
Cognitive bias benchmarks should incorporate insights from ecological rationality

David Thorstad

Abstract

Recent studies suggest that large language models (LLMs) exhibit a range of cognitive biases familiar from human cognition. This paper argues that **cognitive bias benchmarks should incorporate insights from ecological rationality**, a leading paradigm in the study of cognitive bias. Three strategies for resisting some existing allegations of cognitive bias are extracted from the ecological rationality paradigm. Each strategy is used to generate a concrete recommendation for improving cognitive bias benchmarks.

1 Introduction

Recent studies suggest that large language models (LLMs) exhibit a range of cognitive biases familiar from human cognition (Binz and Schulz 2023; Dasgupta et al. 2022; Jones and Steinhardt 2022; Talboy and Fuller 2023). These include anchoring biases (Nguyen 2024; Talboy and Fuller 2023), availability biases (Lin and Ng 2023; Jones and Steinhardt 2022), and framing effects (Horton 2023; Jones and Steinhardt 2022), among many others. Findings of cognitive bias have been used to understand (Binz and Schulz 2023; Shiffrin and Mitchell 2023) and guide implementation of LLMs in professional settings (Schmidgall et al. 2024), as well as to construct benchmarks for model evaluation (Echterhoff et al. 2024; Malberg et al. 2025; Shaikh et al. 2024).

Ensuring the validity (Eriksson et al. 2025; Raji et al. 2021; Reuel et al. 2024) of cognitive bias benchmarks requires benchmarks to measure genuine biases and exclude unbiased behaviors. This paper shows how the ecological rationality paradigm (§2) in cognitive science can be used to improve the validity of cognitive bias benchmarks by including a higher proportion of biased behaviors within them. On this basis, it argues that **cognitive bias benchmarks should incorporate insights from ecological rationality**.

This argument has a negative part and a positive part. The negative argument (§3) extracts three strategies for mitigating the appearance of cognitive bias in LLMs by seeking appropriate normative standards (§4), clarifying bias concepts (§5) and attending to ecological validity (§6). This suggests that existing benchmarks could be improved by incorporating insights from ecological rationality.

The positive argument makes good on this suggestion by using each strategy to generate a concrete recommendation for improving benchmarks of cognitive bias in LLM cognition: using rational analysis to derive normative standards from computational models (§7.1), using taxonomy-based clarification to disambiguate biases along leading scientific taxonomies (§7.2), and sampling stimuli from natural environments to ensure representative experimental design (§7.3). This shows how bias benchmarks can be strengthened by incorporating insights from ecological rationality. The study closes by examining alternative positions (§8) and their relationship to the ecological rationality paradigm.

Strategy	Description	Example
Broadening norms	Normative standards governing heuristic cognition are broadened to incorporate normatively relevant features of situations.	Experiments aim to maximize value of information, rather than to decisively disconfirm rules. (Oaksford and Chater 1994)
Clarifying biases	Definitions of cognitive heuristics (and associated biases) are clarified so that their presence or absence can be reliably empirically detected.	Availability bias does not include reliance on contents of retrieved information or on retrieval latencies (Schwarz and Vaughn 2002).
Ecological validity	Judgmental cues are assessed for their ecological validity: correlation with relevant quantities in task environments.	Player name recognition is an ecologically valid guide to Wimbledon victory chances. (Serwe and Frings 2006)

Table 1: Three ecological strategies

2 Background and related work

2.1 Heuristics and biases

Humans often make judgments and decisions using a toolbox of fast-and-frugal heuristic strategies (Gigerenzer and Selten 2001; Gigerenzer and Gaissmaier 2011). Heuristics have *biases*, understood as situations in which they systematically deviate from axiomatic rationality criteria. The heuristics and biases program of Daniel Kahneman and Amos Tversky pioneered a dual usage of cognitive bias studies: as a descriptive method for establishing the use of corresponding heuristics, and a normative criterion for evaluating their performance (Gilovich et al. 2002; Kahneman et al. 1982; Tversky and Kahneman 1974).

For example, Kahneman and Tversky proposed that agents often estimate quantities using a heuristic called *anchoring and adjustment*. Agents *anchor* on an initial estimate, then iteratively *adjust* the estimate upwards and downwards to incorporate relevant information. Judgments made by anchoring and adjustment show an *anchoring effect*: they tend to be skewed in the direction of the anchor. Kahneman and Tversky proposed that many anchoring effects should be understood as *anchoring biases*. Descriptively, they took anchoring bias as evidence that agents had employed a cognitive process of anchoring and adjustment. Normatively, they took anchoring bias as a problematic feature of human cognition in need of correction.

2.2 Ecological rationality

A leading approach to the rationality of heuristic cognition holds that rationality is *ecological*, or environment-relative (Todd and Gigerenzer 2012; Gigerenzer 2019). All problem-solving strategies perform well in some environments, and poorly in others. For this reason, theories of ecological rationality take rationality to be a matter of choosing strategies that are appropriate in a given environment. While ecological rationality theorists concede that agents sometimes use inappropriate strategies, they argue that strategies which appear inappropriate are often in fact well-suited to the environments where they are used.

In particular, the *narrow norms and vague heuristics critique* of Gerd Gigerenzer and collaborators holds that the heuristics and biases program often relies on inappropriately narrow normative standards, and insufficiently precise definitions of heuristics (Gigerenzer 1991; Gigerenzer et al. 1991; Gigerenzer 1996). Once normative standards are broadened and heuristics are clarified, we see that many seemingly irrational heuristics fall within correct normative standards.

3 Three ecological strategies

3.1 Broadening norms

This paper shows how three strategies from the ecological rationality paradigm (Table 1) can be used to resist some allegations of cognitive bias in LLMs. The first strategy, *broadening norms*, takes inspiration from the first part of the narrow norms and vague heuristics critique. This strategy suggests that some proposed biases rest on inappropriately narrow normative standards. By broadening normative standards, this strategy seeks to show that some proposed biases do not violate correct normative standards.

For example, early studies assumed that agents gather information with the aim of decisively falsifying logical rules (Wason 1968; Rips 1994). Agents were accused of irrational confirmation bias for seeking information that could only confirm, but not falsify, logical rules. More recent studies suggest that agents should sometimes gather information with the aim of maximizing the value of information (Elqayam and Over 2013; Oaksford and Chater 2007), understood in terms of maximizing expected Shannon information gain (Shannon 1948; Oaksford and Chater 1994). Information gain can be produced by confirming as well as falsifying hypotheses. As a result, these studies suggest that some early allegations of confirmation bias rest on an inappropriate normative standard of seeking falsification rather than maximizing the value of received information.

3.2 Clarifying biases

The second strategy, *clarifying biases*, takes inspiration from the second part of the narrow norms and vague heuristics critique. This strategy suggests that some proposed heuristics and the associated biases are vaguely defined. By clarifying bias definitions, this strategy seeks to show that some proposed biases fall outside them.

For example, we will see in Section 5 that Tversky and Kahneman originally defined the availability heuristic as “estim[ing] frequency or probability by the ease with which instances or associations could be brought to mind” (Tversky and Kahneman 1973, p. 208). Availability was initially understood to encompass both *objective availability*, the speed of memory retrieval and the contents of retrieved information, and *subjective availability*, the feeling of subjective ease during memory retrieval.

Early critics charged that objective availability should be excluded from the definition of availability bias, because the contents of retrieved information and the speed of retrieval are often reliable guides to frequency and probability (Hertwig and Pachur 2005; Schwarz and Vaughn 2002). This led to clarification of availability bias as involving only subjective availability, which excluded many early allegations of availability bias in human cognition.

3.3 Ecological validity

The final strategy, attending to *ecological validity*, defines biases in an environment-relative way. Cues are ecologically valid to the extent that they indicate relevant quantities in task environments. This strategy suggests that some apparent biases reflect appropriate reliance on ecologically valid cues.

To illustrate, Serwe and Frings (2006) show that name-recognition of tennis players is a strong predictor of their likelihood of winning early Wimbledon matches. In this context, agents may employ the *recognition heuristic* of judging that a player will win if they recognize her but not her opponent (Goldstein and Gigerenzer 1999, 2002; Gigerenzer and Goldstein 2011). Rather than classifying the recognition heuristic as an irrational bias, due to potential misfires or manipulation, this strategy suggests that reliance on recognition can be rational in environments where recognition is an ecologically valid cue.

3.4 Taking stock

Sections 2-3 introduced the ecological rationality paradigm and extracted three strategies from this paradigm. The next order of business is to show how these strategies can be applied to some recent studies of cognitive bias in LLMs (§§4-6).

4 Broadening norms

The first ecological strategy involves broadening normative standards to show that some proposed biases fall within them. By way of illustration, ecological rationality theorists hold that many purported biases involve accuracy-increasing inferences from relevant information contained in prompts. Experiments wrongly impose a normative standard that forbids inferences based on this relevant information. Revising normative standards to allow appropriate use of relevant information may lessen some appearances of cognitive bias.

For example, Shaikh et al. (2024) assess anchoring bias using the following prompt:

Estimate the number of jellybeans in a hidden jar. Your friend guesses 750. What's your estimate? (A) 50. (B) 200. (C) 800. (D) 1200.

Models are judged to exhibit anchoring bias when they frequently choose option (C). Shaikh and colleagues take this to reveal anchoring bias through reliance on the anchor value of 750.

This anchor is irrelevant only to the extent that the friend's judgment is uncorrelated with the true number of jellybeans. An established research tradition into *wisdom of crowds* phenomena suggests that lay estimates of quantities such as jellybeans in a hidden jar are nontrivially correlated with the truth, and indeed that averaging many such estimates is an excellent way to estimate unknown quantities of this sort (Hong and Page 2004; List and Goodin 2001; Surowiecki 2004). For example, Treynor (1987) ran two experiments in which investment students were asked to estimate the average number of beans in a jar. Mean estimates were 841 and 871 beans, against true values of 810 and 850 beans.

If this is right, then Shaikh et al. (2024) enforce an inappropriate normative standard of avoiding reliance on relevant anchor information by forbidding reliance on the friend's relevant estimate. Because the friend's estimate is evidentially relevant, it is normatively appropriate to show some reliance on the friend's estimate, particularly when the friend's estimate is the only piece of relevant information provided in the prompt.

Similarly, Knipper et al. (2025) assess representativeness bias using the following prompt:

You're introduced to two people— you know one's an electrician, the other's a therapist. One is wearing overalls and boots; the other, dress pants and a shirt. Who's the therapist?

Knipper and colleagues propose that choosing the second individual “reflects bias by stereotyping their appearances to their occupations, despite equal likelihood for either to be the therapist.”

An ecological perspective challenges the normative requirement to assign equal likelihood for each individual to be the therapist. From an ecological perspective, the likelihood that each individual is a therapist should be matched to evidence about the correlation between clothing and occupations. While it is important not to lean on ungrounded stereotypes, there is good evidence that clothing choice correlates with occupation (Hester and Hehman 2023) and that this correlation can be learned by even relatively rudimentary models (Song et al. 2011). In this perspective, it is inappropriate to impose a normative standard of assigning equal likelihood to each individual being a therapist, because this normative standard requires models to neglect evidence of correlation between clothing choice and occupation.

In summary, a first ecological strategy of seeking appropriate normative standards is illustrated by the strategy of recovering defensible inferences from relevant information. In case studies of representativeness and anchoring bias, LLMs are inappropriately judged using normative standards that forbid them from making relevant inferences using provided information.

5 Clarifying biases

The second part of the narrow norms and vague heuristics critique charges that many heuristics, alongside their associated biases, are too broadly defined. This motivates a second ecological strategy: clarifying bias concepts to evaluate which claimed biases fall within them.

By way of illustration, consider the availability heuristic. Tversky and Kahneman define the availability heuristic as “estimat[ing] frequency or probability by the ease with which instances or associations could be brought to mind” (Tversky and Kahneman 1973, p. 208). It is now widely recognized that this definition encompasses two distinct senses of availability (Hertwig and Pachur 2005; Schwarz and Vaughn 2002).

Objective availability involves reliance on the contents of information retrieved from memory, or on objectively observable features of memory retrieval such as response latencies. For example, agents rely on objective availability if they judge that white swans are numerous because examples of white swans are frequently or quickly retrieved from memory. *Subjective availability* involves subjective or experiential features of the memory retrieval process such as feelings of ease or fluency. For example, agents rely on subjective availability if they judge that white swans are numerous because examples of white swans can be retrieved without strain or felt disfluency.

While some theorists express concern about the distorting potential of subjective availability, few theorists hold that bare reliance on objective availability is an irrational bias. Relying on the contents of information retrieved from memory is precisely the purpose of memory search. And the speed of memory retrieval is often a strong predictor of the commonality of items in memory as well as in the world, since items that are common in the world are more likely to be frequently and deeply encoded into memory (Hertwig and Pachur 2005; Schwarz and Vaughn 2002). However, some leading studies of availability bias in LLMs study objective rather than subjective availability.

One common finding is that LLMs rely on the contents of information stored in memory. In particular, in the absence of individuating information, LLMs tend to judge the frequency of items in the world to be similar to their frequency in the training data (Lin and Ng 2023; Lyu et al. 2025). For example, Lin and Ng (2023) ask models trained on a dataset of drug interactions to predict the interaction type of heavily redacted descriptions such as the following:

[Drug name] competes with a N/A of N/A for N/A N/A N/A notably N/A N/A
N/A N/A N/A [drug name] N/A N/A N/A and N/A N/A.

Lin and Ng find that models most commonly predict the redacted interactions to be negative, which is by far the most common type of interaction in the corpus of medical texts on which the model was trained.

Lin and Ng treat this finding as a type of availability bias, in which the objective availability of negative interactions stored in memory wrongly influences models to predict negative interactions when faced with redacted descriptions. But such reliance on the contents of retrieved information, in the form of observed frequencies, is precisely what many statistical learning rules would recommend (Icard 2018; Vul et al. 2014). As in Section 4, there is no uncontroversial normative requirement to ignore relevant information, and the contents of training data stored in memory are relevant to future predictions.

Other authors find that LLMs rely on the latency of memory retrieval. For example, Malberg et al. (2025) find a positive correlation between the recall latency of an event and estimates of the probability that this event will occur in the future. In many contexts, events which are frequent in the world are more deeply and frequently encoded in memory, and hence have lower recall latencies (Hertwig and Pachur 2005; Schwarz and Vaughn 2002). In these contexts, correlation between estimated event probabilities and recall latencies will correctly reflect empirical regularities, and cannot be taken without further argument to constitute an irrational bias.

In summary, the distinction between objective and subjective availability illustrates the potential of a second ecological strategy: clarifying bias concepts. Once bias concepts are clarified, it may be observed that some experimental observations fall outside of them, as in the observations that models rely on response latencies or frequencies of occurrence in training data.

6 Ecological validity

In the ecological tradition, cues are *ecologically valid* to the extent that they indicate relevant quantities in task environments. Rational inference relies on cues in proportion to their ecological validity. A third ecological strategy consists of studying the ecological validity of cues to assess when LLMs are using them appropriately.

By way of illustration, Knipper et al. (2025) test for anchoring bias using the following prompt:

You read that 130,000 people traveled on Florida interstates last year; likely mostly tourists. How many people do you think actually live in Florida?

Noting that the actual population of Florida exceeds twenty million, Knipper and colleagues treat population estimates in the hundred thousands as evidence of overreliance on the anchoring quantity 130,000.

In ecological perspective, the relevant question is how interstate highway use correlates with state population size. If the anchoring quantity of 130,000 travelers is abnormally low for a state with population beyond twenty million, then learning the anchor should lead models to revise their estimate of Florida's population downwards. Two figures suggest that the provided anchor is indeed abnormally low.

First, in 2024 a total of 36 billion vehicle-miles were logged on Florida interstates (Federal Highway Administration 2024). For those miles to have been logged by 130,000 people would require each to have driven over 270,000 miles on the interstate alone, which amounts to over twelve hours of daily interstate driving at a pace of sixty miles per hour. Second, if most of the 130,000 travelers on Florida interstates were tourists, then mere tens of thousands of Florida residents used the interstate highway system during the entire year. That would be surprising given a population in the tens of millions.

Given these figures, the anchoring quantity may provide statistical grounds for downwards revisions to Florida population estimates. In this way, low estimates for the population of Florida may reflect ecologically valid reliance on cues about highway travel provided by the prompt.

Similarly, Lou and Sun (2026) assess anchoring bias by asking models to predict the Tesla stock price next Tuesday. Models are told either that last Tuesday's high price for Tesla was \$343.55 or that last Tuesday's low price for Tesla was \$326.00. Models are deemed to show anchoring bias to the extent that they predict greater future stock prices when shown the high price rather than the low price.

In ecological perspective, the relevant question is how daily high and low stock prices are related. If a daily low of \$326.00 or below is relatively unsurprising given a daily high of \$343.55, then models should not make substantially more optimistic predictions when told the high price instead of the low price. By contrast, if a daily high price of \$343.55 tends to be accompanied by daily low prices substantially above \$326.00, then predictions should be more optimistic when told the high price instead of the low price.

When markets are not especially volatile, daily price fluctuations are often significantly smaller than this (Andersen et al. 2001). For example, in the years 2020-5, the average daily gap between high and low Tesla stock prices was \$5.29, with fluctuations of at least \$17.55 (the amount described by the prompt) occurring on 4.11% of days (Yahoo Finance 2026). This suggests that failure to correct for the rare phenomenon of abnormally large price fluctuations may not reflect anchoring bias, but rather ecologically valid inference given historical price fluctuations.

7 Recommendations

Sections 4-6 had the negative project of showing how three ecological rationality strategies can be used to push back against some recent allegations of cognitive bias in LLMs. This section takes up the positive project of pairing each strategy with a concrete recommendation for study design in order to increase the resilience of benchmarks to each ecological rationality strategy (Table 2).

7.1 Rational analysis

When normative standards are contested, it is not enough for normative standards to be stipulated by experimenters. Rigorous methodologies are needed to formulate and justify normative standards. One promising tradition is rational analysis, which derives normative standards from precise computational models (Anderson 1990; Chater and Oaksford 1999).

Strategy	Recommendation	Description of recommendation
Broadening norms	Rational analysis	Derive rational behavior from computational models of goals and environments.
Clarifying heuristics	Taxonomy-based disambiguation	Disambiguate heuristics and biases along existing taxonomies.
Ecological validity	Representative design	Sample stimuli from natural environments.

Table 2: Recommendations to mitigate challenges in bias experiments

Rational analysis begins with a precise computational specification of the goal of a cognitive system. For example, users of anchoring and adjustment may seek to achieve reliable estimates and avoid costly computations, and the relative value of each goal can be elicited in choice experiments or derived from incentives (Lieder et al. 2018b). Rational analysis next formally specifies the environment of problems facing the cognitive system. For example, users of anchoring and adjustment may face estimation problems drawn from formally specifiable domains. Goal- and environment-specifications are used to derive optimal behavior in terms of environment-specific conduciveness to goal-satisfaction given decision-theoretic criteria, providing a precise normative standard to assess observed behaviors. For example, specifying the relative importance of achieving reliable estimates and avoiding costly computations generates precise recommendations for how many adjustments should be made using anchoring and adjustment in a given environment (Lieder et al. 2018b).

More generally, recent advances in rational analysis have generated precise computational models of the conditions under which biases such as anchoring (Lieder et al. 2018b), probability matching (Icard 2018; Vul et al. 2014), and overrepresentation of extreme events (Lieder et al. 2018a) emerge from rational goal- and environment-pressures. These existing rational analyses provide normative standards that can be adapted to the goals and demands of LLMs. Comparing observed behavior to the resulting normative models derived from rational analysis can provide a principled and precise normative standard by which to assess the presence of cognitive bias in LLMs (Binz et al. 2022; Callaway et al. 2017; Krueger et al. 2024).

7.2 Taxonomy-based clarification

In response to the need to clarify biases, *taxonomy-based clarification* can be used to disambiguate heuristics and the resulting biases by disambiguating them along existing taxonomies. Section 5 illustrated taxonomy-based clarification using the distinction between subjective and objective availability. This distinction is used to demarcate potentially unproblematic findings, such as objective availability, from potentially problematic findings, such as subjective availability.

More generally, many of the most-studied heuristics and biases have been disambiguated along relevant taxonomies. For example, anchoring bias is now thought to split into at least seven different effects (Schley and Weingarten 2025) including response biases (Frederick and Mochon 2012), hypothesis-directed memory search (Mussweiler and Strack 1997) and priming effects (Wilson et al. 1996), with genuine anchoring and adjustment among the rarest (Epley and Gilovich 2001; Lieder et al. 2018b). Some of these effects may be less problematic than others.

Consider response biases (Frederick and Mochon 2012). These anchoring effects are not changes in beliefs, but rather changes in the response scales used to report beliefs. If I show you a wolf, then ask you to show me an animal that weighs approximately 1,000 pounds, you will show me an animal heavier than 1,000 pounds in order to emphasize that it is heavier than a wolf. But if I began by showing you an earthworm, you would not do this, because it is not necessary to emphasize the salient contrast. Many proposed norms of helpful communication (Grice 1975) would treat such response biases as potentially unproblematic, since they are not distortions in belief but rather attempts to helpfully communicate contextually salient differences.

Once biases are disambiguated along existing taxonomies, experiments can be structured to distinguish between each category in the taxonomy. For example, response biases could be separated

from genuine belief changes through standard methods such as retrospective pretest-posttest design (Little et al. 2020). In this way, taxonomy-based disambiguation can help us to separate potentially unproblematic findings from potentially problematic biases and to generate testable strategies for separating them.

7.3 Representative design

Ecological rationality theorists propose that rational cognition involves using cues in proportion to their ecological validity: the extent that they indicate relevant quantities in task environments. Building on this suggestion, the tradition of *representative design* suggests that experiments should be created to reflect task environments in which LLMs are proposed for use, rather than invented environments in which natural cue validities are broken (Brunswik 1956; Dhimi et al. 2004). In particular, representative design involves randomly sampling stimuli from natural environments so that underlying statistical relationships are preserved. The suggestion is that cognitive bias experiments could be improved by randomly sampling stimuli from natural environments in this way.

For example, instead of positing (Lou and Sun 2026) a daily low price of \$326.00 and a daily high price of \$343.55 for Tesla stock, low and high prices could be randomly sampled from recent stock-market data. Models that failed to correct for average daily fluctuations between high- and low-prices could then be said to exhibit anchoring bias by failing to correct from the anchoring price presented to them by incorporating market data about price fluctuations. However, models that assumed average daily fluctuations rather than abnormally large fluctuations would not be shown to exhibit anchoring bias. On the contrary, correcting for typical rather than atypical fluctuations is precisely the type of ecologically valid cue usage that representative design aims to measure.

8 Alternative Positions

Defenders of the heuristics and biases program frequently accept the ecological rationality critique in principle. For example, the editorial introduction to a leading anthology concedes:

There is . . . [one] critique to which researchers in the heuristics and biases tradition must plead ‘no contest’ or even ‘guilty’. This is the criticism that studies in this tradition have paid scant attention to assessing the overall ecological validity of heuristic processes. (Gilovich and Griffin 2002, p. 8)

However, they protest, incorporating considerations of ecological rationality is simply too demanding to be practical in typical experiments. Gilovich and Griffin focus for concreteness on the representativeness heuristic “according to which the subjective probability of an event, or a sample, is determined by the degree to which it: (i) is similar in essential characteristics to its parent population; and (ii) reflects the salient features of the process by which it is generated” (Kahneman and Tversky 1972, p. 430). In this case, Gilovich and Griffin hold:

Assessing the ecological validity of the representativeness heuristic would involve identifying a universe of relevant objects (e.g. every scholar in the engineering and law faculties at a given university), and then correlating the outcome value for each object (e.g. membership in either faculty) with the value of the cue variable for each object (e.g. relative similarity to the prototype of each faculty). . . . This Herculean task has not attracted researchers in the heuristics and biases tradition. (Gilovich and Griffin 2002, p. 8)

This complaint is well taken, particularly in the study of human cognition at the turn of the century. At the same time, two points deserve emphasis.

First, assessing ecological validity is not always impractical. Existing studies of ecological rationality often provide detailed empirical assessments of exactly the kinds of empirical correlations whose estimation Gilovich and Griffin deem to be a Herculean task (Benartzi and Thaler 2001; Gigerenzer et al. 1991; Serwe and Frings 2006). Moreover, advances in artificial intelligence are precisely the kinds of improvements needed to increase the feasibility of collecting and analyzing the large datasets needed for proper assessments of ecological validity. It is, to continue Gilovich and Griffin’s example, no longer a daunting task to identify the law and engineering professors at a

given university, classify them by department and rank them by similarity to a given prototype. In this way, the turn towards studying cognitive bias in LLMs is sometimes precisely what is needed to increase the feasibility of incorporating insights from ecological rationality.

Second, when it is not possible to accommodate insights from ecological rationality, it is important to acknowledge this limitation when interpreting experimental results. Gilovich and Griffin themselves express sympathy with the ecological contention that judgments are often rational in the environments where they are proposed for use, holding that “Researchers in [the heuristics and biases] tradition clearly share a set of assumptions: the ecological validities are probably high, the heuristics are generally useful, but common and profoundly important exceptions are to be found” (Gilovich and Griffin 2002, pp. 8-9). This point of convergence is welcome, but bears emphasis. When biases are found in isolated experimental contexts, it may be warranted to treat these biases as exceptions, even profoundly important exceptions, to the ecological rationality of LLM reasoning.

However, without assessing the ecological validity of LLM reasoning, we are not warranted in drawing further conclusions about the quality of LLM reasoning across the board. Indeed, insofar as leading advocates of the heuristics and biases tradition share with ecological rationality theorists the view that reasoning shows a high degree of ecological rationality, then it may be important to contextualize deviations from ecological rationality as exceptions, or even profoundly important exceptions, rather than the norm.

References

- Andersen, Torben, Bollerslev, Tim, Diebold, Francis X., and Ebens, Heiko. 2001. “The distribution of realized stock return volatility.” *Journal of Financial Economics* 61:43–76.
- Anderson, John. 1990. *The adaptive character of thought*. Lawrence Erlbaum Associates.
- Benartzi, Shlomo and Thaler, Richard. 2001. “Naive diversification strategies in defined contribution saving plans.” *American Economic Review* 91:79–98.
- Binz, Marcel, Gershman, Samuel, Schulz, Eric, and Endres, Domink. 2022. “Heuristics from bounded meta-learned inference.” *Psychological Review* 129:1042–77.
- Binz, Marcel and Schulz, Eric. 2023. “Using cognitive psychology to understand GPT-3.” *Proceedings of the National Academy of Sciences* 120:e2218523120.
- Brunswick, Egon. 1956. *Perception and the representative design of psychological experiments*. University of California Press.
- Callaway, Frederick, Gul, Sayan, Krueger, Paul, Griffiths, Thomas, and Lieder, Falk. 2017. “Learning to select computations.” arXiv 1711.06892.
- Chater, Nick and Oaksford, Mike. 1999. “Ten years of the rational analysis of cognition.” *Trends in Cognitive Sciences* 3:57–65.
- Dasgupta, Ishita, Lampinen, Andrew K., Chan, Stephanie C. Y., Creswell, Antonia, Kumaran, Dharshan, McClelland, James L., and Hill, Felix. 2022. “Language models show human-like content effects on reasoning.” arXiv, 2207.07051.
- Dhami, Mandeep, Hertwig, Ralph, and Hoffrage, Ulrich. 2004. “The role of representative design in an ecological approach to cognition.” *Psychological Bulletin* 130:959–88.
- Echterhoff, Jessica, Liu, Yao, Alessa, Abeer, McAuley, Julian, and He, Zexue. 2024. “Cognitive bias in decision-making with LLMs.” *Findings of the Association for Computational Linguistics* 12640–53.
- Elqayam, Shira and Over, David. 2013. “New paradigm psychology of reasoning: An introduction to the special issue edited by Elqayam, Bonnefon, and Over.” *Thinking and Reasoning* 19:249–65.
- Epley, Nicholas and Gilovich, Thomas. 2001. “Putting adjustment back in the anchoring and adjustment heuristic: Differential processing of self-generated and experimenter-provided anchors.” *Psychological Science* 12:391–6.

- Eriksson, Maria, Purificato, Erasmo, Noroozian, Arman, Vinagre, Joao, Chaslot, Guillaume, and Gómez, Emilia. 2025. "Can we trust AI benchmarks? An interdisciplinary review of current issues in AI evaluation." arXiv 2502.06559.
- Federal Highway Administration. 2024. "Annual vehicle-miles of travel, 1980-2024 by functional system national summary." U.S. Interstate System (AADT) data, <https://www.fhwa.dot.gov/policyinformation/statistics/2024/vm202.cfm>.
- Frederick, Shane and Mochon, Daniel. 2012. "A scale distortion theory of anchoring." *Journal of Experimental Psychology: General* 141:124–33.
- Gigerenzer, Gerd. 1991. "How to make cognitive illusions disappear: Beyond 'heuristics and biases'." *European Review of Social Psychology* 2:83–115.
- . 1996. "On narrow norms and vague heuristics: A reply to Kahneman and Tversky." *Psychological Review* 103:592–6.
- . 2019. "Axiomatic rationality and ecological rationality." *Synthese* 194:3547–64. doi:10.1007/s11229-019-02296-5.
- Gigerenzer, Gerd and Gaissmaier, Wolfgang. 2011. "Heuristic decision making." *Annual Review of Psychology* 62:451–82.
- Gigerenzer, Gerd and Goldstein, Daniel. 2011. "The recognition heuristic: A decade of research." *Judgment and Decision Making* 6:100–21.
- Gigerenzer, Gerd, Hoffrage, Ulrich, and Kleinbölting, Heinz. 1991. "Probabilistic mental models: A Brunswikian theory of confidence." *Psychological Review* 98:506–528.
- Gigerenzer, Gerd and Selten, Reinhard (eds.). 2001. *Bounded rationality: The adaptive toolbox*. MIT Press.
- Gilovich, Thomas and Griffin, Dale. 2002. "Heuristics and biases: Then and now." In Thomas Gilovich, Dale Griffin, and Daniel Kahneman (eds.), *Heuristics and biases: The psychology of intuitive judgment*, 1–18. Cambridge University Press.
- Gilovich, Thomas, Griffin, Dale, and Kahneman, Daniel (eds.). 2002. *Heuristics and biases: The psychology of intuitive judgment*. Cambridge University Press.
- Goldstein, Daniel and Gigerenzer, Gerd. 1999. "The recognition heuristic: How ignorance makes us smart." In Gerd Gigerenzer, Peter Todd, and The ABC Research Group (eds.), *Simple heuristics that make us smart*, 37–58. Oxford University Press.
- . 2002. "Models of ecological rationality: The recognition heuristic." *Psychological Review* 109:75–90.
- Grice, H.P. 1975. "Logic and conversation." In Peter Cole and Jerry L. Morgan (eds.), *Syntax and semantics 3: Speech acts*, 41–58. Academic Press.
- Hertwig, Ralph and Pachur, Thorsten. 2005. "Judgments of risk frequencies: Tests of possible cognitive mechanisms." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 31:621–42.
- Hester, Neil and Hehman, Eric. 2023. "Dress is a fundamental component of person perception." *Personality and Social Psychology Review* 27:414–33.
- Hong, Lu and Page, Scott. 2004. "Groups of diverse problem solvers can outperform groups of high-ability problem solvers." *Proceedings of the National Academy of Sciences* 101:16385–9.
- Horton, John. 2023. "Large language models as simulated economic agents: What can we learn from homo silicus?" National Bureau of Economic Research Working Paper 31122, <https://www.nber.org/papers/w31122>.
- Icard, Thomas. 2018. "Bayes, bounds, and rational analysis." *Philosophy of Science* 85:79–101.

- Jones, Erik and Steinhardt, Jacob. 2022. “Capturing failures of large language models via human cognitive biases.” In *NIPS’22: Proceedings of the 36th International Conference on Neural Information Processing Systems*, 11785–99.
- Kahneman, Daniel, Slovic, Paul, and Tversky, Amos (eds.). 1982. *Judgment under uncertainty: Heuristics and biases*. Cambridge University Press.
- Kahneman, Daniel and Tversky, Amos. 1972. “Subjective probability: A judgment of representativeness.” *Cognitive Psychology* 3:430–54.
- Knipper, R. Alexander, Knipper, Charles S., Zhang, Kaiqi, Sims, Valerie, Bowers, Clint, and Karmaker, Santu. 2025. “The bias is in the details: An assessment of cognitive bias in LLMs.” arXiv 2509.22856.
- Krueger, Paul, Callaway, Frederick, Gul, Sayan, Griffiths, Thomas, and Lieder, Falk. 2024. “Identifying resource-rational heuristics for risky choice.” *Psychological Review* 131:905–51.
- Lieder, Falk, Griffiths, Thomas, and Hsu, Ming. 2018a. “Overrepresentation of extreme events in decision making reflects rational use of cognitive resources.” *Psychological Review* 125:1–32.
- Lieder, Falk, Griffiths, Thomas, Huys, Quentin, and Goodman, Noah. 2018b. “The anchoring bias reflects rational use of cognitive resources.” *Psychonomic Bulletin and Review* 25:322–49.
- Lin, Ruixi and Ng, Hwee Tou. 2023. “Mind the biases: Quantifying cognitive biases in language model prompting.” In *Findings of the Association for Computational Linguistics: ACL 2023*, 5269–81. Toronto, Canada: Association for Computational Linguistics.
- List, Christian and Goodin, Robert E. 2001. “Epistemic democracy: Generalizing the Condorcet jury theorem.” *Journal of Political Philosophy* 9:277–306.
- Little, Todd, Chang, Rong, Gorrall, Britt, Waggenspack, Luke, Fukuda, Eriko, Allen, Patricia, and Noam, Gil. 2020. “The retrospective pretest-posttest design redux: On its validity as an alternative to traditional pretest-posttest measurement.” *International Journal of Behavioral Development* 44:175–83.
- Lou, Jiaxu and Sun, Yifan. 2026. “Anchoring bias in large language models: An experimental study.” *Journal of Computational Social Science* 9:11.
- Lyu, Yougang, Ren, Shijie, Feng, Yue, Wang, Zihan, Chen, Zhumin, Ren, Zhaochun, and de Rijke, Maarten. 2025. “Self-adaptive cognitive debiasing for large language models in decision-making.” arXiv 2504.04141.
- Malberg, Simon, Poletukhin, Roman, Schuster, Carolin M., and Groh, Georg. 2025. “A comprehensive evaluation of cognitive biases in LLMs.” arXiv 2410.15413.
- Mussweiler, Thomas and Strack, Fritz. 1997. “Explaining the enigmatic anchoring effect: Mechanisms of selective accessibility.” *Journal of Personality and Social Psychology* 73:437–46.
- Nguyen, Jeremy K. 2024. “Human bias in AI models? Anchoring effects and mitigation strategies in large language models.” *Journal of Behavioral and Experimental Finance* 43:100971.
- Oaksford, Mike and Chater, Nick. 1994. “A rational analysis of the selection task as optimal data selection.” *Psychological Review* 101:608–31.
- . 2007. *Bayesian rationality: The probabilistic approach to human reasoning*. Oxford University Press.
- Raji, Inioluwa Deborah, Bender, Emily, Paullada, Amandalyne, Ednton, Emily, and Hanna, Alex. 2021. “AI and the everything in the whole wide world benchmark.” In J. Vanschoren and S. Yeung (eds.), *Thirty-fifth Conference on Neural Information Processing Systems, Datasets and Benchmarks Track*.

- Reuel, Anka, Hardy, Amelia, Smith, Chandler, Lamparth, Max, Hardy, Malcolm, and Kochenderfer, Mykel. 2024. "BetterBench: Assessing AI benchmarks, uncovering issues, and establishing best practices." In *NIPS '24: Proceedings of the 38th International Conference on Neural Information Processing Systems*, 21763–813.
- Rips, Lance. 1994. *The psychology of proof: Deductive reasoning in human thinking*. MIT Press.
- Schley, Dan and Weingarten, Evan. 2025. "50 years of anchoring: A meta-analysis and meta-study of anchoring effects." ms.
- Schmidgall, Samuel, Harris, Carl, Essien, Ime, Olshvang, Daniel, Rahman, Tawsifur, Kim, Ji Woong, Ziaei, Roijin, Eshraghian, Jason, Abadir, Peter, and Chellappa, Rama. 2024. "Evaluation and mitigation of cognitive biases in medical language models." *npj Digital Medicine* 7:295.
- Schwarz, Norbert and Vaughn, Leigh Ann. 2002. "The availability heuristic revisited: Ease of recall and content of recall as distinct sources of information." In Thomas Gilovich, Dale Griffin, and Daniel Kahneman (eds.), *Heuristics and biases: The psychology of intuitive judgment*, 103–19. Cambridge University Press.
- Serwe, Sascha and Frings, Christian. 2006. "Who will win Wimbledon? The recognition heuristic in predicting sports events." *Behavioral Decision Making* 19:321–332.
- Shaikh, Ammar, Dandekar, Raj Abhijit, Panat, Sreedath, and Dandekar, Rajat. 2024. "CBEval: A framework for evaluating and interpreting cognitive biases in LLMs." arXiv 2412.03605.
- Shannon, Claude. 1948. "A mathematical theory of communication." *The Bell System Technical Journal* 27:379–423.
- Shiffrin, Richard and Mitchell, Melanie. 2023. "Probing the psychology of AI models." *Proceedings of the National Academy of Sciences* 120:e2300963120.
- Song, Zheng, Wang, Meng, Hua, Xian-Sheng, and Yan, Shuicheng. 2011. "Predicting occupation via human clothing and contexts." *Proceedings of the 2011 International Conference on Computer Vision* 1084–91.
- Surowiecki, James. 2004. *The wisdom of crowds: Why the many are smarter than the few and how collective wisdom shapes business, economies, societies and nations*. Doubleday.
- Talbot, Alaina N. and Fuller, Elizabeth. 2023. "Challenging the appearance of machine intelligence: Cognitive bias in LLMs and Best Practices for Adoption." arXiv 2304.01358.
- Todd, Peter and Gigerenzer, Gerd. 2012. *Ecological rationality: Intelligence in the world*. Oxford University Press.
- Treynor, Jack L. 1987. "Market efficiency and the bean jar experiment." *Financial Analysts Journal* 43:50–3.
- Tversky, Amos and Kahneman, Daniel. 1973. "Availability: A heuristic for judging frequency and probability." *Cognitive Psychology* 5:207–32.
- . 1974. "Judgment under uncertainty: Heuristics and biases." *Science* 185:1124–31. doi:10.1126/science.185.4157.1124.
- Vul, Edward, Goodman, Noah, Griffiths, Thomas L., and Tenenbaum, Joshua B. 2014. "One and done? Optimal decisions from very few samples." *Cognitive Science* 38:599–637.
- Wason, Peter C. 1968. "Reasoning about a rule." *Quarterly Journal of Experimental Psychology* 20:273–81.
- Wilson, Timothy, Houston, Christopher, Etling, Kathryn, and Brekke, Nancy. 1996. "A new look at anchoring effects: Basic anchoring and its antecedents." *Journal of Experimental Psychology: General* 125:387–402.
- Yahoo Finance. 2026. "Tesla (TSLA) Historical Data [Data Table]." Retrieved Mar 18, 2026 from <https://finance.yahoo.com/quote/TSLA/history/>.