] (see Figure 1 in main text).

Glossary

Benchmark: a fixed statistical model of the environment that an idealized version of a person could apply to maximize potential gains. It is typically constructed either in line with the best representation the experimenter could think of, or in line with the best representation the experimenter thought the participant could think of.

Correspondence problem: the difficulty of generating representations that map accurately onto the environment. The problem is always present for decision making under uncertainty (and thus virtually in all real-life circumstances).

Flexible error correction: processes aimed at the detection and elimination of potential errors in any aspect of representations. Necessarily has to imply the ability to generate new representations, including ways to modify existing representations.

Intuitive scientist: a metaphor that considers people capable of flexibly improving their representations, besides applying them. It entails capacity for detecting and correcting errors in any aspect of their representations.

Intuitive statistician: a metaphor that (at least implicitly) considers people's representations highly specific, and their choice as the inflexible application of these representations.

Representation: mental abstractions of external environments that are isomorphic to those environments in certain ways. They also allow for the simulation of counterfactuals, which can aid people in making advantageous decisions.

be carefully considered which environmental regularities are to be expressed in variables and error distributions, and how. Importantly, for any type of representation, these considerations need to be addressed by the person; yet, the way people would attempt to solve these problems is often entirely ignored in intuitive statistician models of decision making.

The problem is reflected clearly in people's capability to represent the same environment in infinitely many potential ways [31]. An example of the potential multiplicity of representations that people can have of a repeated-choice gambling experiment is depicted in Figure 1 (Figure 1, Key Figure). Even such simple experimental environments can be represented across infinitely many different dimensions: for example, in terms of long-run frequencies of the outcomes (Figure 1B); in terms of the temporal structure of the outcomes (Figure 1C); or even in terms of the computer software that generates them (Figure 1D). Some of these representations may be more accessible to the participants than others (e.g., given their background knowledge, their level of interest, etc.), yet people's capability to represent the environment in these (and an infinite number of more) fundamentally different ways is evident. This capability is not explained in intuitive statistician theories of decision making – instead a collection of potential representations is introduced as *a priori* assumptions.

Flexibility in Intuitive Statistician Theories

Substituting the explanation of how people would solve the correspondence problem with the researcher's own specific solution grants a large degree of flexibility to intuitive statistician theories. This is evidenced by how such theories are advanced to explain mismatches. In particular, whenever mismatches arise between what is expected on the basis of the researcher's benchmark solution (Figure 1 and Box 1) and the participants' observed behavior, the fixed benchmark (and therefore people's assumed representation) is simply updated to remain consistent with the experimenter's changed understanding of the environment. This form of flexible theory change sidesteps how people address the correspondence problem.

The flexibility of this theory-change regime can be seen in that people can always be assumed to weight variables differently [13,36,37], in that their choices can always be subject to various different sources of noise [15,38,39], or in the possibility of adding a new heuristic to people's toolbox that designate different aspects of the environment as important [19,40–42]. Any of these and similar *ad hoc* assumptions can be freely deployed to salvage the central assumption that the participant's representation is largely similar to that of the researcher, and to easily explain away any mismatch between expected and observed behavior.

Moreover, the possibility of such *ad hoc* assumptions allows researchers to assume arbitrary complexity and specificity of the knowledge captured in the representations. This can range from the assumption of people having almost perfect knowledge of and behaving almost fully in accordance with probability theory [15], to them deciding only on a whim [43]. What knowledge is assumed to be captured in the representations can then give rise to conflicting interpretations of the same observations: people's behavior can be considered irrational [14,44], rational [45,46], or boundedly rational [17,47–49].

Although on the surface rational and bounded rational approaches appear different, they both exploit the same type of flexibility resulting from a reliance on fixed benchmark solutions to the correspondence problem. That is, instead of explaining how people might solve the correspondence problem, these theories start at the point where the experimenter has already declared that problem solved. In addition, the researcher's ability to perpetually update their benchmark solutions in the face of conflicting observations limits the usefulness of experimental testing. If

CellPress REVIEWS

Key Figure

Possible Representations of Repeated-Choice Tasks

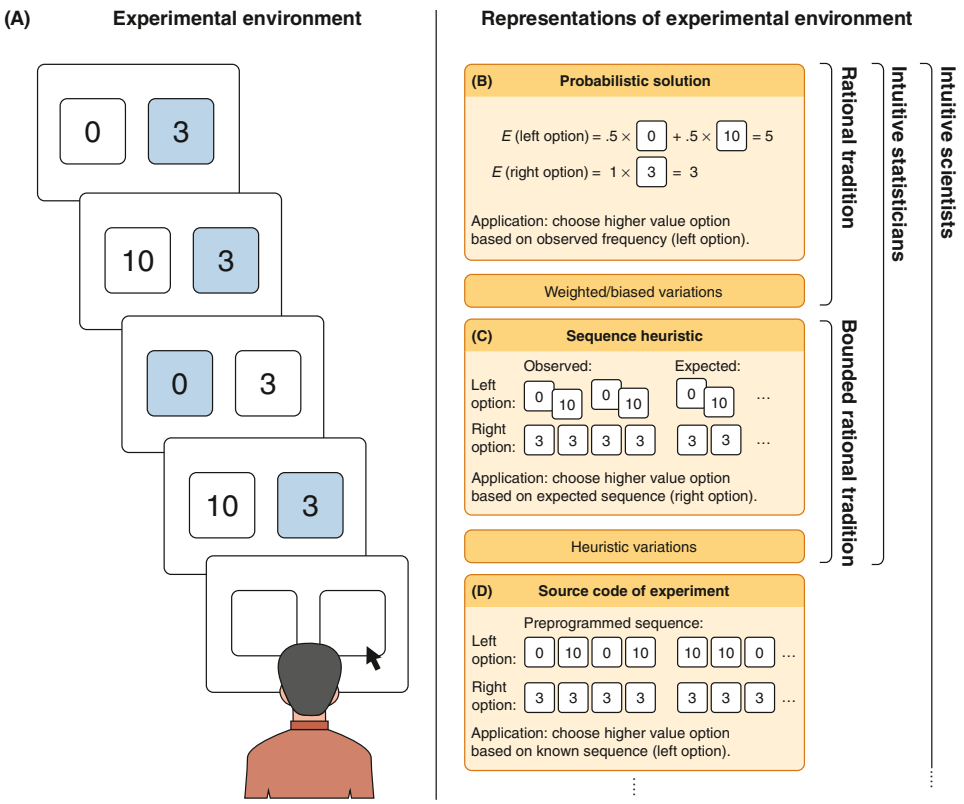

Figure 1. Illustration of possible representations of a simple decision-making task. The left-side panel (A) depicts a simple repeated-choice task [37,82]. Participants make choices between two options in a sequence of trials to collect points that are converted to monetary payoffs at the end of the experiment. The outcomes of the left option are risky (randomly generated: 10 points with $p = 0.5$, 0 points otherwise), right option is safe (3 points with $p = 1$). Participants observe outcomes of chosen and non-chosen options (history of choices and observations on previous four trials are shown; chosen options highlighted). The right-side panel depicts possible representations of the experimental environment, their applications, and the related research traditions. (B) Example of a rational representation: expected value, E, of options calculated based on the rolling average of the outcome values for each choice option. Informed by the probabilistic outcome generation process of the experimental design. (C) Example of a bounded rational representation: expected value of options based on a sequence heuristic that tracks the temporal pattern of the outcomes. Informed by the structure of many natural environments [1,83]. (D) Example of a representation relying on background knowledge of the experimental environment: expected value of options based on observing the preprogrammed sequence of outcomes in the source code of the experiment. Although this representation includes background knowledge regarding computer programs and psychology experiments (not conventionally considered in theories of people's representations), it is possible to represent the environment this way. The multiplicity of possible ways to represent the same environment illustrates the need to consider people's capability to create them in theories of decision making.

the explanation based on the currently assumed representation does not fit what people did, the addition of a biasing influence or changing what is treated as important can always make it fit.

Where Do Statistical Representations Come From?

The explanations for where statistical representations can come from, although not completely ignored, are not well developed. Researchers rarely explain the sources in more depth than

alluding to the potential developmental, learning-based, and evolutionary origins of the representations. Perhaps the most well-articulated of these explanations is the evolutionary one which assumes that statistical representations evolved through natural selection [50]. Such selection processes are assumed to be responsible for the generation of models that specify what the important parts of the environment are, and how those parts are to be interpreted and used to guide choices.

The problem with such explanations lies in their rigidness: evolutionary processes can only produce highly specialized behavioral adaptations, and, therefore, they can only explain highly specialized aspects of human behavior. For instance, deciding which apartment to buy or rent (a frequent scenario in decision-making experiments) does not closely resemble any ancient evolutionary environment where a particular heuristic – say the take-the-best strategy [51,52] – could have evolved. Yet, such rigid strategies can only be used in contexts similar to the original evolutionary environment. If people can use such heuristics in novel environments, there needs to be an additional explanation for how they can flexibly recognize the similarities between the original evolutionary environment and the current decision.

More generally, people's apparent representational flexibility remains unexplained in the solely model-application accounts of decision making. The rapid progress we can observe in our everyday life, for instance developments in science or engineering, can only be achieved by a cognitive (and an accompanying cultural) system that allows the improvement of its representations of the environment. Of course, the system that solves this problem had to evolve through evolution. Yet it is not a far-fetched thought that natural selection would favor a cognitive system that is capable of flexibly improving its representations of the environment, instead of one that uses a bag of static tricks for that purpose.

Intuitive Scientists and the Correspondence Problem

A potential way to better integrate the correspondence problem into theories of decision making, following on the research tradition of other fields of psychology (e.g., categorization and developmental psychology [26–29]), is to replace the intuitive statistician with the more apt metaphor of intuitive scientists. In our view, the benefit of this metaphor lies in providing a framework in which **flexible error correction** can occur – a necessary prerequisite for people's capability to address the correspondence problem. People are intuitive scientists in the sense that they are able to address the correspondence problem by attempting to identify and eliminate errors in any aspect of their representations. This is to say that, in contrast with certain variants of the intuitive scientist approach, we take this metaphor to be about people's capability to address the correspondence problem, and not about any particular approach of science (e.g., experimentation or the use of mathematical-statistical models).

Flexibility in error correction is important, because the possibility of errors in any aspect of representations is ever-present [53,54]. This entails a broader view of error than intuitive statistician theories allow. While intuitive statistician theories provide ways in which errors can be detected, potential errors are defined relatively narrowly in advance. One example of this is the way in which different fixed representations specify potential errors in gambling scenarios. These errors are only specified with respect to the aspect of the environment deemed important: for example, the errors that a sequence tracking heuristic (Figure 1C) would recognize are only related to the temporal structure of the outcomes (e.g., whether the sequence is longer or shorter). The problem is that there is nothing within such fixed representations of the task that would recognize potential errors outside of the representations. The representation can however contain infinitely many errors that cannot be defined in advance, such as the lack of including relevant knowledge in

the task representation (e.g., knowledge of the experimenter's intentions), or the lack of considering a more useful level of approximation (e.g., the level of the computer software that the experiment was implemented in). That none of these errors could be handled from within a fixed representation is a general problem that pervades all statistical representations (Box 1), and it highlights the importance of the flexibility of error-correction processes.

Because the potential for errors is infinite, they cannot be specified *a priori*, which necessitates processes that continually attempt to identify and eliminate errors [53,54]. First, there needs to be processes aimed at identifying errors in existing representations, for example, by comparing environmental feedback to expectations [55,56] or by assessing representations by other criteria [57]. Second, there needs to be processes aimed at correcting those errors, for example, by generating new representations (e.g., through modifying parts of existing representations). The point of these processes is not to guarantee that a 'correct' representation will be achieved or even that some errors will be corrected. But rather that their continual nature enables representations to vary and thus to potentially become more accurate – more accurate in the sense that they encapsulate aspects of the particular environment more adequately than competing representations for the particular problem at hand [17].

In the following sections, we explore what are in our opinion the most intriguing immediate consequences of this metaphor. This includes both an examination of generally underappreciated implications of the framework across different fields, as well as a discussion of how existing research based on this metaphor could inform and inspire theoretical and experimental work in decision-making research.

Rationality without Fixed Benchmarks

An important general implication of the specific intuitive scientist view we advocate is the elimination of fixed benchmarks for rationality. The conventional way of assessing rationality based on comparing people's behavior to fixed benchmarks is too myopic, because there is always a possibility of discovering a different but more useful representation of the environment, and therefore a better benchmark solution to some problem can be provided [35,58–61]. That is, the same way as there is no fixed method that scientists could use to apply, generate, and improve their theories, there is no fixed way in which people should do the same in everyday life.

Rational behavior, in this view, is taking the course of action that the most adequate representation for the particular decision environment dictates in line with whatever one's goal is (applying the representation), while actively facilitating the detection and correction of errors in those representations (enabling improvement of the representation). Existing fixed benchmarks for rational behaviors could not be considered rational under this framework, because they do not include mechanisms for detecting and correcting errors in any aspect of people's representations.

A related consequence of this view is that, if we accept the ever-present possibility of representational error, motivational factors may need to take on essential roles in explanations of rational behavior (Box 2). This is because changes in the cost of error correction for unforeseen problems or in light of new discoveries cannot be calculated in advance, and thus a purely cost–benefit analysis of rational behavior becomes inadequate. To resolve this problem, future theories of rationality may need to take alternative sources of motivation into account.

An Explanation for Mismatches

The intuitive scientist metaphor can also help resolving the puzzling mismatches between people's expected and observed behavior: they may well stem from different solutions to the

Box 2. Motivation and Cost of Thinking

Differentiating between error-correction and application processes with respect to environmental representations in decision making can lead to two qualitatively different ways to think about motivation and costs of thinking. Specifically, the view that all cognitive processes are governed by cost–benefit analyses [48,87] needs to be revisited. This is because a cost–benefit analysis only makes sense when an existing representation is being applied in a known environment. In such cases, the computational steps a person needs to take can be calculated based on the task at hand, the length of the process can be estimated, and the sources of possible errors can be known and be minimized in advance. Consequently, a cost can be calculated, and monetary incentives can be used to improve performance. This is the case in many 'robotic' tasks, such as alternated pressing of two keyboard buttons [88].

However, under uncertainty, there is no blueprint that a person can use to generate representations. In such cases, the outcome of the process is in-principle unknowable [89], and thus the cost of the process cannot be calculated in advance. Moreover, there are no guarantees that the error-correction processes lead to better solutions than the existing problematic solution. Explaining that people still attempt to generate representations under such circumstances therefore necessarily needs to shed light on how goal directedness can be coupled with forms of non-monetary motivation, such as intrinsic motivation or curiosity [90–92]. Such factors are necessary, at least in addition to standard monetary incentives, to explain why people engage in error-correction processes.

Overall, however, it is not clear whether there is any 'pure application' problem when it comes to people and the real world – arguably everything can be construed as learning and deciding under uncertainty. Even in simple repeated-choice experiments, people tend to generate and improve different candidate hypotheses [93]. If representational error-correction and application processes can be separated, it would be interesting to see how people can suppress intrinsic motivation or curiosity to behave more like a (present-day) computer.

correspondence problem. Differences in background knowledge (even when coupled with highly detailed experimental instructions) can easily lead to different solutions. Such differences can make it difficult for participants to infer that an environment is randomly generated; or that there exists no connection between events that at some point appeared related; or that information presented in an experiment might really be uninformative. In fact, it is possible that participants solved the correspondence problem differently even when they appear to behave in the way the experimenter expects, because substantially different representations can lead to the same observed behavior [62–64]. Thus, the fact that nothing necessitates that participants think about the experiment in similar terms as the researchers do needs to be taken more seriously into account in experimental investigations (Box 3).

Box 3. Studying Representations and Their Generation

The fact that people continually have to address the correspondence problem – and thus can have substantially different representations of the environment from the experimenter and from each other – presents a problem for the experimental study of decision making. This results from the change in what we should expect to be invariant features of human cognition: instead of expecting invariant representations of the environment, invariance should be expected in the processes that allow the generation and evaluation of representations. Measurements and manipulations in experimental studies should target factors that contribute to these processes.

One avenue is to manipulate experimental environments across the different ways in which they can be explained. The causal structure is just one example [80,94] – even within the context of such experiments, there seem to exist many other factors that people could consider, such as the simplicity and breadth of explanations [57]. Another avenue is to examine what people find important in an environment and how they try to identify and understand those aspects: investigating how people guide their own learning and how they ask questions about environments appear to be good starting points [73,95,96]. Experiments focusing on these questions can illuminate how people determine whether they should generate new or improve their existing explanations, or what they should explain to begin with.

Relatedly, researchers need to improve the sensitivity of their measurements of people's representations. The use of introspective methods can aid us in this goal by the virtue of people's privileged access to their own representations about the world [64,97]. Although misusing such techniques in weak psychological theorizing led to an underutilization of this method, such general disdain is undeserved when the method is used appropriately. Eliciting diagnostic verbal reports with respect to how people solve tasks [98–100], or what they think of certain features of real or experimental environments [5,93,101] can, and already has, contributed to our understanding of the underlying psychology. These practices can help to constrain researchers' explanation of participants' representations, and they can also serve as sensitive manipulation checks.

This pervasive possibility of diverging representations renders the predominant experimental approach of documenting specific behavioral patterns or their deviations from rational or bounded rational benchmarks necessarily limited (e.g., research questions taking the form of 'Which of these benchmarks describes people's behavior best?' or 'Do people behave (in)consistently with this benchmark?'). If, as we argued, people's behavior is largely determined by their best available representations, measuring the way in which they spontaneously respond to certain tasks does not need to reveal important aspects of human psychology. People's best available representations may well depend on contemporary cultural factors – for example, on the current state of cultural background knowledge, such as scientific theories, technological advances, or mathematical methods – rather than on meaningful psychological factors [65]. Without an explanation that takes the correspondence problem into account, simply documenting the matches or mismatches between scientists' and (lay)people's representations is unlikely to lead to a deeper understanding of human choice.

Representations as Explanations?

What is the nature of representations if they are not statistical models? As we have argued, representations need to contain flexible ways to detect and correct errors. We speculate that explanations [54,66] – statements of different complexity and specificity answering why and how questions – could be naturally suitable for this purpose. There are two major ways in which regarding representations as explanations can help improve decision-making research.

First, explanations naturally embody the flexibility needed for representation generation because of their compositional structure [67–69], (Box 4). Assuming such a structure for representations allows for the specification of processes that can generate new representations from existing ones: for example, by defining the rules by which changes can be effected in existing representations.

Box 4. Origin and Structure of Representations

An important direction for future research and theorizing is to provide a clear explanation of where representations come from. We expect that the idea that representations are built from primitive units of representations and rules by which they can be combined [67] will play an important role. Such compositional explanations [68,69] often posit some building blocks and rules that are inborn, but both of which can be 'upgraded' throughout development and even in adulthood. In such explanations, representational flexibility is an emergent feature of algorithms that run operations on primitive information processing units, in contrast with highly specific representations where flexibility has to be assumed.

Although compositionality is necessary for the kind of representational flexibility we argued people are capable of, it is not sufficient in itself. Even when statistical models take such form [68], their usefulness still depends on *a priori* determination of important aspects of the environment (e.g., 'observable' and 'hidden' variables). Yet, in many cases this cannot be done in advance (e.g., when there is no agreement on what 'hidden' variables to look for in the first place), (see Figure 1 in main text). Instead, the explanations need to incorporate how important aspects are selected and can be overwritten.

For an example of what we have in mind, consider Carey and Barner's [102] explanation for how people's representation of integer concepts develops. In brief, they propose that efficient and flexible representation of integers need to implement the concept of one-to-one correspondence (where number words refer to tallies of number sets), and the successor function (the concept of the next number). Children learn to associate number words with number sets early in development (implementing one-to-one correspondence, which allows the representation of small integers), which is later succeeded by an exact counting algorithm (implementing both concepts, which allows the representation of all integers).

There are several advantages of this kind of explanation: (i) it provides a clear explanation of the required concepts that need to be implemented by the representation; (ii) it provides a plausible compositional implementation that allows these concepts to flexibly emerge instead of having to be assumed; and (iii), importantly, in this general view of number representations (where learned algorithms contribute to how people represent the concepts), improvement can readily be integrated in the form of learning another algorithm. For instance, learning about or inventing different rules can actually improve people's representation of mathematical concepts – for example, extend their range of representations to rational or real numbers.

Another reason in favor of regarding representations as explanations is that they allow error detection and correction to be flexible. This is because, by intrinsically incorporating the aspects of the environment that are relevant, they open up the possibility of identifying errors in any aspect of them. By contrast, statistical models (even when they consider the generation and evaluation of representations, e.g., [70,71]), reduce the potential for error correction by introducing simplifying assumptions and approximations. For example, defining representations in terms of probability distributions reduces the potential for error correction, because the mechanisms by which people made those simplifications, why they chose to make those particular ones, and how they can undo them are not specified. Explanations can go further and allow for such error correction through arguments in many other aspects of representations including in their background assumptions or in general expectations about the quality of explanations [57].

Regarding representations as explanations reveals that even making a simple choice based on a single feature – say, between apartments based on the price – is more complicated than a statistical representation of it would make it appear. Solutions to a problem such as this, although seemingly determined by a simple proximal feature of the problem ('how much does the apartment cost'), are always dependent on extensive distal background knowledge. Such distal assumptions can include, for instance, why one needs an apartment ('to have a place where I am safe from harm'), why price is an important feature ('money is a scarce resource'), why price is the only important feature ('I cannot afford to spend enough to consider other features'), and so on. Solving other decision-making problems similarly depends on having such complex proximal and distal explanatory structures. Learning the causal structure of environments [72], learning in a self-directed way [73], taking communicative intent into account in one's decisions [74–77], or the selective relevance and interpretation of environmental feedback [55,56] are all good examples of how much decision making relies on explanatory structures that are more complex than what a fixed representation of the proximal problem would suggest.

Extensive enquiry about the structure and function of explanations already exists [66,78], and we know much about how they are evaluated [57], about the role they play in causal reasoning and judgment [79,80], and about the constraints the limitations of cognitive systems place on how people generate them [81], amongst many others. But there is a whole range of questions waiting to be answered related to people's ability to flexibly generate and improve their representations (see Outstanding Questions). We hope and expect that answering such questions – and especially the unifying role error-correction processes play in them – will bring us closer to a more complete explanation of explanation.

Outstanding Questions

How do people become and remain motivated to execute error correction? Can this be sufficiently explained solely in terms of a cost–benefit analysis if there is no way to calculate the benefits (because the success of the correction is not guaranteed)? If not, what kind of motivation is required to engage error-correction processes?

What are the developmental origins and trajectories of the ability to flexibly generate and improve representations? What are the processes by which novel forms of representations supersede existing ones?

How pervasive is the problem that there can always be a mismatch between the researcher's and the participant's representations? Can something still be learned from comparing researchers' (supposedly optimal) solutions and the participants' solutions for a certain task?

Concluding Remarks

The difficulties associated with constructing accurate environmental representations – the correspondence problem – receives undeservedly little attention in theories of human choice. The prevailing approach is to introduce the researcher's own representation of the decision environment into the theories of people's representations. This is problematic, because people's representations often differ from the researcher's due to differences in their background knowledge, as well as because people – similar to researchers – are capable of representing environments in multiple different ways and of improving their representations. Focusing on how people address the challenge associated with the correspondence problem via flexible detection and correction of representational error can provide a fruitful avenue for future research.

Acknowledgments

We are grateful to Nick Chater, Andrei Cimpian, Chris Donkin, Craig McKenzie, Christin Schulze, and Shlomi Sher for illuminating discussions and comments on a previous version of this paper. We also thank the anonymous referees and the editor, Lindsey

Drayton, for their thoughtful and constructive feedback. This work was supported by the Australian Research Council (DP160101186, DP190101675).

References

1. Gaissmaier, W. and Schooler, L.J. (2008) The smart potential behind probability matching. *Cognition* 109, 416–422
2. Navarro, D.J. *et al.* (2016) Learning and choosing in an uncertain world: an investigation of the explore–exploit dilemma in static and dynamic environments. *Cogn. Psychol.* 85, 43–77
3. Hahn, U. and Warren, P.A. (2009) Perceptions of randomness: why three heads are better than four. *Psychol. Rev.* 116, 454–461
4. Ayton, P. and Fischer, I. (2004) The hot hand fallacy and the gambler's fallacy: two faces of subjective randomness? *Mem. Cogn.* 32, 1369–1378
5. Lagnado, D.A. *et al.* (2006) Insight and strategy in multiple-cue learning. *J. Exp. Psychol. Gen.* 135, 162–183
6. Fiedler, K. *et al.* (2013) Pseudocontingencies. *Curr. Dir. Psychol. Sci.* 22, 324–329
7. Schulze, C. *et al.* (2017) Hold it! The influence of lingering rewards on choice diversification and persistence. *J. Exp. Psychol. Learn. Mem. Cogn.* 43, 1752–1767
8. Loewenstein, G. (2006) The Pleasures and Pains of Information. *Science* 312, 704–706
9. Iigaya, K. *et al.* (2016) The modulation of savouring by prediction error and its effects on choice. *eLife* 5, e13747
10. Bennett, D. *et al.* (2016) Intrinsic valuation of information in decision making under uncertainty. *PLoS Comput. Biol.* 12, e1005020
11. von Neumann, J. and Morgenstern, O. (1947) *Theory of games and economic behavior* (2nd edn.), Princeton University Press
12. Savage, L.J. (1954) *The Foundations of Statistics*, John Wiley and Sons
13. Kahneman, D. and Tversky, A. (1979) Prospect theory: an analysis of decision under risk. *Econometrica* 47, 263–292
14. Tversky, A. and Kahneman, D. (1974) Judgment under uncertainty: heuristics and biases. *Science* 185, 1124–1131
15. Costello, F. and Watts, P. (2014) Surprisingly rational: probability theory plus noise explains biases in judgment. *Psychol. Rev.* 121, 463–480
16. Oaksford, M. and Chater, N. (2007) *Bayesian Rationality: The Probabilistic Approach to Human Reasoning*, Oxford University Press
17. Simon, H.A. (1972) Theories of bounded rationality. *Decis. Organ.* 1, 161–176
18. Gigerenzer, G. and Todd, P.M. (1999) *Simple heuristics that make us smart*, Oxford University Press
19. Gigerenzer, G. *et al.* (2011) *Heuristics: The Foundations of Adaptive Behavior*, Oxford University Press
20. Fiedler, K. and Juslin, P. (2006) *Information sampling and adaptive cognition*, Cambridge University Press
21. Le Mens, G. and Denrell, J. (2011) Rational learning and information sampling: on the "naivety" assumption in sampling explanations of judgment biases. *Psychol. Rev.* 118, 379–392
22. Gallistel, C.R. (2001) Mental representations; psychology of. In *International Encyclopedia of the Social and Behavioural Sciences* (Smelser, N.J. and Baltes, P.B., eds), pp. 9691–9695, Elsevier
23. Hammond, K.R. *et al.* (2000) Coherence and correspondence theories in judgment and decision making. In *Judgment and Decision Making: An Interdisciplinary Reader* (2nd edn.) (Conolly, T. *et al.*, eds), pp. 53–65, Cambridge University Press, NY
24. Dunwoody, P.T. (2009) Theories of truth as assessment criteria in judgment and decision making. *Judgm. Decis. Mak.* 4, 116–125
25. Arkes, H.R. *et al.* (2016) How bad is incoherence? *Decision* 3, 20–39
26. Murphy, G.L. and Medin, D.L. (1985) The role of theories in conceptual coherence. *Psychol. Rev.* 92, 289–316
27. Kuhn, D. (1989) Children and adults as intuitive scientists. *Psychol. Rev.* 96, 674–689
28. Gopnik, A. and Meltzoff, A.N. (1997) *Words, Thoughts, and Theories*, MIT Press
29. Carey, S. and Spelke, E. (1996) Science and core knowledge. *Philos. Sci.* 63, 515–533
30. Peterson, C.R. and Beach, L.R. (1967) Man as an intuitive statistician. *Psychol. Bull.* 68, 29–46
31. McKenzie, C.R.M. and Soll, J.B. (1996) Which reference class is evoked? *Behav. Brain Sci.* 19, 34–35
32. Brette, R. (2019) Is coding a relevant metaphor for the brain? *Behav. Brain Sci.* 42, e215
33. Szollosi, A. and Newell, B.R. (2020) What is the purpose of cognition? *Behav. Brain Sci.* 43, e25
34. Navarro, D.J. (2018) Between the devil and the deep blue sea: tensions between scientific judgement and statistical model selection. *Comput. Brain Behav.* 2, 28–34
35. Felin, T. *et al.* (2017) Rationality, perception, and the all-seeing eye. *Psychon. Bull. Rev.* 24, 1040–1059
36. Markowitz, H. (1952) The utility of wealth. *J. Polit. Econ.* 60, 151–158
37. Hertwig, R. *et al.* (2004) Decisions from experience and the effect of rare events in risky choice. *Psychol. Sci.* 15, 534–539
38. Fox, C.R. and Hadar, L. (2006) "Decisions from experience" = sampling error + prospect theory: reconsidering Hertwig, Barron, Weber & Erev (2004). *Judgm. Decis. Mak.* 1, 159–161
39. Bhatia, S. and Loomes, G. (2017) Noisy preferences in risky choice: a cautionary note. *Psychol. Rev.* 124, 678–687
40. Newell, B.R. *et al.* (2003) Empirical tests of a fast-and-frugal heuristic: not everyone "takes-the-best". *Organ. Behav. Hum. Decis. Process.* 91, 82–96
41. Newell, B.R. (2005) Re-visions of rationality? *Trends Cogn. Sci.* 9, 11–15
42. Szollosi, A. and Donkin, C. (2019) Neglected sources of flexibility in psychological theories: from replicability to good explanations. *Comput. Brain Behav.* 2, 190–192
43. Chater, N. (2018) *The mind is flat: the illusion of mental depth and the improvised mind*, Penguin
44. Ariely, D. (2008) *Predictably irrational: the hidden forces that shape our decisions*, HarperCollins
45. Anderson, J.R. (1990) *The adaptive character of thought*, Lawrence Erlbaum Associates, Inc.
46. Costello, F. and Watts, P. (2019) The rationality of illusory correlation. *Psychol. Rev.* 126, 437–450
47. Gigerenzer, G. and Selten, R. (2001) *Bounded rationality: the adaptive toolbox*, MIT Press
48. Lieder, F. and Griffiths, T.L. (2020) Resource-rational analysis: understanding human cognition as the optimal use of limited computational resources. *Behav. Brain Sci.* 43, e1
49. Gershman, S.J. *et al.* (2015) Computational rationality: a converging paradigm for intelligence in brains, minds, and machines. *Science* 349, 273–278
50. Gigerenzer, G. *et al.* (1998) Ecological intelligence: an adaptation for frequencies. In *The Evolution of Mind* (Cummins, D.D. and Allen, C., eds), pp. 9–29, Oxford University Press
51. Gigerenzer, G. and Goldstein, D.G. (1996) Reasoning the fast and frugal way: models of bounded rationality. *Psychol. Rev.* 103, 650–669
52. Newell, B.R. and Shanks, D.R. (2003) Take the best or look at the rest? Factors influencing "one-reason" decision making. *J. Exp. Psychol. Learn. Mem. Cogn.* 29, 53–65
53. Popper, K. (1963) *Conjectures and refutations: the growth of scientific knowledge*, Routledge
54. Deutsch, D. (2011) *The Beginning of Infinity: Explanations That Transform the World*, Allen Lane
55. Hogarth, R.M. *et al.* (1991) Learning from feedback: exactingness and incentives. *J. Exp. Psychol. Learn. Mem. Cogn.* 17, 734–752
56. Newell, B.R. *et al.* (2015) *Straight choices: the psychology of decision making* (2nd edn.), Psychology Press
57. Lombrozo, T. (2016) Explanatory preferences shape learning and inference. *Trends Cogn. Sci.* 20, 748–759
58. McKenzie, C.R.M. (2003) Rational models as theories – not standards – of behavior. *Trends Cogn. Sci.* 7, 403–406
59. Popper, K. (1985) Knowledge without authority (1960). In *Popper Selections* (Miller, D., ed.), pp. 46–57, Princeton University Press

60. Chater, N. *et al.* (2018) Mind, rationality, and cognition: an interdisciplinary debate. *Psychon. Bull. Rev.* 25, 793–826
61. Rahnev, D. and Denison, R.N. (2018) Suboptimality in perceptual decision making. *Behav. Brain Sci.* 41, e223
62. Dulany, D.E. *et al.* (1984) A case of syntactical learning and judgment: how conscious and how abstract? *J. Exp. Psychol. Gen.* 113, 541–555
63. Shanks, D.R. and St. John, M.F. (1994) Characteristics of dissociable human learning systems. *Behav. Brain Sci.* 17, 367–395
64. Newell, B.R. and Shanks, D.R. (2014) Unconscious influences on decision making: a critical review. *Behav. Brain Sci.* 38, 1–19
65. Simon, H.A. (1992) What is an "explanation" of behavior? *Psychol. Sci.* 3, 150–161
66. Keil, F.C. (2006) Explanation and understanding. *Annu. Rev. Psychol.* 57, 227–254
67. Newell, A. *et al.* (1958) Elements of a theory of human problem solving. *Psychol. Rev.* 65, 151–166
68. Lake, B.M. *et al.* (2017) Building machines that learn and think like people. *Behav. Brain Sci.* 40, e253
69. Frankland, S.M. and Greene, J.D. (2020) Concepts and compositionality: in search of the brain's language of thought. *Annu. Rev. Psychol.* 71, 273–303
70. Navarro, D.J. and Perfors, A.F. (2011) Hypothesis generation, sparse categories, and the positive test strategy. *Psychol. Rev.* 118, 120–134
71. Dasgupta, I. *et al.* (2017) Where do hypotheses come from? *Cogn. Psychol.* 96, 1–25
72. Waldmann, M.R. *et al.* (2006) Beyond the information given. *Curr. Dir. Psychol. Sci.* 15, 307–311
73. Gureckis, T.M. and Markant, D.B. (2012) Self-directed learning. *Perspect. Psychol. Sci.* 7, 464–481
74. Grice, H.P. (1975) Logic and conversation. In *Syntax and Semantics 3: Speech Arts* (Cole, P. and Morgan, J.L., eds), pp. 41–58, Academic Press
75. Hilton, D.J. (1995) The social context of reasoning: conversational inference and rational judgment. *Psychol. Bull.* 118, 248–271
76. Sher, S. and McKenzie, C.R.M. (2006) Information leakage from logically equivalent frames. *Cognition* 101, 467–494
77. Oaksford, M. and Chater, N. (2020) New paradigms in the psychology of reasoning. *Annu. Rev. Psychol.* 71, 305–330
78. Lombrozo, T. (2006) The structure and function of explanations. *Trends Cogn. Sci.* 10, 464–470
79. Hilton, D.J. (1996) Mental models and causal explanation: judgements of probable cause and explanatory relevance. *Think. Reason.* 2, 273–308
80. Rehder, B. and Hastie, R. (2001) Causal knowledge and categories: the effects of causal beliefs on categorization, induction, and similarity. *J. Exp. Psychol. Gen.* 130, 323–360
81. Horne, Z. *et al.* (2019) Explanation as a cognitive process. *Trends Cogn. Sci.* 23, 187–199
82. Camilleri, A.R. and Newell, B.R. (2011) When and why rare events are underweighted: a direct comparison of the sampling, partial feedback, full feedback and description choice paradigms. *Psychon. Bull. Rev.* 18, 377–384
83. Newell, B.R. and Schulze, C. (2016) Probability matching. In *Cognitive Illusions* (2nd edn.) (Pohl, R.F., ed.), pp. 62–78, Psychology Press
84. Kahneman, D. and Tversky, A. (1973) On the psychology of prediction. *Psychol. Rev.* 80, 237–251
85. Koehler, J.J. (1996) The base rate fallacy reconsidered: descriptive, normative, and methodological challenges. *Behav. Brain Sci.* 19, 1–17
86. Brase, G.L. and Hill, W.T. (2017) Adding up to good Bayesian reasoning: problem format manipulations and individual skill differences. *J. Exp. Psychol. Gen.* 146, 577–591
87. Lieder, F. and Griffiths, T.L. (2017) Strategy selection as rational metareasoning. *Psychol. Rev.* 124, 762–794
88. DellaVigna, S. and Pope, D. (2018) What motivates effort? Evidence and expert forecasts. *Rev. Econ. Stud.* 85, 1029–1069
89. Knight, F.H. (1921) *Risk, uncertainty, and profit*, Houghton and Mifflin
90. Berlyne, D.E. (1966) Curiosity and exploration. *Science* 153, 25–33
91. Deci, E. (1972) Intrinsic motivation, extrinsic reinforcement, and inequity. *J. Pers. Soc. Psychol.* 22, 113–120
92. Loewenstein, G. (1994) The psychology of curiosity: a review and reinterpretation. *Psychol. Bull.* 116, 75–98
93. Szollosi, A. *et al.* (2019) Simultaneous underweighting and overestimation of rare events: unpacking a paradox. *J. Exp. Psychol. Gen.* 148, 2207–2217
94. Hayes, B.K. *et al.* (2014) The role of causal models in multiple judgments under uncertainty. *Cognition* 133, 611–620
95. Rothe, A. *et al.* (2018) Do people ask good questions? *Comput. Brain Behav.* 1, 69–89
96. Coenen, A. *et al.* (2019) Asking the right questions about the psychology of human inquiry: nine open challenges. *Psychon. Bull. Rev.* 26, 1548–1587
97. Lovibond, P.F. and Shanks, D.R. (2002) The role of awareness in Pavlovian conditioning: empirical evidence and theoretical implications. *J. Exp. Psychol. Anim. Behav. Process.* 28, 3–26
98. Ericsson, K.A. and Simon, H.A. (1980) Verbal reports as data. *Psychol. Rev.* 87, 215–251
99. Fox, M.C. *et al.* (2011) Do procedures for verbal reporting of thinking have to be reactive? A meta-analysis and recommendations for best reporting methods. *Psychol. Bull.* 137, 316–344
100. Szaszi, B. *et al.* (2017) The cognitive reflection test revisited: exploring the ways individuals solve the test. *Think. Reason.* 23, 207–234
101. Konstantinidis, E. and Shanks, D.R. (2014) Don't bet on it! Wagering as a measure of awareness in decision making under uncertainty. *J. Exp. Psychol. Gen.* 143, 2111–2134
102. Carey, S. and Barner, D. (2019) Ontogenetic origins of human integer representations. *Trends Cogn. Sci.* 23, 823–835