Alternatives to Bayesian Updating

Pietro Ortoleva[1]

[1]Department of Economics and SPIA, Princeton University, Princeton, NJ, USA, 08544; email: pietro.ortoleva@princeton.edu; ORCID: 0000-0002-5943-6621.

Annual Review of Economics 2022.
15:1–28

https://doi.org/10.1146/annurev-economics-100223-050352

Keywords

Bayes' rule, non-Bayesian Behavior, base-rate neglect, conservatism, cognitive noise, hypothesis testing

Abstract

We discuss models of updating that depart from Bayes' rule even when it is well-defined. After reviewing Bayes' rule and its foundations, we begin our analysis with models of non-Bayesian behavior arising from a bias, a pull towards suboptimal behavior due to a heuristic or a mistake. Next, we explore deviations caused by individuals questioning the prior probabilities they initially used. We then consider non-Bayesian behavior resulting from the optimal response to constraints on perception, cognition, or memory, as well as models based on motivated beliefs or distance minimization. Finally, we briefly discuss models of updating after zero probability events.

Contents

1. Introduction

The rule on how to update beliefs upon the receipt of new information introduced in the 18th century by Thomas Bayes is one of the fundamental normative benchmarks of decision-making. It prescribes a method to update beliefs that adheres to the principles of probability theory and is the unique rule that guarantees several normative properties, like dynamic consistency. This normative appeal is coupled with a positive one: Bayes' rule provides a very useful framework for understanding a broad set of behaviors and, in several instances, even a reasonable approximation of the actual evolution of beliefs. Bayes' rule is at the core of any treatment of information in many disciplines, including economics—from game theory to rational expectations in macroeconomics.

Despite its appeal, the updating procedure suggested by Bayes has several known limitations. Particularly relevant for a model of decision-making, robust evidence shows that humans consistently and substantially depart from its prescriptions. To account for these deviations, many alternative updating rules have been developed in several disciplines.

The goal of this article is to review the most prominent alternatives to Bayesian updating, with a specific focus on those most relevant to economics. We cover several different kinds of departures from the normative benchmark, from biases to to cognitive limitations. While some deviations can be naturally classified as mistakes, for others it will be less clear. We also briefly mention models that extend Bayes' rule to capture the reaction to zero probability events.

Before embarking on our discussion, we must clarify our scope. We will cover merely a slice of the colossal literature on belief updating in economics, psychology, neuroscience, philosophy, mathematics, and statistics. We focus exclusively on models of updating, with only indirect references to empirical evidence; in economics, we refer to Benjamin (2019) for a recent review. We concentrate solely on theories in which *i*) individuals possess a well-defined belief before and after receiving information, and follow subjective expected utility when necessary; *ii*) there exists a unique and precise way to interpret the new information, which can be formalized as an event in the appropriately defined state space; and *iii*) beliefs are not incorrect or misspecified. Only at the end, we very briefly mention models that extend Bayes' rule to capture reactions to zero probability events. Therefore, we do not discuss updating when subjects do not have unique priors or posteriors, such as with ambiguity aversion,[1] or when beliefs are not formed, as in reinforcement learning; with more general types of information, such as vague, ambiguous, or imprecise information or information of the form "Event A is more likely than event B;"[2] or with misspecified beliefs, such as with overconfidence or overly simple models.[3] We only briefly mention the broader implications of non-Bayesian learning.[4] Even within our restricted focus, we emphasize models that have been particularly influential in economics or have the potential to become so. Due to space limitations, instead of delving deeply into a few models, we cover more but be more brief, aiming to convey core insights while leaving details to the original papers.

These pages are organized as follows. In Section 2, we introduce the framework and review Bayes' rule, highlighting some of the reasons for its appeal. Moving on to models of non-Bayesian behavior, Section 3 examines models that attribute such behavior to a *bias*. In Section 4, we explore deviations from Bayes' rule due to individuals questioning their priors. Section 5 delves into models that link non-Bayesian behavior to limitations in memory, perception, or generic cognitive noise, while in Section 6, we discuss models that connect it to utility derived from beliefs or updating based on minimizing a distance. Section 7 briefly reviews models that extend Bayes' rule to capture reactions to information to which the prior assigned probability zero. We conclude in Section 8 with a discussion on the state of the literature.

2. Framework and Bayes' Rule

Consider a non-empty set Ω of states of the world with σ-algebra Σ, a probability measure π—what we refer to as our "prior"—and an event $A \in \Sigma$—our "information." Denote by $\Delta(\Omega)$ the set of probability distributions. We are interested in *updating rules*: rules that prescribe how to update π if we are given the information that the true state of the world lies inside a given event A. Formally, an updating rule maps each prior and event to an updated belief, that is, maps from $\Delta(\Omega) \times \Sigma$ to $\Delta(\Omega)$. Denote by π_A this updated belief, often called the "posterior."

[1] Updating under ambiguity is particularly interesting, as it presents a case where non-Bayesian behavior may be natural; see Gilboa & Marinacci (2013) for a review.

[2] These cases have been extensively studied in several disciplines and include classical approaches such as Jeffrey's rule or AGM belief revisions (Alchourrón et al. 1985, Diaconis & Zabell 1986). For recent contributions in economics, see Dominiak et al. (2021), Zhao (2022).

[3] Learning with misspecified beliefs is a classical problem in many disciplines, and recently has received particular attention in economic theory; Section 5 includes a brief discussion. We do include approaches in which individuals act as if Bayesian with the wrong model (what Rabin 2013 calls *quasi-Bayesian*) when that can be seen as a way to represent the agent's mistake in updating (e.g., Rabin 2002's Law of Small Numbers; see Section 3.1.4).

[4] A few papers explore implications in areas such as social learning (DeGroot 1974, Molavi et al. 2018) or persuasion and information design (Galperti 2019, de Clippel & Zhang 2022).

Several papers in economics and psychology adopt a special case of this setup with a set of payoff relevant states S, a set of possible messages M, a prior probability p over S, and a conditional probability of M given S; we are interested in the updated beliefs over S given message $m \in M$, which we denote p_m. This setup may appear different because the message m is not a subset of the state space S, but we only need to define the state space correctly: Letting $\Omega = S \times M$ and appropriately constructing Σ and π on Ω, it is easy to see that this is a special case in which the underlying state space has a product structure, and we are interested only in updating after a specific subset of possible events and only in the beliefs on a specific dimension of the state space. Because the general framework is common in formal treatments and this special case in some modeling, we consider both.

2.1. Bayes' rule

The universal benchmark for updating rules is Bayes' rule,[5] according to which

$$\pi_A(B) = \frac{\pi(A \cap B)}{\pi(A)} =: \pi_A^{\mathrm{BU}}(B)$$

for all $B \in \Sigma$ and for all A such that $\pi(A) \neq 0$. Since $\pi(A \cap B) = \pi(A|B)\pi(B)$, this is equivalent to

$$\pi_A(B) = \frac{\pi(A|B)\pi(B)}{\pi(A)}.$$

This formula follows directly from the definition of conditional probabilities: If we require $\pi_A(B)$ to coincide with $\pi(B|A)$, we immediately obtain $\frac{\pi(A\cap B)}{\pi(A)}$. This constitutes a strong argument for its role as a normative benchmark: Insofar as updating coincides with conditional probabilities, it *must* be Bayesian.

Bayes' rule is also intuitive. Suppose we are told the state lies in A; what are the chances that it also belongs to B? If $A \cap B = \emptyset$, they should be zero—as Bayes' rule prescribes. What if $A \cap B \neq \emptyset$? Then, to compute the likelihood of B, only the states in $A \cap B$ remain relevant; hence, a good starting point for $\pi_A(B)$ is $\pi(A \cap B)$. But this cannot be the final answer: For π_A to be a probability measure, we must have $\pi_A(\Omega) = 1$; thus, we need to normalize the probabilities. Since $\pi(A \cap \Omega) = \pi(A)$, dividing everything by $\pi(A)$ gives the correct normalization. We obtain $\pi_A(B) = \frac{\pi(A\cap B)}{\pi(A)}$, that is, Bayes' rule.

Another distinguishing feature of Bayes' rule is that updated probabilities maintain the correct relative likelihoods. Take $B, C \in \Sigma$ and suppose $B, C \subset A$. Before information, their relative likelihood is $\frac{\pi(B)}{\pi(C)}$. Because $B, C \subset A$, the relative likelihood should not change with the information that the true state is in A. Indeed, under Bayes' rule, $\frac{\pi_A(B)}{\pi_A(C)} = \frac{\frac{\pi(A\cap B)}{\pi(A)}}{\frac{\pi(A\cap C)}{\pi(A)}} = \frac{\pi(A\cap B)}{\pi(A\cap C)} = \frac{\pi(B)}{\pi(C)}$, as sought. If B and C are not subsets of A, we want relative likelihoods to maintain a ratio that is equal to the measures remaining after we take the intersection with A—again, precisely what Bayes' rule prescribes.

[5] Typically ascribed to the English revered Thomas Bayes (1701-1761), the updating rule did not appear until after his death, in 1763, in an article edited by Richard Price (1723-1791), who may have contributed. The same rule is believed to have been discovered independently by Pierre-Simon, Marquis de Laplace (1749-1827), and published in 1774 (McGrayne 2011). Interestingly, Laplace is also known to be the first to document gamblers' fallacy, noting that if early in a month more boys are born, people expect more girls to be born in the rest of the month.

Despite these advantages, one core limitation of Bayes' rule is already apparent: Updating is not defined after zero probability events. Indeed, the rule requires $\pi(A) > 0$. This is unavoidable: The rule aims to preserve relative likelihoods of all subsets of A, but when $\pi(A) = 0$, all such likelihoods are "flattened" to zero; there is no relative ranking to preserve. Put another way, the prior π does not include any information on the relative likelihood of subsets of all events of probability zero—and therefore, it cannot offer guidance on how to construct a posterior. The only way to construct such a posterior is to invoke directions not contained in the original prior, departing from a fundamental tenant of Bayes' rule.

2.2. On its appeal

We now briefly review some additional well-known reasons for Bayes' rule appeal.[6]

2.2.1. It derives from the rules of probability theory. We have seen that if individuals update beliefs following conditional probabilities, they must follow Bayes' rule. This is probably the main reason why many scholars consider Bayesian updating a manifestation of rationality and any deviation from it as evidence of mistakes/biases/limitations.

However, there is an important caveat to this argument. In some instances, the initial belief π is objectively given to the decision-maker: For example, in many lab experiments, subjects are told the correct prior probability distribution. In these cases, indeed, insofar as decision-makers trust this objective prior and the way probabilities work, they should follow Bayes' rule. But this normative argument becomes significantly weaker when the prior probability is *subjective*, that is, when individuals are not given a prior belief but need to form one themselves. For example, people may have a prior belief about the behavior of the stock market, but this is quite different from an objectively correct belief; it is subjectively assessed and likely to be wrong, and individuals know it. In these cases, one may argue that reasonable individuals may well deviate from Bayes' rule. For example, an individual who receives information that was very unlikely given her chosen prior—e.g., an unexpected financial crisis—may decide, quite reasonably, that they may have used the wrong prior in the first place; because the prior was subjectively chosen, this seems like something they are entitled to do without being considered "irrational." But this may lead them to a (subjective) posterior that is not the Bayesian update of the (subjective) prior; this is the approach taken by the models in Section 4. In general, when beliefs are subjective, it is not obvious that non-Bayesian behavior should be classified as a mistake; alas, in the vast majority of decision problems outside the laboratory, subjective beliefs are all that we have. While this point is known, it is also typically ignored, and the normative appeal of Bayes' rule is rarely questioned. We find that surprising. For a related discussion in economics, see Epstein (2006).

2.2.2. Behavioral Foundations and Dynamic Consistency. Next, we discuss the behavioral foundations of Bayes' rule. Under expected utility, preferences are directly related to beliefs, and we study how preferences change as we change information. We will show that there

[6] This is but a brief review. See also Williams (1980) for a relation to the Principle of Minimal Information, Jaynes (2003) for a textbook that discusses the relation to logic, or Perea (2009) for a relation to dynamic consistency in games.

is a tight relationship between the Dynamic Consistency of preferences and the Bayesian updating of beliefs.[7]

Assume that Ω is finite and consider a set of prizes X, e.g., monetary amounts. Following the literature, we call an *act* any map from Ω to X, that is, a map connecting a prize to each state; denote by $\mathcal{F}$ the set of all acts. Clearly, preferences over acts are related to beliefs over states, and we want to study how preferences over acts change with information. To this end, fix an event A and consider two preference relations on $\mathcal{F}$: $\succcurlyeq$ denotes the preferences before information, and $\succcurlyeq_A$ denotes the preferences after the agent has been told that the true state lies in A.[8] Assume that both $\succcurlyeq$ and $\succcurlyeq_A$ admit expected utility representations with utility u over X and beliefs π and π_A, respectively, and that $\pi(A), \pi_A(A) \neq 0$.[9]

We are now ready to introduce our two key behavioral postulates. Consider two acts, f and g, and suppose that $f(\omega) = g(\omega)$ for all $\omega \in A$: Inside A, f and g are the same. Suppose we tell the individual that the state lies in A. Then, the individual should be indifferent between f and g. This is *Consequentialism.*[10]

Axiom 1 (Consequentialism)**.** *For any $f, g \in \mathcal{F}$, if $f(\omega) = g(\omega)$ for all $\omega \in A$, then $f \sim_A g$.*

To introduce our next postulate, we need an extra bit of notation: For any $f, g \in \mathcal{F}$, denote fAg the act that is equal to $f(\omega)$ if $\omega \in A$ and $g(\omega)$ otherwise. Suppose $fAg \succeq g$. Because fAg and g differ only within A, $fAg \succeq g$ indicates that our agent prefers what f returns to what g returns within A—what they return outside is the same and "cancels out." But then, it stands to reason that if we tell the agent that the true state lies in A, the agent should prefer f to g, that is, $f \succcurlyeq_A g$. The converse should also hold true: If f is preferred to g knowing that A is true, fAg should be preferred to g before information. This gives us our final postulate.

Axiom 2 (Dynamic Consistency)**.** *If $\pi(A) > 0$, for any $f, g \in \mathcal{F}$, $fAg \succeq g \;\Leftrightarrow f \succcurlyeq_A g$.*

This postulate imposes a form of consistency in preferences with information: After receiving information A, preferences cannot change for those acts that differed only inside A. Indeed, this is exactly the notion of dynamic consistency. These two postulates give us Bayesian updating (for the proof, see Ghirardato 2002).

Theorem 1. *$\succcurlyeq$ and $\succcurlyeq_A$ satisfy Consequentialism and Dynamic Consistency if and only if $\pi_A = \pi_A^{\mathrm{BU}}$.*

The behavioral foundations for Bayes' rule are Consequentialism—akin to the agent believing the information—and Dynamic Consistency—akin to preferences not changing for acts that are identical outside the information. Theorem 1 shows that not only does Bayes' rule satisfy these normatively appealing postulates: It is the *only* updating rule that does so. Any other updating rule must violate at least one of these postulates.

[7] This discussion follows Ghirardato (2002), but simplifies the analysis; see that paper and to the many references therein for a more complete treatment and for proofs.

[8] A preference relation R on a set Y is a binary relation on Y ($R \subseteq Y \times Y$) that is complete (for all $x, y \in Y$, either xRy or yRx or both) and transitive (for all $x, y, z \in Y$, xRy and yRz imply xRz). As usual, we adopt $\sim$ and $\succ$ to denote the symmetric and anti-symmetric parts.

[9] A preference relation R on $\mathcal{F}$ admits an expected utility representation if there exist $u : X \to \mathbb{R}$ and $\pi \in \Delta(\Omega)$ s.t. for all $f, g \in \mathcal{F}$, fRg if and only if $\sum_{\omega \in \Omega} \pi(\omega) u(f(\omega)) \geq \sum_{\omega \in \Omega} \pi(\omega) u(g(\omega))$. See Kreps (1988) for a textbook treatment that includes foundations.

[10] This requirement may appear almost obvious, almost a test of whether the individual believes the information. Yet, several models discussed in the next sections will violate it.

2.2.3. The Bayesian paradigm in cognitive sciences and the Bayesian Brain. While a vast literature documents systematic deviation from Bayes' rule in updating tasks, a separate and growing literature in cognitive psychology and neuroscience has instead argued that the human mind is close to being "Bayes optimal" in a wide range of environments, like perception, motor control, memory, language, and several cognitive tasks (Ernst & Banks 2002, Körding & Wolpert 2004, Tenenbaum et al. 2006, Griffiths & Tenenbaum 2006, Oaksford & Chater 2007, Doya 2007, Griffiths et al. 2008). While choices are, of course, frequently not accurate, this literature suggests that the mind does properly account for the variability and noise in the information it receives and forms "Bayesian optimal" behavior accounting for such noise; this leads some authors to argue that behavior may be more rational than previously thought (Oaksford & Chater 2007, Preface). In general, the Bayesian approach has proved to be a useful and effective way of capturing how the mind constructs complex, abstract, and broadly correct models of reality using only the noisy data from perception, providing both convenient functional forms and empirical fit. This has led to the formulation of what some call the *Bayesian Brain* hypothesis, according to which the brain is an "inference machine" (Friston 2009, p. 294) that uses a generative model of the world to infer causes of sensations and make predictions (Knill & Pouget 2004, Doya 2007, Friston 2012, Sanborn & Chater 2016, Kording 2014). Some authors have connected this to the principles of free energy (Friston & Stephan 2007, Friston 2009).

There is, of course, no contradiction between non-Bayesian behavior in updating experiments and Bayes optimality in, say, vision. We have learned the latter with experience, without being given, and consciously thinking of, priors, noise of information, etc.; these aspects may be encoded (or not, see below) by our brain, but we do not consciously think of probabilities and noise. In updating experiments, instead, subjects receive information about the environment in the form of probabilities exogenously given—not experienced—and are asked to perform tasks like reporting a probability that requires them to obtain a probability representation. These are very different processes.

The Bayesian approach to cognitive sciences is not uncontroversial. First of all, there is a fundamental issue of tractability, given how quickly the dimensions of the probability space grow, particularly in high-level cognitive tasks (Sanborn et al. 2010, Tenenbaum et al. 2011, Jones & Love 2011, Bowers & Davis 2012); it is unclear how the brain implements or approximates computationally intractable updating (Kwisthout et al. 2011). Moreover, even within the framework, the issue remains open as to whether the brain does indeed encode a generative model, likelihood functions, and probabilities (the Bayesian Coding hypothesis) or whether this is an "as if" representation—because Bayesian responses are optimal, by construction any process that delivers optimality or near optimality can be represented as if it were Bayesian. The brain may also be able to perform Bayesian-like calculations without actually representing distributions; for example, some authors have suggested the "brain is a Bayesian sampler" (Sanborn & Chater 2016, p. 883). In general, there is also a distinction between the claim that behavior in perception or motor tasks is Bayesian and the claim that high-level reasoning is Bayesian, for which, thus far, the evidence is less substantial. It should also be emphasized that the Bayesian modeling of high-level cognitive tasks has many degrees of freedom in the choice of where noise is, the priors, the nature of signals, etc.; therefore, empirical fit alone may be insufficient.

This approach is becoming increasingly influential in economics (Woodford 2020). In Section 5.1, we discuss recent attempts to argue that non-Bayesian behavior may actually have a Bayesian rationale considering the noise in the system.

3. Biases in Updating

We are now ready to begin our discussion of alternatives to Bayes' rule, and we start with the traditional approach, which ascribes deviations from normative benchmarks to a "bias": a pull towards sub-optimal behavior due to heuristics or mistakes.

3.1. Biases in how Bayes' rule is applied

The most immediate way to model biases in updating is to posit that posteriors are computed using a modified updating rule. A very useful functional form is the so-called Grether rule (Grether 1980, 1992), introduced to interpret experimental data but later adopted as a model of behavior. Following the literature, for the rest of this section, we focus on the special case of payoff-relevant states in S and messages in M, both finite (see Section 2). According to the Grether rule, for each $s \in S$ and $m \in M$

$$p_m(s) = \frac{p(s)^\alpha p(m|s)^\beta}{\sum_{s' \in S} p(s')^\alpha p(m|s')^\beta}$$

for some $\alpha, \beta > 0$.[11] Clearly, $\alpha = \beta = 1$ returns Bayes' rule. The parameter α controls the weight on the prior, or, more precisely, how the posterior depends on how much the prior differentiates among elements of S;[12] instead, the parameter β controls the weight on the new evidence. As we will see, several conflicting combinations of parameters have been suggested in the literature: $\alpha < 1$ (base-rate neglect), $\alpha > 1$ (in some cases, confirmation bias), $\beta > 1$ (overreaction), $\beta < 1$ (underreaction). As pointed out by Benjamin (2019), we can have α and β both smaller (or larger) than 1: The values of both α and β matter.

3.1.1. Conservatism. One of the most classical deviations from Bayesian behavior is introduced in Edwards (1968), Phillips & Edwards (1966), and Phillips et al. (1966) and corresponds to the case $\alpha = 1$ and $\beta < 1$. Individuals correctly weigh the relative odds given by the prior but underweight new evidence. A classical interpretation ascribes it to difficulty in aggregating different sources (Slovic & Lichtenstein 1971).

Several papers in macroeconomics and finance derive the implications of a different model in which the updated belief is a convex combination of the prior and the Bayesian posterior (Mankiw & Reis 2002, Coibion & Gorodnichenko 2012, Bouchaud et al. 2019). Kovach (2021) shows that this updating behavior is axiomatized by a property called Dynamic conservatism. One implication is that if investors observe current cash flows that are informative about future ones, conservatism can generate the profitability anomaly, where past profits forecast future returns (Bouchaud et al. 2019). Intuitively, because investors are conservative in updating, their beliefs do not fully incorporate that high cash flows today imply high cash flows in the future, so they are likely to be positively surprised.

3.1.2. Confirmation bias. Another classic deviation is the tendency of individuals to overdraw from information that reinforces the current prior and not from evidence that weakens

[11] Virtually all the literature assumes S is binary and defines the Grether rule using likelihood ratios. Here, we consider the general case of finite S and extend the rule appropriately (but we cannot use likelihood ratios). The extension above guarantees that p_m is a probability measure.

[12] Indeed, α becomes irrelevant when the prior is uniform: if $p(s) = \frac{1}{|S|}$ for all $s \in S$, then $p_m(s) = \frac{p(s)^\alpha p(m|s)^\beta}{\sum_{s' \in S} p(s')^\alpha p(m|s')^\beta} = \frac{\frac{1}{|S|}^\alpha p(m|s)^\beta}{\sum_{s' \in S} \frac{1}{|S|}^\alpha p(m|s')^\beta} = \frac{p(m|s)^\beta}{\sum_{s' \in S} p(m|s')^\beta}$.

it, a broad umbrella of behavior that usually goes under the name *confirmation bias.*[13] Although the distinction is at times less clear, confirmation bias is different from conservatism because it implies that some evidence is interpreted incorrectly, which means that new information may be detrimental to learning and lead beliefs further astray.

A classical explanation in psychology is that individuals evaluate the consistency of the information with the hypothesis at hand but do not account for the consistency with other hypotheses (Fischhoff & Beyth-Marom 1983). Confirmation bias can be modeled using the Grether rule if β is made to depend on whether the evidence is confirmatory or not; Charness & Dave (2017) use one specification and Benjamin (2019, p. 155–156) discusses others. The most well-known model in economics is the simple and tractable formalization introduced in Rabin & Schrag (1999). Consider a binary payoff-relevant state and binary space of informative messages, and assume that individuals correctly apply Bayes' rule when messages confirm their current belief, that is, when the message supports the state they currently consider more likely. If, however, they receive a message against their current belief, with probability q, they "misperceive" this message as if it were the opposite, supportive message. Note that q is assumed to be fixed and independent of prior strength; moreover, individuals are severely biased in that they perceive the *opposite*, supporting evidence, as opposed to simply ignoring disconfirming messages.

With $q = 1$, only the first message matters—after it, beliefs will slowly converge to the state it pointed to. In general, $q > 0$ leads, on average, to overconfidence: Intuitively, beliefs will be more extreme than that of a Bayesian because individuals receive (on average) more supporting messages. Learning is affected even in the long run. As the number of messages grows to infinity, if messages are not too informative and q is high enough, there is a positive probability that the decision-maker may hold (and reinforce) the belief in the wrong state. To see why, suppose that the decision-maker has received a first message pointing towards the wrong state; if messages are sufficiently noisy and q is high, they are more likely to perceive messages that confirm this wrong state than messages that disagree with it, so they may become gradually convinced of the wrong state. Pouget et al. (2017) examines the consequences of this model on financial markets, connecting them to several empirical patterns.

Other models explain confirmatory bias as a desire for consistency. Yariv (2005) considers a model with binary states and three periods in which individuals must choose an action after observing one of two informative messages. However, their utility is affected not only by the value of the action but also by the consistency of beliefs: they receive a utility boost if their prior convictions become stronger, as they like to think that they have made the right choice in previous periods. Depending on the agent's sophistication—their awareness of their own biases—the model can generate under- and overconfidence, excess stickiness or volatility in choice, and even a preference to avoid information.

Confirmation bias is likely to be particularly relevant when some of the messages are ambiguous—instead of pointing to specific states, they are open to interpretation. Fryer et al. (2019) studies a model with binary state space in which individuals receive either a message supporting one of the states or an ambiguous message. Individuals interpret the latter using their current belief and treat it as if it were the message that is more likely

[13] This broad term is connected to belief polarization and other findings. While several papers pointed out how the initial evidence of polarization can be compatible with Bayes' rule, other evidence seems impossible to fit within the Bayesian framework. See Benjamin (2019, p. 156).

given their belief. When ambiguous messages are sufficiently frequent, there is a positive probability that two agents that observe the same messages but start with different priors converge to being sure of opposite states, and a positive probability that both converge to the same wrong conclusion. The intuition is similar to that of Rabin & Schrag (1999): If enough messages are distorted by the prior, this will overcome the informativeness of undistorted messages and the prior directs the final outcome.

Finally, a form of confirmation bias may operate not only during information processing but also during information *seeking*: Individuals may seek information that confirms their beliefs or that confirms their previous choice. This may happen because subjects have a tendency to seek information where they believe their optimal choice will be, but they recall only their previous choice, not their previous belief. Together, these tendencies lead agents to seek information that confirms their previous choice. Kaanders et al. (2022) and Sepulveda et al. (2020) present evidence as well as basic models.

3.1.3. Base-rate neglect. Motivated by experimental evidence that beliefs do not vary enough with prior information (Kahneman & Tversky 1973), other papers study individuals who underweight the prior—the base-rate—while correctly incorporating new signals.

Benjamin et al. (2019) introduce a model that corresponds to the Grether rule with $\alpha \in [0,1)$ and $\beta = 1$. Here, predictions become sensitive to how the state space is constructed. Consider two scenarios. In the first, $S = \{$“Car A is bigger”, “Car B is bigger”$\}$, each with probability .5. In the second, “Car A is bigger” is divided into two states, “Car A is bigger and Red” and “Car A is bigger and Blue,” each with probability .25. Consider messages that inform us of the size of the car and not the color. Under Bayes' rule, the belief that “Car B is bigger” is the same in the two scenarios. This is no longer the case in the model of Benjamin et al. (2019): “Car B is bigger” is less likely in the second scenario because the individual underweights the fact that this state has higher prior probability.

In a dynamic setting with a sequence of messages over time, more recent messages are given more weight: past messages enter the prior of the following round and are underweighted, their influence decreasing with each new round. Beliefs fluctuate more than in the Bayesian benchmark, and individuals may become relatively sure of wrong beliefs after a few rounds if they receive a stream of coherent evidence. Because they overweight new evidence, they may also fail to converge to full confidence in one state even with infinite messages or fail to learn the true state even if their prior assigns positive probability to it.

3.1.4. Representativeness. Kahneman & Tversky (1972) introduce the *representativeness* heuristic: the idea that humans report the likelihood of A given B not by computing the conditional probability of A given B, but by trying to assess the extent to which A is “representative” of B, capturing the salient features of B. In principle, this could be a bias on how probabilities are computed or on how people answer questions about probabilities. Several papers have shown how this can capture many violations of Bayes' rule, like the Law of Small Numbers (LSN)—the belief that small samples should reflect the actual probabilities (Kahneman & Tversky 1972)—sample-size and base-rate neglects (Kahneman & Tversky 1973), and the conjunction fallacy—the belief that the intersection (conjunction) of two events is more likely than one of them by itself (Kahneman & Tversky 1983).

While the intuitive notion of representativeness may explain several biases, it needs to be precisely defined for the model to have real predictive ability. Kahneman & Tversky (1973) discuss several characteristics that judgments of representativeness should have but leave an exact definition open (see also Kahneman & Frederick 2002). Griffin & Tversky (1992) suggest that judgments of probabilities after observing some evidence are a combination of the evidence's *strength*—how extreme the information is, for example, the fact that a sample of coin tosses contains only heads—and its *weight*—how significant the evidence is, e.g., the size the sample of coin tosses. In their models, individuals give too little importance to the weight of evidence and too much to the strength, which representativeness affects. Individuals should then be overconfident when strength is high but weight is low—they over-infer from unreliable evidence that is very representative—and underconfident in the opposite case. However, this model only partially contributes to defining what representativeness actually means in general. Other papers have suggested using likelihood or similarity, but both have several limitations (see Tenenbaum et al. 2001 for a discussion). Finally, Tenenbaum et al. (2001) define the representativeness of data d given a hypothesis h as the logarithm of the likelihood ratio $\frac{P(d|h)}{P(d|h^c)}$, where h^c is the negation of the hypothesis h. Intuitively, data is representative of a hypothesis the higher the ratio that this data is generated under that hypothesis than under the alternative. This captures several intuitive features of representativeness but relies on the specification of an alternative hypothesis, which may be seen as either a welcome degree of freedom or a limit on predictive power.

In economics, several papers have tried to formalize the precise meaning of representativeness. Probably the most comprehensive attempt is the model of local thinking and diagnostic expectations of Gennaioli & Shleifer (2010) and Bordalo et al. (2016), related to Tenenbaum et al. (2001), which we discuss in depth in Section 3.3. Zhao (2018) gives axiomatic foundations of a model where individuals use how "similar" A is to B instead of how likely A is given B. The paper suggests a specific form for the similarity function, of which a convenient special case is $\pi(A|B)^a\pi(B|A)^{1-a}$ for some $a \in (0, 1]$; individuals mistakenly also consider $\pi(B|A)$. The paper shows how this approach captures base-rate neglect and the conjunction and disjunction fallacies.

Rabin (2002) studies the Law of Small Numbers in a model in which agents face independent and identical draws of a sequence of one of two messages a or b, where they are trying to learn the probability θ of message a, given a finite set of possibilities. Instead of treating each draw as independent, these agents update as if these draws were made *without replacement* from an urn of size N containing θN a messages and $(1-\theta)N$ b messages. They act as if draws were correlated, generating the gambler's fallacy: The second draw is believed to be negatively correlated with the first.

The paper then explores two economic implications. First, the LSN implies overinference after a short sequence of messages: Because the LSN implies that agents believe that even short sequences must reflect the underlying rate, they will exaggerate how informative each small sample is. Moreover, when agents observe messages from different sources with different rates, they believe that rates are more heterogeneous than they really are—they ascribe differences in small samples to large differences in rates.

Rabin & Vayanos (2010) introduce a related model in a more applied setting, in which individuals receive i.i.d. messages but treat them as if negatively correlated over time, with more recent realizations having a stronger impact. The paper considers the general case in which the underlying state can be constant or follow an A-R process and shows how this model can generate not only the gambler's fallacy but also, when the state is constant, the hot-hand fallacy (the mistaken belief that streaks will continue).

Finally, Noor & Payró (2022) proposes an axiomatic model of the Law of Small Numbers canonical coin-tossing environment. The paper proposes the postulate of Mean Reversion, which formalizes the belief that the sample mean will tend to stay close to the bias of the coin along the entire sequence, and obtain a representation in which beliefs evolve as if the bias of the coin is path-dependent and self-correcting.

3.2. Biases on how conditional probabilities are computed

Another class of models studies deviations from Bayes' rule that originate from mistakes in how conditional probabilities are computed. (Models of representativeness can also be considered part of this class.) Benjamin et al. (2016) study the Non-Belief in the Law of Large Numbers (NBLLN), the erroneous belief that the proportion of binary signals may differ from the population mean even in large samples. Agents observe a clump of size N of binary signals a or b, generated by a binomial distribution that gives signal a with rate θ. The agent, however, treats them as if generated in two phases: First, a subjective rate β is drawn from a full-support distribution f_θ, centered at θ; second, the sample is generated from a binomial with rate β. Given this mistaken belief, the agent applies Bayes' rule. This has several implications. First, if $N > 1$, beliefs about the ratio of a signals are mean-preserving spreads of the correct ratio. Intuitively, if N grows to infinity, a Bayesian agent will be sure of seeing exactly θ% a signals, while NBLLN agents believe that the proportion of a signals comes with density f_θ, which is not degenerate. Moreover, when predicting θ, such agents under-infer from data: even with infinite N, they believe that each θ implies a full-support distribution instead of a precise percentage; they treat the data as noisier than it is and believe that noise does not disappear even with infinite data. Because NBLLN agents under-infer from data, they are less willing to pay for data and may resort to smaller datasets. NBLLN agents are also less willing to bet on many independent identical gambles because they fail to appreciate that they entail little risk for the law of large numbers.

3.3. Local thinking and diagnostic expectations

Motivated by representativeness, Gennaioli & Shleifer (2010) and Bordalo et al. (2016) introduce the *local thinking* model, which connects the idea of representativeness with features of memory retrieval, with an approach very related to Tenenbaum et al. (2001). To illustrate, consider a doctor who runs a test for a disease carried by 1% of the population. Patients are either sick (s) or healthy (h), and the test is either positive (+) or negative (-). The test has few false negatives but many false positives: Only 1% of sick patients test negative, while 30% of healthy patients test positive. In this case, the chances that a patient who tests positive carries the disease is low, at $\frac{1}{31}$ per Bayes' rule. Local thinking suggests that individuals *inflate* this conditional belief: Because positive tests are more common among the sick than the healthy, it becomes "representative" of the sick.

3.3.1. Local Thinking. Formally, consider payoff-relevant states S and a message $m \in M$. To define the model, we also need to specify a comparison group, defined by another possible message $\overline{m}$ (which may or not coincide with $M \backslash \{m\}$); we discuss possible choices below. According to local thinking, instead of using the correct $p(s|m)$, individuals use

$$\hat{p}(s|m) = \frac{1}{Z} p(s|m) \left(\frac{p(s|m)}{p(s|\overline{m})} \right)^\theta$$

where $\theta \geq 0$ is a parameter while Z is a normalization term that guarantees that $\hat{p}$ is a probability measure (summing to 1). Intuitively, $\frac{p(s|m)}{p(s|\overline{m})}$ denotes how representative state

s is of message m, and θ modulates how much this matters: Individuals overestimate the chances of s given m if s is more common given m than $\overline{m}$.

In our example of medical tests, recall that $p(s|+) = \frac{1}{31}$ but note that we also have $p(s|-) = \frac{1}{6931}$: Because there are few false negatives, the probability of being sick with a negative test is very low. Being sick is therefore more representative of those who tested positive than those who tested negative. Taking as a comparison group people with negative tests, the model of local thinking posits that, instead of $p(s|+) = \frac{1}{31}$, individuals use

$$\hat{p}(s|+) = \frac{1}{Z}p(s|+)\left(\frac{p(s|+)}{p(s|-)}\right)^{\theta} = \frac{1}{Z}\frac{1}{31}(\frac{\frac{1}{31}}{\frac{1}{6931}})^{\theta} = \frac{1}{Z}\frac{1}{31}\left(\frac{6931}{31}\right)^{\theta},$$

inflating the probability of being sick given a positive test. Note how this grows arbitrarily large with fewer false negatives: If $p(s|-)$ goes to zero, $\hat{p}(s|+)$ goes to one independently of the real $p(s|+)$ and for any θ.

This example shows how local thinking can generate base-rate neglect: The individual is not giving enough importance to the prior information that very few patients are sick. A similar logic shows how the model generates representativeness, conjunction, or disjunction fallacies. (These are intuitive; we refer to the original papers for a discussion.)

As discussed in Gennaioli & Shleifer (2010) and Bordalo et al. (2021a), local thinking can be motivated by known features of memory retrieval. Individuals associate trait s with message m the more frequently s and m are seen together (memory recall is associative), but this is subject to interference of two kinds: It is weaker the more m is observed not associated with s, and the more s is associated with $\overline{m}$. In our example, the association of being sick with a positive test is weaker the more sick people have a negative test and the more healthy people have a positive test. Because the latter is exceedingly rare, the interference is very weak, and the association between sick and positive is easy to recall.

An important aspect of this model is that individuals distort beliefs by inflating true characteristics—in our example, testing positive does make it more likely that the patient is sick, but this aspect is inflated. Distorted beliefs are therefore built on a "kernel of truth." Bordalo et al. (2016) show how this approach delivers a model of stereotypes: If individuals use local thinking to construct opinions, they will develop stereotypes, that is, inaccurate beliefs about the characteristics of different groups. Individuals take existing differences but exaggerate their importance depending on how representative they are of a given group.

3.3.2. Diagnostic Expectations. Bordalo et al. (2018, 2019) extend this idea to a dynamic framework, leading to the model of diagnostic expectations. (See also Gennaioli & Shleifer 2018, Bordalo et al. 2020, 2021b, 2022, forthcoming.) In one version of the model, in each period of an infinite horizon in discrete time, an individual is trying to predict a state s_t using the information available at that period, denoted I_t. As in local thinking, individuals overestimate the chances of states that are representative of a given information I_t. But what is the comparison group? Diagnostic expectations propose to use the information from the previous period, I_{t-1}. Denoting $f(s|I_t)$ as the true conditional probability, under diagnostic expectations individuals use the distribution

$$\hat{f}(s|I_t) = \frac{1}{Z}f(s|I_t)\left(\frac{f(s|I_t)}{f(s|I_t = I_{t-1})}\right)^{\theta}$$

where Z is the normalizing term guaranteeing this is a density. To illustrate, suppose s indicates the value of a stock and I_t contains good news. Then, the higher θ, the more individuals *overreact*: States corresponding to a high price are more representative of the new information than the old, and their likelihood is inflated. Indeed, this can give foundations to extrapolative expectations since good news makes even better news seem more likely.[14] For example, suppose that s_t follows an AR(1) process with Normal shocks, $s_{t+1} = bs_t + \epsilon_t$; in this case, the only information is the price itself. Suppose the price increases, $s_t > s_{t-1}$, and consider some $\hat{s}_{t+1} > s_t$. Because $\hat{s}_{t+1}$ is more likely under s_t than it was under s_{t-1}, it is representative of s_t and its chances are inflated. In fact, in this setup, it is easy to show that if $\mathbb{E}_t$ denotes the correct expectations at time t and $\hat{\mathbb{E}}_t$ the expectations under diagnostic expectations, we must have (Bordalo et al. 2018, Prop. 1):

$$\hat{\mathbb{E}}_t[s_{t+1}] = \mathbb{E}_t[s_{t+1}] + \theta\left(\mathbb{E}_t[s_{t+1}] - \mathbb{E}_{t-1}[s_{t+1}]\right).$$

Individuals overreact to news, extrapolating current trends more than a Bayesian would do. Yet, they maintain a kernel of truth: They overreact to news that is genuinely there. Note how this mechanism can also generate reversals: When good news stops coming, beliefs cool down even without negative news, leading to belief change not driven by fundamentals. (See Bianchi et al. forthcoming for more implications.)

3.3.3. Advantages and disadvantages. This general approach has several advantages. Most importantly, it provides a broad organizing principle derived from known properties of memory retrieval that applies in very different contexts, from lab experiments to expectations in finance, capturing a wide range of empirical patterns. It also adopts a tractable functional form with few parameters, suitable for several applications.

At the same time, it has a few limitations, some of which have already been discussed (Benjamin 2019, p. 153). The first one pertains to the choice of the comparison group, which is sometimes not obvious. In our example about medical tests, should the comparison group be those who tested negative (as we assumed) or the untested (as in one example in Gennaioli & Shleifer 2018)? The two approaches give very different predictions.[15] On the one hand, this degree of freedom allows the model's implications to depend on context, and Bordalo et al. (2016) argue that this flexibility is important to fit experimental data. On the other hand, the predictive power of the model is unavoidably reduced by the need for an additional parameter, the choice of which is sometimes difficult. Bordalo et al. (2021a) relates this to a notion of similarity, which may be manipulated exogenously.

Furthermore, this approach explains only some evidence of non-Bayesian behavior. Of course, this is not problematic, for we should not expect a model to explain all evidence. However, consider base-rate neglect. Local thinking can explain some real-world instances but not many lab experiments, where memory is less likely to play a role. Therefore, we need a separate model for base-rate neglect in the lab. But isn't it more natural to assume that the mechanism giving rise to base-rate neglect in the lab is the same as that operating outside? If so, local thinking may not be the full explanation for base-rate neglect.

Finally, while this approach can be motivated by features of memory and recall, this remains a model of a *bias*—it models non-Bayesian behavior as a mistake. This aspect sets this approach apart from the attempts to capture non-Bayesian behavior as optimal

[14] See Barberis (2018, p. 104) for more discussion on representativeness and base-rate neglect as foundations of "extrapolative beliefs."

[15] For example, when the rate of false negatives goes to zero, we get unbounded distortion *for any* θ when the comparison group is those who test negative, but a more limited one if the comparison group is the untested.

reactions to cognitive limitations, discussed in Section 5. (Of course, future research may show how this bias is an optimal reaction to cognitive bounds; to our knowledge, however, this has not been done.)

3.4. Biases due to temptation

Another type of deviation from Bayes' rule takes the form of temptation and is connected to the literature on temptation and self-control that originated with Gul & Pesendorfer (2001). For example, individuals who receive a particularly good signal may overreact and become overly optimistic, or they may underreact and fail to appreciate the good news. In either case, they act as if they updated a prior different from the one they used before information.

Epstein (2006) introduces this idea and studies a three-period model in which, at time 0, agents have a prior over the states and a belief about the connection between the signals in period 1 and the state. However, agents also know that they will be tempted to deviate from Bayes' rule, e.g., being too exuberant or too cautious. As in Gul & Pesendorfer (2001), this conflict is captured by studying period 0 preferences over menus faced in period 1. Epstein (2006) provides a very neat representation theorem and shows how the model accommodates under- and overreaction, base-rate neglect, sample bias, as well as representativeness. Epstein et al. (2008) extends the model to infinite horizon settings, while Epstein et al. (2010) studies additional long-run implications.

4. Questioning the Prior

Typical experiments on updating provide subjects with the correct prior belief and all conditional likelihoods, and deviations from Bayes' rule can be reasonably classified as mistakes. However, this normative appeal is less clear in natural environments where beliefs are subjective. Individuals who need to make choices that depend say, on the return of a stock or on GDP growth, do not have access to an objective prior belief and will have to form a *subjective* prior. But updating subjective priors may be different. For example, suppose that our individual receives some unexpected news, information that was unlikely given their adopted prior, such as an unusually high earning announcement for a stock. In this case, the individual may go beyond Bayes' rule: They may reconsider whether they were using the right prior in the first place. Because the prior was subjectively chosen, they may well wonder if they had made the right choice. Clearly, this is a different type of non-Bayesian behavior than the ones discussed in the previous section.

It is worth emphasizing that this approach is not unusual in the development of science, including in the social sciences. Scholars often use models to understand processes and typically look at existing data to estimate model parameters and make predictions; they may entertain multiple models simultaneously, but rarely consider the universe of *all* possible models; for example, to our knowledge, all central banks use a finite (and not extensive) subset of macroeconomic models to predict inflation (which is a model itself). Every so often, new data emerges that seems incompatible with the adopted model. In most cases, such data was not ruled out by the model (many have full support), but it is sufficiently unlikely that the model itself is questioned and sometimes replaced. This is a central process in the development of science, including social sciences and economics, as highlighted in classical work in epistemology (Kuhn 1962). Yet, it is clearly non-Bayesian.

4.1. The Hypothesis Testing model

The intuition above is formalized in the Hypothesis Testing (HT) model of Ortoleva (2012).[16] Individuals have a prior $\pi \in \Delta(\Omega)$, but also a second-order belief (a prior over priors) $\rho \in \Delta(\Delta(\Omega))$, as well as a threshold $\epsilon \in [0, 1)$.[17] The second-order belief ρ indicates which priors are considered and their relative likelihood, and the initial prior π is the one with the highest probability, $\{\pi\} = \underset{\pi' \in \Delta(\Omega)}{\arg\max}\ \rho(\pi')$.

When new information A arrives, individuals first test to determine if they were using the right prior, using the threshold ϵ. If A is not unlikely given π, that is, $\pi(A) > \epsilon$, then π is kept and updated following Bayes' rule: When business proceeds as usual, the agent is Bayesian. However, when the new information was unexpected given the prior, $\pi(A) \leq \epsilon$, then the prior is questioned: Our agent goes back to the prior over priors ρ, updates it with Bayes' rule, and chooses the prior to which the updated ρ gives the highest likelihood (which may also coincide with the previous one). That is, when a "model" performs poorly, the individual questions it and may replace it with the model that appears most likely given the new data (and the prior over models).

Formally, recall that π_A^{BU} is the Bayesian' update of π following A; denote by ρ_A^{BU} the Bayesian' update of ρ following A, that is, $\rho_A^{\text{BU}}(\bar{\pi}) = \frac{\bar{\pi}(A)\rho(\bar{\pi}))}{\int_{\Delta(\Omega)} \pi'(A)\rho(\mathrm{d}\pi')}$ for all $\bar{\pi} \in \Delta(\Omega)$. We say that the updated beliefs $\{\pi_A\}_{A \in \Sigma}$ admit an HT representation (π, ρ, ϵ) if for all $A \in \Sigma$,

$$\pi_A = \begin{cases} \pi_A^{\text{BU}} & \text{if } \ \pi(A) > \epsilon \\ \\ \bar{\pi}_A^{\text{BU}} & \text{otherwise} \end{cases}$$

where $\{\bar{\pi}\} = \underset{\pi' \in \Delta(\Omega)}{\arg\max}\ \rho_A^{\text{BU}}(\pi')$.[18]

Behavior in this model differs from that of a Bayesian who starts from a prior over priors: In that case, the behavior is Bayesian with a belief that is the expectation of the prior over priors, accounting for *all* beliefs in the support of the prior over prior. Instead, the HT model uses only the belief to which the prior over priors assigns the highest likelihood. This may appear irrational. Indeed, this could be the behavior followed by a boundedly rational individual with a "cost of considering models." Investors trying to forecast GDP growth may not account for *all* possible models in economics and finance, as doing so may be exceedingly costly. Instead, they may focus on a combination of those more likely to be correct, which is a model itself; indeed, even central banks focus on a selection of models. This model would then be used and updated until new data emerges that induces our decision-maker to question it and possibly pick a new one.

We conclude by noting how the HT model does not require the new information to be in the support of the original prior ($\pi(A) > 0$). Indeed, the HT updating rule is well-defined after zero probability events, as we will discuss in Section 7.

[16] A related idea appears in the game theory: In Foster & Young (2003), players test their priors about the opponent's behavior and revise them if rejected. However, no prescription is given for how the new prior is chosen. See also Weinstein (2011).

[17] Ortoleva (2012) studies preferences as primitives, but for simplicity, we rewrite everything in terms of beliefs. Because the original paper assumes subjective expected utility, very little is lost; the mapping is routine.

[18] Note how $\bar{\pi}$ is the unique maximizer. Ortoleva (2012) shows how ρ can always be constructed to allow a unique maximizer. Alternatively, one may include a decision rule that indicates which prior should be used in the case of indeterminacy.

4.2. Foundations

We now discuss the foundations of the HT model.

4.2.1. Dynamic Coherence.

Consider two events A and B and suppose that, if told that the state lies in A, the individual is sure it is also in B ($\pi_A(B) = 1$), and also suppose that, if told that the state is in B, they are sure it is also in A ($\pi_B(A) = 1$). This is a special situation in which, after either event, the individual is also sure of the other event. Then, one could argue that the informational content of A and B is the same—in both cases, it is as if the individual has been told $A \cap B$. But then, we should have $\pi_A = \pi_B$. When A and B are given positive probability by the prior, this is implied by Bayes' rule (this happens when the prior assigns probability 0 to $A \backslash B$ and $B \backslash A$). But this is a much weaker requirement than Bayes' rule, as it applies only to special situations. This is the core requirement of Dynamic Coherence, which extends this idea to a sequence of events: If after A_i we are sure of A_{i+1}, and after A_n we are sure of A_1, we have a cycle in beliefs: then, the informational content of all events is the same, and beliefs should be the same.

Axiom 3 (Dynamic Coherence)**.** *For any* $A_1, \ldots, A_n \in \Sigma$, *if* $\pi_{A_i}(A_{i+1}) = 1$ *for* $i = 1, \ldots, (n-1)$ *and* $\pi_{A_n}(A_1) = 1$, *then* $\pi_{A_1} = \pi_{A_n}$.

Bayesian agents naturally satisfy Dynamic Coherence whenever the events $A_1, \ldots, A_n$ are given positive probability by the prior. But of course, Bayes' rule is much stronger. In fact, there is a conceptual difference between this axiom and Dynamic Consistency, the core foundation of Bayes' rule. The latter compares beliefs before and after information, the prior and the posteriors; as such, it cannot restrict reaction to zero probability events. Instead, Dynamic Coherence imposes consistency of posteriors after different events *without going through the prior*, and also applies to events to which the prior gives zero probability. At the same time, for events that have non-zero probability, Dynamic Coherence is much weaker than Dynamic Consistency. The two axioms are, therefore, not nested and can be imposed together, as we will see momentarily.

We are now ready to state the representation theorem. To this end, it is useful to define a *minimal* HT model—an HT model such that no strictly smaller ϵ represents the same behavior. The following theorem is proved in Ortoleva (2012).

Theorem 2. π *and* $\{\pi_A\}_{A\in\Sigma}$ *satisfy Consequentialism and Dynamic Coherence if and only if they admit a minimal HT model representation* (π, ρ, ϵ). *Moreover,* $\epsilon = 0$ *if and only if* $\{\pi_A\}_{A\in\Sigma}$ *also satisfies Dynamic Consistency.*

The HT model is obtained by simply replacing Dynamic Consistency with Dynamic Coherence. Because the two postulates are not nested, adding Dynamic Consistency gives us the special case of the HT model when $\epsilon = 0$, where behavior coincides with Bayes' rule whenever defined, but where the reaction to zero probability events is also prescribed.

4.2.2. Generality.

Dominiak et al. (2023) show that the HT model is equivalent to Inertial Updating (see Section 6.2), which they show generalizes several well-known updating rules, including the Grether rule. This implies that several models discussed in Section 3—all those admitting representations using the Grether rule—also admit an HT representation. Indeed, Dynamic Coherence and Consequentialism put only limited restrictions on behavior. To see this in the functional form, note that if ϵ is chosen to be close to 1, updating is entirely controlled by ρ, a high-dimensional object with many degrees of freedom. A drawback of the HT model is, therefore, its generality—the other side of the coin of its foundation on weak axioms. On the other hand, the model predictions coincide with Bayes' rule for all events with a likelihood above ϵ, which means that the generality can be modulated by the choice of ϵ. In general, the HT model can be understood as a useful framework to study non-Bayesian behavior, but in applications, more structure should be given to ρ and ϵ to increase predictive power.

5. Deviations due to Bounds to Rationality, Perception, Cognition, or Memory

A third class of models of non-Bayesian behavior ascribes it to limitations in individuals' ability to acquire or process information. This approach differs conceptually from the models that assume a bias. Here, individuals react *optimally* and are, therefore, much closer to rationality, except they have additional constraints of a cognitive or perceptual nature. Moreover, biases can be corrected—for example, subjects can be instructed to be mindful of base-rate neglect—and a benevolent social planner may want to do so. Instead, deviations due to the optimal reaction to limitations seem much harder to correct.

The models in this section belong to the classical literature in bounded rationality and to the more recent literature that attempts to explain several biases as optimal reactions to constraints, deeply related to the models of the Bayesian paradigm in cognitive sciences discussed in Section 2.2.3; see Gabaix (2019) and Woodford (2020) for recent reviews.

5.1. Bayesian cognitive noise

Recent literature suggests that deviations from Bayes' rule may be due to the optimal reaction to limitations in perception or processing. To illustrate, consider a problem unrelated to updating: An individual needs to guess the weight of a package. They can hold it in their hand for a few seconds to obtain some information. Assume that our individual has a prior about the weight and that the information is the true weight plus a symmetric noise of known variance. Then, the posterior will be a combination of the prior and of the signal: The more informative the signal, the closer the posterior to the signal and the further from the prior. This is the standard Bayesian way of modeling beliefs with imprecise information, where subjects know the exact noise.

Several recent papers suggest that this approach can explain many deviations from normative benchmarks for risk and time preferences (Bhui & Gershman 2018, Woodford 2020, Gabaix & Laibson 2022, Enke & Graeber 2023). Interestingly, this approach can also explain deviations from Bayes' rule. Suppose that individuals have an imprecise perception of the outcome of Bayesian updating: They may receive a noisy signal of the correct posterior and be aware of such noise. Assuming for simplicity a binary state space S, suppose that the prior belief about the correct posterior has mean d and precision τ, while the signal has an error with mean zero and precision η. Then, the posterior is a convex combination of the signal and d, giving more weight to the latter the higher η is and the smaller τ is.[19] This has two implications. First, choices will be stochastic because of the noise of the signals. Second, because the weight on the signal is less than one and the rest of the weight is given to a value independent of the problem, this mechanically reduces sensitivity to both the

[19] Enke & Graeber (2023) present in Appendix A a formalization with Beta and Binomial distributions, where closed-form solutions are attained.

base rate and the likelihood ratio. That is, the optimal behavior of a Bayesian individual who has a noisy perception of the optimal action is non-Bayesian.

Alternatively, the imprecision in cognition may lie in the strength of the signal. Augenblick et al. (2023) propose a model in which signal informativeness is perceived with noise: Individuals know the direction they should update in but are unsure of how much. Like above, individuals have an initial belief about informativeness with mean μ and receive noisy information about it. (This could be noise in the information or uncertainty about how to incorporate it.) This generates stochasticity in choice and a bias toward the prior in the dimension in which there is noise: In this case, individuals will bias their belief about the informativeness of signals towards the average μ. Mechanically, this implies that individuals overestimate informativeness when informativeness is below the expectation μ and underestimate it otherwise. In turn, this means that individuals overreact after weakly informative signals and underreact after strongly informative ones.

Other papers present related models. Ba et al. (2023) combine the effects of cognitive noise with higher attention to more representative states (thus relating to the approach of Bordalo et al. 2016). Azeredo da Silveira & Woodford (2019) and Azeredo da Silveira et al. (2020) present models in which cognitive noise applies to the memory of past states, showing how this generates, among other things, overreaction in forecasts.

5.1.1. What is non-Bayesian?.

A crucial aspect of these approaches is that individuals are rational and follow Bayesian prescriptions given the noise they face. Indeed, they offer a fully Bayesian explanation for non-Bayesian behavior. Some may even argue that the documented behaviors were not really deviations from Bayes' rule because we were not accounting for the genuine noise in the information. (Of course, this is mostly a matter of definitions.) Another advantage is that they explain non-Bayesian behavior using a modeling framework that, as mentioned above, can also explain several other deviations from normative benchmarks in risk and time preferences, offering the possibility of a unified explanation. This unified explanation, however, needs to be subject to rigorous empirical tests (e.g., like those in Chapman et al. 2023; see also Enke & Graeber 2023). Moreover, empirical tests should be mindful not only of matching average behavior but of accounting for heterogeneity—because behavior is often multimodal, and it is essential that models capture the possible modes instead of a generic average that no individual subject chooses.

On the other hand, it is important to highlight the degrees of freedom inherent in these modeling approaches. At an abstract level and focusing on standard choice without external manipulations, their sole general implication is that choices must be biased towards the prior in the dimension in which we assume that there is noise. By carefully selecting the prior or the dimension of noise, one can give a "Bayesian rationale" to a very wide range of behaviors. At the same time, more tests are possible (e.g., manipulating the prior or the noise), and some have been adopted in the literature. To our knowledge, a careful study of the broad empirical implications of this approach has not been attempted.

5.2. Other forms of cognitive limitations

Many other papers have suggested bounds to rationality, memory, or perception that generate violations of Bayes' rule. We focus on a few well-known or recent examples.

In the models discussed in Section 5.1, agents optimally react to noisy information, but the information is exogenous. Other models also include the choice of attention. Gabaix

(2014) introduces the sparse-max model, where the attention agents pay to a variable depends on how useful they believe it is; as discussed in Gabaix (2019, 2.3), this may generate several departures from Bayes' rule. In Schwartzstein (2014), individuals predict a variable using observables, but attend to a covariate only if they (currently) think it is sufficiently likely to be predictive. They follow Bayes' rule but are naïve in the sense of not attempting to infer what the unobserved variables may have been. In Khaw et al. (2017), individuals need to estimate the likelihood of a varying state that governs a stream of observations, and the decision of whether and how much to adjust the estimates is the solution of a rational inattention problem (Sims 2003). This leads to infrequent non-Bayesian adjustments.

Wilson (2014) studies the optimal behavior of agents with finite memory states who face a binary decision and a stochastic sequence of signals; the model can generate several deviations from Bayes' rule, including confirmation bias and polarization. (See also Hellman & Cover 1970.) Mullainathan (2002) presents a model of bounded memory recording and retrieval that may also generate over- and under-reactions.

The models in Fryer & Jackson (2008) and Jakobsen (2021) are based on "coarse updating": Decision-makers simplify the problem by considering only a subset of possible probabilities or categories (Mullainathan et al. 2008 apply a similar approach to persuasion). Jakobsen (2021), for example, gives the axiomatic foundations of a model in which individuals partition the probability simplex in (convex) cells, each of which has a representative distribution; after information, they adopt the belief that corresponds to the representative distribution in the cell where the Bayesian posterior belongs.

Finally, several models in finance show how forms of bounded rationality can lead to extrapolative beliefs. In the model of Hong & Stein (1999), while some investors receive information about fundamentals, others observe only price changes; the latter group will, therefore, follow price changes to infer fundamentals but will continue to do so even when all information is incorporated into the price, leading the price to "overshoot." See Barberis (2018) for a review of models of extrapolative beliefs in finance.

Finally, a growing literature studies learning with misspecified beliefs, which may generate behavior that appears non-Bayesian. Several of these papers connect misspecification to bounded rationality, with decision-makers using overly simple models. See, among many, Barberis et al. (1998), Daniel et al. (1998), Hong et al. (2007), Fuster et al. (2010), Ortoleva & Snowberg (2015), Esponda & Pouzo (2016), Spiegler (2016), Glaeser & Nathanson (2017), Heidhues et al. (2018), Eliaz & Spiegler (2020), Gagnon-Bartsch et al. (2021), Gagnon-Bartsch & Bushong (2022), Montiel Olea et al. (2022), Ba (2023), Frick et al. (2023). As mentioned in the introduction, however, we will not discuss these papers here.

6. Other Models of Non-Bayesian Behavior

6.1. Utility from beliefs

A very different reason why people may be non-Bayesian is because they *want* to believe that a specific state is true and update more after "good news" than "bad news." (See Benjamin 2019, Section 9 for a discussion of the complex evidence.) Like confirmation bias, this asymmetry can be easily captured with a modified Grether rule, where updating strength depends on the direction of the evidence (Benjamin 2019, Section 9.1). However, several other approaches have been proposed.

In Bénabou & Tirole (2002), individuals can choose to ignore (or fail to remember) signals that point to states of the world they like less, in a "game of self-deception." (See

Gottlieb 2010 for a dynamic model.) In Bénabou (2013), a group of individuals participates in a joint enterprise that gives a payoff that depends on their actions and on a joint productivity shock, about which they receive a public binary signal. Agents derive anticipatory utility from future payoffs and can choose to incorporate the signal or manipulate it to encode it as if it were the opposite (at a cost). The main insight is that information manipulation is a strategic substitute when agents benefit from others' over-optimism, and a strategic complement when agents are made worse off by others' mistakes. In the latter case, we can have multiple equilibria and a "contagious" denial.

Möbius et al. (2022) propose a model in which individuals also experience a direct utility benefit from believing that one state has higher chances. Following Brunnermeier & Parker (2005), agents can commit to a biased updating rule to maximize total utility, trading off the desire to believe in a given state with the cost of inaccurate choices. The paper shows that the optimal rule is asymmetric—it responds differently to good versus bad news—as well as conservative—agents underreact to all signals. In Caplin & Leahy (2019), individuals can distort the mapping between messages and states to increase their belief in a given state but have to pay a cost that depends on the extent of distortion.

6.2. Updating as minimizing a distance

A different way of thinking about updating is as distance minimization; in economics, this appears in Perea (2009) and Dominiak et al. (2023). Following the latter, for finite Ω, define a distance function with respect to the prior as any function that assigns the smallest value to the prior itself: $d : \Delta(\Omega) \to \mathbb{R}$ is a distance function with respect to $\pi \in \Delta(\Omega)$, denoted d_π, if $d_\pi(\pi) < d_\pi(\pi')$ for all $\pi' \in \Delta(\Omega)\backslash\{\pi\}$. (Note that d_π may violate symmetry or the triangle inequality and may therefore not be a distance; we follow the paper in maintaining this terminology.) An individual follows Inertial Updating if there is a distance function d_π such that for all $A \in \Sigma$, $\pi_A = \underset{\pi' \in \Delta(A)}{\arg\min}\, d_\pi(\pi')$. Simply, the updated belief after A is the belief that is closest to the prior among those that have support only in A. Considering different ds, this approach can capture several updating rules. Bayes' rule is obtained if $d_\pi(\pi') = -\sum_{\omega\in\Omega} \pi(\omega)\sigma\left(\frac{\pi'(\omega)}{\pi(\omega)}\right)$ where σ is strictly increasing and strictly concave. (If $\sigma(x) = \ln(X)$, this gives the Kullback-Leibler divergence, connecting this approach to very well-known results in information theory, e.g., Berk 1966; see also Zhao 2022, 2.3.) The Euclidean distance gives $\pi_A(s) = \pi(s) + \frac{1-\pi(A)}{|A|}$: deviations from Bayes' rule in which prior probabilities are reallocated uniformly. The paper also shows that an appropriately chosen distance delivers the Grether rule and interesting generalizations.

Surprisingly, Dominiak et al. (2023) show that Inertial Updating is behaviorally equivalent to the HT model discussed in Section 4: Preferences admit an Inertial Updating representation if and only if they admit an HT representation if and only if they satisfy Consequentialism and Dynamic Coherence. This provides behavioral foundations and illustrates how distance minimization may be a convenient form to study non-Bayesian behavior.

7. Reaction to zero probability events

A well-known limitation of Bayes' rule is that it does not prescribe updating after information to which the prior assigns probability zero.This limitation has relevant consequences, especially in game theory, since in most notions of equilibria, beliefs assign probability zero to off-equilibrium behavior. Bayes' rule, therefore, imposes no discipline on off-equilibrium beliefs, a well-known concern of Bayesian Nash Equilibrium that led to the development of several refinements. We now briefly review the main extensions of Bayes' rule to reactions to zero probability events in economics.

The Conditional Probability System (CPS) of Myerson (1986a,b) defines $\{\pi_A\}_{A\in\Sigma}$ such that $\pi_A(\omega) = \pi_B(\omega)\pi_A(B)$ for all $\omega \in B \subseteq A$. While Bayes' rule implies this condition if $\pi_A(B) \neq 0$, all conditional probabilities are defined in CPS. Another approach is the Lexicographic Conditional Probability Systems (LCPS) of Blume et al. (1991), according to which individuals adopt a vector of beliefs that they use lexicographically; this updating rule is obtained by relaxing (Archimedean) continuity.

Two of the models discussed earlier in the paper also include updating rules after probability zero events. The HT model (Section 4) specifies the posteriors after any event. When $\epsilon = 0$, the HT model coincides with Bayes' rule whenever it is defined but extends it after probability zero events, when individuals will necessarily have to reconsider their prior; the model proposes a way to select a new prior based on updating the prior over priors. Dominiak & Lee (2023) use the HT model to define a refinement of Perfect Bayesian Nash Equilibria, showing how it relates to, but does not coincide with, well-known refinements like the Intuitive Criterion.

Because the HT model is behavioral equivalent to Intertial Updating (Section 6.2), also the latter specifies behavior after probability zero events. Perhaps surprisingly, Dominiak et al. (2023) show that the CPS model is, in fact, a special case of both the HT model and Intertial Updating. (LCPS are instead not continuous and therefore not nested.)

8. Discussion and Conclusions

We have discussed several alternatives to Bayes' rule. We began with models of non-Bayesian behavior due to a bias— a pull towards suboptimal behavior related to a heuristic or a mistake. Next, we analyzed non-Bayesian behavior due to individuals questioning their prior: When the prior is subjectively chosen and individuals are presented with evidence that appears unlikely given the prior, they may wonder if they were using the right prior to begin with. We then considered models of non-Bayesian behavior due to bounds to cognition, memory, or perception, and models of non-Bayesian behavior due to motivated beliefs and updating as minimizing a distance. Finally, we briefly reviewed models that extend Bayes' rule to capture reactions to information to which the prior assigned probability zero.

Overall, the literature has proposed very different kinds of models. Some models provide simple and tractable functional forms with few parameters; this is the case with several models meant to capture specific biases, such as the Law of Small Numbers or the NBLLN. Other models are more complex or have more complex parameters, often because they are derived from postulates that aim to be general; this is the case with the HT model, for which, in applications, one needs to assume a simple functional form for the prior over priors. While most models we discussed focus on updating, some relate non-Bayesian behavior to other phenomena, offering the possibility of a unified, broader explanation; this is the case with the models of cognitive noise. Several models are tailored to capture only very specific deviations, like the Law of Small Numbers, while others attempt to be much more general, trying to derive several empirical regularities from a unique cause. This is the case with local thinking, which derives a variety of regularities in lab experiments and in forecasting from biases in memory, or the models of cognitive noise, based on a framework that can provide a "Bayesian" explanation for several regularities. These attempts have the advantage of generality, although typically at the cost of more flexibility. Rigorous empirical tests will need to test these approaches and offer ways to limit their flexibility.

From this brief overview, it is clear that the literature has proposed *many* different models, each with advantages and disadvantages. In general, research on models of non-Bayesian behavior does not seem to have reached the level of maturity of those areas of behavioral economics where a "leading" model has emerged—for better or worse, depending on one's opinion of the leading model. This may be inevitable: One may argue that non-Bayesian behavior is, in fact, a broad umbrella that refers to several different phenomena that are only tangentially related, and one should not hope, or even try, to obtain a unified model. Yet, the literature has not converged even for specific important phenomena: There is no "leading" model of base-rate neglect or of overreaction to unexpected news. This is problematic, for it hinders the development of applications: What model should we use to study the implications of base-rate neglect on, say, matching or finance? Indeed, very few papers study the implications of non-Bayesian behavior on several economic choices, especially compared to other areas of behavioral economics like reference dependence or non-exponential discounting, for which extensive literature has analyzed the implications in a wide range of settings. This may be because we have yet to agree on the most important violations of Bayes' rule, or on which are the correct models, or it may be because economists find it difficult to accept non-Bayesian behavior despite overwhelming evidence.

Surely, more empirical work is needed. While several deviations have been robustly documented, only modest progress has been made in documenting which are most consequential for economics. Indeed, outside of finance and related forecasting, comparably little empirical work has studied belief updating in real-world behavior, where individuals are not provided exogenous priors and may need to seek and integrate information; surprisingly little is known about how behavior differs from bookbag-and-poker-chips lab experiments, and which are the prominent biases. Moreover, comparatively little empirical work has studied the heterogeneity of behavior: several papers analyze the empirical fit for average behavior, but this may be misleading if we have heterogeneous, multimodal responses; instead of fitting average behavior chosen by no individual subject, we should aim to understand the right model, or combination of models, that capture the most important modes.

More theoretical work is, however, also necessary. More work is needed to develop models that provide the right balance of tractability, empirical fit, and a convincing story, possibly connecting non-Bayesian behavior to other biases. And much more work is needed to study the implications of non-Bayesian behavior in various economic environments. This is great news for researchers, for an important area has yet to be understood in full.

DISCLOSURE STATEMENT

The author is not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

We thank Ned Augenblick, Dan Benjamin, Mark Dean, Benedetto De Martino, David Dillenberger, Adam Dominiak, Nicola Gennaioli, Daniel Gottlieb, Matthew Kovach, Eben Lazarus, Collin Raymond, Andrei Shleifer, Michael Thaler, Gerelt Tserenjigmid, Michael Woodford, and Chen Zhao for their useful comments and suggestions.

LITERATURE CITED

Alchourrón CE, Gärdenfors P, Makinson D. 1985. On the logic of theory change: Partial meet contraction and revision functions. *Journal of Symbolic Logic* 50(2):510–530

Augenblick N, Lazarus E, Thaler M. 2023. Overinference from weak signals and underinference from strong signals. Mimeo, University of California Berkeley

Azeredo da Silveira R, Sung Y, Woodford M. 2020. Optimally imprecise memory and biased forecasts. National Bureau of Economic Research

Azeredo da Silveira R, Woodford M. 2019. Noisy memory and over-reaction to news. *AEA Papers and Proceedings* 109:557–561

Ba C. 2023. Robust misspecified models and paradigm shift. Mimeo, University of Pennsylvania

Ba C, Bohren JA, Imas A. 2023. Over-and underreaction to information. Mimeo, University of Pennslyvania

Barberis N. 2018. Psychology-based models of asset prices and trading volume. In *Handbook of behavioral economics: applications and foundations 1*, vol. 1. Elsevier, 79–175

Barberis N, Shleifer A, Vishny R. 1998. A model of investor sentiment. *Journal of financial economics* 49(3):307–343

Bénabou R. 2013. Groupthink: Collective delusions in organizations and markets. *Review of economic studies* 80(2):429–462

Bénabou R, Tirole J. 2002. Self-confidence and personal motivation. *Quarterly Journal of Economics* 117(3):871–915

Benjamin D, Bodoh-Creed A, Rabin M. 2019. Base-rate neglect: Foundations and implications. Mimeo, University of California Berkeley

Benjamin DJ. 2019. Errors in probabilistic reasoning and judgment biases. *Handbook of Behavioral Economics: Applications and Foundations 1* 2:69–186

Benjamin DJ, Rabin M, Raymond C. 2016. A model of nonbelief in the law of large numbers. *Journal of the European Economic Association* 14(2):515–544

Berk RH. 1966. Limiting behavior of posterior distributions when the model is incorrect. *The Annals of Mathematical Statistics* 37(1):51–58

Bhui R, Gershman SJ. 2018. Decision by sampling implements efficient coding of psychoeconomic functions. *Psychological Review* 125(6):985

Bianchi F, Ilut CL, Saijo H. forthcoming. Diagnostic business cycles. *Review of Economic Studies*

Blume L, Brandenburger A, Dekel E. 1991. Lexicographic probabilities and choice under uncertainty. *Econometrica* :61–79

Bordalo P, Coffman K, Gennaioli N, Schwerter F, Shleifer A. 2021a. Memory and representativeness. *Psychological Review* 128(1):71

Bordalo P, Coffman K, Gennaioli N, Shleifer A. 2016. Stereotypes. *Quarterly Journal of Economics* 131(4):1753–1794

Bordalo P, Gennaioli N, Kwon SY, Shleifer A. 2021b. Diagnostic bubbles. *Journal of Financial Economics* 141(3):1060–1077

Bordalo P, Gennaioli N, La Porta R, Shleifer A. forthcoming. Belief overreaction and stock market puzzles. *Journal of Political Economy*

Bordalo P, Gennaioli N, Ma Y, Shleifer A. 2020. Overreaction in macroeconomic expectations. *American Economic Review* 110(9):2748–2782

Bordalo P, Gennaioli N, Porta RL, Shleifer A. 2019. Diagnostic expectations and stock returns. *Journal of Finance* 74(6):2839–2874

Bordalo P, Gennaioli N, Shleifer A. 2018. Diagnostic expectations and credit cycles. *Journal of Finance* 73(1):199–227

Bordalo P, Gennaioli N, Shleifer A. 2022. Overreaction and diagnostic expectations in macroeconomics. *Journal of Economic Perspectives* 36(3):223–244

Bouchaud JP, Krueger P, Landier A, Thesmar D. 2019. Sticky expectations and the profitability anomaly. *Journal of Finance* 74(2):639–674

Bowers JS, Davis CJ. 2012. Bayesian just-so stories in psychology and neuroscience. *Psychological bulletin* 138(3):389

Brunnermeier MK, Parker JA. 2005. Optimal expectations. *American Economic Review* 95(4):1092–1118

Caplin A, Leahy JV. 2019. Wishful thinking. NBER WP 25707

Chapman J, Dean M, Ortoleva P, Snowberg E, Camerer C. 2023. Econographics. *Journal of Political Economy Microeconomics* 1(1):115–161

Charness G, Dave C. 2017. Confirmation bias with motivated beliefs. *Games and Economic Behavior* 104:1–23

Coibion O, Gorodnichenko Y. 2012. What can survey forecasts tell us about information rigidities? *Journal of Political Economy* 120(1):116–159

Daniel K, Hirshleifer D, Subrahmanyam A. 1998. Investor psychology and security market under- and overreactions. *Journal of Finance* 53(6):1839–1885

de Clippel G, Zhang X. 2022. Non-bayesian persuasion. *Journal of Political Economy* 130(10):2594–2642

DeGroot M. 1974. Reaching a consensus. *Journal of the American Statistical Association* 69(345):118–121

Diaconis P, Zabell S. 1986. Some alternatives to bayes's rule, In *Proc. Second University of California, Irvine, Conference on Political Economy*, pp. 25–38

Dominiak A, Kovach M, Tserenjigmid G. 2021. Inertial updating with general information. Mimeo, UCSC

Dominiak A, Kovach M, Tserenjigmid G. 2023. Inertial updating. Mimeo, UCSC

Dominiak A, Lee D. 2023. Testing rational hypotheses in signaling games. *European Economic Review* 160:104610

Doya K. 2007. Bayesian brain: Probabilistic approaches to neural coding. MIT press

Edwards W. 1968. Conservatism in human information processing. *Formal representation of human judgment*

Eliaz K, Spiegler R. 2020. A model of competing narratives. *American Economic Review* 110(12):3786–3816

Enke B, Graeber T. 2023. Cognitive uncertainty. *Quarterly Journal of Economics* 138(4):2021–2067

Epstein LG. 2006. An axiomatic model of non-Bayesian updating. *Review of Economic Studies* 73(2):413–436

Epstein LG, Noor J, Sandroni A. 2008. Non-Bayesian updating: a theoretical framework. *Theoretical Economics* 3(2):193–229

Epstein LG, Noor J, Sandroni A. 2010. Non-Bayesian Learning. *The B.E. Journal of Theoretical Economics (Advances)* 10(1)

Ernst MO, Banks MS. 2002. Humans integrate visual and haptic information in a statistically optimal fashion. *Nature* 415(6870):429–433

Esponda I, Pouzo D. 2016. Berk–nash equilibrium: A framework for modeling agents with misspecified models. *Econometrica* 84(3):1093–1130

Fischhoff B, Beyth-Marom R. 1983. Hypothesis evaluation from a bayesian perspective. *Psychological Review* 90(3):239

Foster DP, Young HP. 2003. Learning, hypothesis testing, and Nash equilibrium. *Games and Economic Behavior* 45(1):73–96

Frick M, Iijima R, Ishii Y. 2023. Belief convergence under misspecified learning: A martingale approach. *The Review of Economic Studies* 90(2):781–814

Friston K. 2009. The free-energy principle: a rough guide to the brain? *Trends in cognitive sciences* 13(7):293–301

Friston K. 2012. The history of the future of the bayesian brain. *NeuroImage* 62(2):1230–1233

Friston KJ, Stephan KE. 2007. Free-energy and the brain. *Synthese* 159:417–458

Fryer R, Jackson MO. 2008. A categorical model of cognition and biased decision making. *The BE*

Journal of Theoretical Economics 8(1)

Fryer RG, Harms P, Jackson MO. 2019. Updating beliefs when evidence is open to interpretation: Implications for bias and polarization. *Journal of the European Economic Association* 17(5):1470–1501

Fuster A, Laibson D, Mendel B. 2010. Natural expectations and macroeconomic fluctuations. *Journal of Economic Perspectives* 24(4):67–84

Gabaix X. 2014. A sparsity-based model of bounded rationality. *Quarterly Journal of Economics* 129(4):1661–1710

Gabaix X. 2019. Behavioral inattention. In *Handbook of behavioral economics: Applications and foundations 1*, vol. 2. Elsevier, 261–343

Gabaix X, Laibson D. 2022. Myopia and discounting. Mimeo, Harvard University

Gagnon-Bartsch T, Bushong B. 2022. Learning with misattribution of reference dependence. *Journal of Economic Theory* 203:105473

Gagnon-Bartsch T, Rabin M, Schwartzstein J. 2021. Channeled attention and stable errors. Mimeo, Harvard University

Galperti S. 2019. Persuasion: The art of changing worldviews. *American Economic Review* 109(3):996–1031

Gennaioli N, Shleifer A. 2010. What comes to mind. *Quarterly Journal of Economics* 125(4):1399–1433

Gennaioli N, Shleifer A. 2018. A crisis of beliefs: Investor psychology and financial fragility. Princeton University Press

Ghirardato P. 2002. Revisiting Savage in a conditional world. *Economic Theory* 20:83–92

Gilboa I, Marinacci M. 2013. Ambiguity and the bayesian paradigm. In *Advances in Economics and Econometrics: Theory and Applications, Tenth World Congress of the Econometric Society*, eds. D Acemoglu, M Arellano, E Dekel, vol. 1. Cambridge University Press, 179–242

Glaeser EL, Nathanson CG. 2017. An extrapolative model of house price dynamics. *Journal of Financial Economics* 126(1):147–170

Gottlieb D. 2010. Will you never learn? self deception and biases in information processing. Mimeo, Princeton University

Grether DM. 1980. Bayes rule as a descriptive model: The representativeness heuristic. *Quarterly Journal of Economics* 95(3):537–557

Grether DM. 1992. Testing Bayes rule and the representativeness heuristic: Some experimental evidence. *Journal of Economic Behavior & Organization* 17(1):31–57

Griffin D, Tversky A. 1992. The weighing of evidence and the determinants of confidence. *Cognitive Psychology* 24(3):411–435

Griffiths TL, Kemp C, Tenenbaum JB. 2008. Bayesian models of cognition. In *The Cambridge handbook of computational cognitive modeling*, ed. R Sun. Carnegie Mellon University, 80–138

Griffiths TL, Tenenbaum JB. 2006. Optimal predictions in everyday cognition. *Psychological science* 17(9):767–773

Gul F, Pesendorfer W. 2001. Temptation and Self-Control. *Econometrica* 69(6):1403–1435

Heidhues P, Kőszegi B, Strack P. 2018. Unrealistic expectations and misguided learning. *Econometrica* 86(4):1159–1214

Hellman ME, Cover TM. 1970. Learning with finite memory. *Annals Mathematical Statistics* 41(3):765–782

Hong H, Stein JC. 1999. A unified theory of underreaction, momentum trading, and overreaction in asset markets. *Journal of Finance* 54(6):2143–2184

Hong H, Stein JC, Yu J. 2007. Simple forecasts and paradigm shifts. *Journal of Finance* 62(3):1207–1242

Jakobsen A. 2021. Coarse bayesian updating. Mimeo, Northwestern University

Jaynes ET. 2003. Probability theory: The logic of science. Cambridge university press

Jones M, Love BC. 2011. Bayesian fundamentalism or enlightenment? on the explanatory status and theoretical contributions of bayesian models of cognition. *Behavioral and brain sciences* 34(4):169–188

Kaanders P, Sepulveda P, Folke T, Ortoleva P, De Martino B. 2022. Humans actively sample evidence to support prior beliefs. *Elife* 11:e71768

Kahneman D, Frederick S. 2002. Representativeness revisited: Attribute substitution in intuitive judgment. *Heuristics and biases: The psychology of intuitive judgment* :49–81

Kahneman D, Tversky A. 1972. Subjective probability: A judgment of representativeness. *Cognitive psychology* 3(3):430–454

Kahneman D, Tversky A. 1973. On the psychology of prediction. *Psychological Review* 80(4):237

Kahneman D, Tversky A. 1983. Extensional vs. intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review* 91:293–315

Khaw MW, Stevens L, Woodford M. 2017. Discrete adjustment to a changing environment: Experimental evidence. *Journal of Monetary Economics* 91:88–103

Knill DC, Pouget A. 2004. The bayesian brain: the role of uncertainty in neural coding and computation. *TRENDS in Neurosciences* 27(12):712–719

Kording KP. 2014. Bayesian statistics: relevant for the brain? *Current opinion in neurobiology* 25:130–133

Körding KP, Wolpert DM. 2004. Bayesian integration in sensorimotor learning. *Nature* 427(6971):244–247

Kovach M. 2021. Conservative updating. Mimeo, Virginia tech

Kreps D. 1988. Notes on the Theory of Choice. Westview Press, Boulder

Kuhn TS. 1962. The structure of scientific revolutions. University of Chicago press

Kwisthout J, Wareham T, Van Rooij I. 2011. Bayesian intractability is not an ailment that approximation can cure. *Cognitive Science* 35(5):779–784

Mankiw NG, Reis R. 2002. Sticky information versus sticky prices: a proposal to replace the new keynesian phillips curve. *Quarterly Journal of Economics* 117(4):1295–1328

McGrayne SB. 2011. The theory that would not die: How bayes' rule cracked the enigma code, hunted down russian submarines, & emerged triumphant from two centuries of c. Yale University Press

Möbius MM, Niederle M, Niehaus P, Rosenblat TS. 2022. Managing self-confidence: Theory and experimental evidence. *Management Science* 68(11):7793–7817

Molavi P, Tahbaz-Salehi A, Jadbabaie A. 2018. A theory of non-bayesian social learning. *Econometrica* 86(2):445–490

Montiel Olea JL, Ortoleva P, Pai M, Prat A. 2022. Competing models. *Quarterly Journal of Economics* 137(4):2419–2457

Mullainathan S. 2002. A memory-based model of bounded rationality. *Quarterly Journal of Economics* 117(3):735–774

Mullainathan S, Schwartzstein J, Shleifer A. 2008. Coarse Thinking and Persuasion. *Quarterly Journal of Economics* 123(2):577–619

Myerson RB. 1986a. Axiomatic foundations of bayesian decision theory. *Center for Mathematical Studies in Economics and Management Science, Northwestern University Working Paper* 671

Myerson RB. 1986b. Multistage games with communication. *Econometrica* 54(2):323–358

Noor J, Payró F. 2022. An axiomatic approach to the law of small numbers. Mimeo, Boston University

Oaksford M, Chater N. 2007. Bayesian rationality: The probabilistic approach to human reasoning. Oxford University Press

Ortoleva P. 2012. Modeling the change of paradigm: Non-bayesian reactions to unexpected news. *American Economic Review* 102(6):2410–2436

Ortoleva P, Snowberg E. 2015. Overconfidence in political behavior. *American Economic Review* 105(2):504–535

Perea A. 2009. A model of minimal probabilistic belief revision. *Theory and Decision* 67:163–222

Phillips LD, Edwards W. 1966. Conservatism in a simple probability inference task. *Journal of experimental psychology* 72(3):346

Phillips LD, Hays WL, Edwards W. 1966. Conservatism in complex probabilistic inference. *IEEE Transactions on Human Factors in Electronics* (1):7–18

Pouget S, Sauvagnat J, Villeneuve S. 2017. A mind is a terrible thing to change: confirmatory bias in financial markets. *The Review of Financial Studies* 30(6):2066–2109

Rabin M. 2002. Inference by believers in the law of small numbers. *Quarterly Journal of Economics* 117(3):775–816

Rabin M. 2013. Incorporating limited rationality into economics. *Journal of Economic Literature* 51(2):528–543

Rabin M, Schrag JL. 1999. First Impressions Matter: A Model of Confirmatory Bias*. *Quarterly Journal of Economics* 114(1):37–82

Rabin M, Vayanos D. 2010. The gambler's and hot-hand fallacies: theory and applications. *Review of Economic Studies* 77(2):730–778

Sanborn AN, Chater N. 2016. Bayesian brains without probabilities. *Trends in cognitive sciences* 20(12):883–893

Sanborn AN, Griffiths TL, Navarro DJ. 2010. Rational approximations to rational models: alternative algorithms for category learning. *Psychological Review* 117(4):1144

Schwartzstein J. 2014. Selective attention and learning. *Journal of the European Economic Association* 12(6):1423–1452

Sepulveda P, Usher M, Davies N, Benson AA, Ortoleva P, De Martino B. 2020. Visual attention modulates the integration of goal-relevant evidence and not value. *Elife* 9:e60705

Sims C. 2003. Implications of Rational Inattention. *Journal of Monetary Economics* 50(3):665–690

Slovic P, Lichtenstein S. 1971. Comparison of bayesian and regression approaches to the study of information processing in judgment. *Organizational behavior and human performance* 6(6):649–744

Spiegler R. 2016. Bayesian networks and boundedly rational expectations. *Quarterly Journal of Economics* 131(3):1243–1290

Tenenbaum JB, Griffiths TL, Kemp C. 2006. Theory-based bayesian models of inductive learning and reasoning. *Trends in cognitive sciences* 10(7):309–318

Tenenbaum JB, Griffiths TL, et al. 2001. The rational basis of representativeness, In *Proceedings of the 23rd annual conference of the Cognitive Science Society*, vol. 6

Tenenbaum JB, Kemp C, Griffiths TL, Goodman ND. 2011. How to grow a mind: Statistics, structure, and abstraction. *science* 331(6022):1279–1285

Weinstein J. 2011. Provisional probabilities and paradigm shifts. Mimeo Northwestern University

Williams PM. 1980. Bayesian conditionalisation and the principle of minimum information. *The British Journal for the Philosophy of Science* 31(2):131–144

Wilson A. 2014. Bounded Memory and Biases in Information Processing. *Econometrica* 82(6):2257–2294

Woodford M. 2020. Modeling imprecision in perception, valuation, and choice. *Annual Review of Economics* 12:579–601

Yariv L. 2005. I'll see it when i believe it? a simple model of cognitive consistency. Mimeo, UCLA

Zhao C. 2018. Representativeness and similarity. Mimeo, University of Hong Kong

Zhao C. 2022. Pseudo-bayesian updating. *Theoretical Economics* 17(1):253–289