Good Guesses

Kevin Dorst and Matt Mandelkern*

Abstract

This paper is about *guessing*: how people respond to a question when they aren't certain of the answer. Guesses show surprising and systematic patterns that the most obvious theories don't explain. We offer a theory that does explain them: we propose that people aim to optimize a tradeoff between accuracy and informativity in forming their guess. After spelling out our theory, we use it to argue that guessing plays a central role in our cognitive lives. In particular, our account of guessing yields new theories of (1) belief, (2) assertion, and (3) the *conjunction fallacy*—the psychological finding that people sometimes rate conjunctions as more probable than their conjuncts. More generally, we suggest that guessing helps explain how boundedly rational agents like us navigate a complex, uncertain world.

1 Take a Guess

Where do you think Latif will go to law school? He's been accepted to the top four: Yale, Harvard, Stanford, and NYU; now he just has to choose. We don't know his preferences, but here's the data on where applicants who've had the same choice have gone in recent years:

	Yale	Harvard	Stanford	NYU	Total
Percentage:	38%	30%	20%	12%	100%

So take a guess: Where do you think he'll go?

Here are some observations. One natural guess is 'Yale'. Another is 'Either Yale or Harvard'. Meanwhile, it's decidedly *un*natural to guess 'not Yale', or 'Yale, Stanford, or NYU'. Though robust, these judgments are immediately puzzling. First, 'Yale' is a fine guess, but its probability is below 50%, meaning that its negation is strictly more probable (38% vs. 62%); nevertheless, 'not Yale' is a weird guess. Moreover, 'Yale or Harvard' is a fine guess—meaning that it's okay to guess something other than the single most likely school—yet 'Yale, Stanford, or NYU' is a weird guess (why leave out 'Harvard'?). This is so despite the fact that 'Yale or Harvard' is less probable than 'Yale, Stanford, or NYU' (68% vs. 70%).

*Authors contributed equally; author order was determined randomly. Dorst: Magdalen College, Oxford, OX1 4AU, United Kingdom, kevindorst0@gmail.com. Mandelkern: All Souls College, Oxford, OX1 4AL, United Kingdom; matthew.mandelkern@gmail.com. This project has benefited tremendously from more than the usual share of feedback. Many thanks to audiences at Dartmouth; All Souls College, Oxford; VirLaWP; Institut Jean-Nicod; the London Group for Formal Philosophy; and the University of Chicago; and to Kyle Blumberg, Lyle Brenner, Molly Brown, Fabrizio Cariani, Emmanuel Chemla, Hasan Dindjer, Cian Dorr, Daniel Drucker, Diego Feinmann, Vera Flocke, Brian Hedden, Daniel Hoek, Ben Holguín, Sophie Horowitz, Thomas Icard, Nathan Klinedinst, Joshua Knobe, Jason Konek, Harvey Lederman, Salvador Mascarenhas, Richard Pettigrew, Jonathan Phillips, Milo Phillips-Brown, Bernhard Salow, Weng Kin San, Ginger Schultheis, and Julia Staffel, for extensive feedback and discussion.

We'll generalize these patterns (§2), and offer an account that explains them (§3). The idea is that guessers aim to optimize a tradeoff between *accuracy* and *informativity*—between saying something that's as likely as possible to be true, but which is also as informative (specific) as possible. These goals directly compete—the more specific an answer is, the less probable it will typically be—and different guessers, in different contexts, will treat this broadly Jamesian tradeoff in different ways (James, 1956). Some will guess 'Yale'; others will guess 'Yale or Harvard'; still others will guess something else. But we'll show that, however they do so, optimizing this tradeoff is guaranteed to satisfy a variety of structural constraints on guesses that we'll bring out.

After offering this account, we'll use it to argue that guessing— along with its accuracy-informativity tradeoff— plays a central role in our cognitive lives.[1] First, we'll argue that our account of guessing offers a theory of *belief* (§4.1). In particular, Holguín (2020) argues that your beliefs are your best guesses, and shows how this account unifies recent observations about both the *weakness* and *question-sensitivity* of beliefs. Our account of guessing augments this account of belief by showing how to explain key structural features of guesses, and hence of beliefs.

Second, we'll show that our theory of guessing helps to both explain and generalize the standard pragmatic story about how conversations proceed (Stalnaker, 1978; Roberts, 2012), suggesting that guessing plays a central role in ordinary exchanges of information (§4.2).

Finally, we'll argue that our theory of guessing helps explain the *conjunction fallacy* (§4.3)—the psychological finding that people sometimes rank a conjunction $p \wedge q$ as more probable than one of its conjuncts p, contrary to the laws of probability (Tversky and Kahneman, 1983). Our explanation is built around the observation that it is perfectly permissible to *guess* an answer that is less likely, but more informative, than an alternative. ('Yale or Harvard' is a fine guess, even though it's less likely than 'Yale, Harvard, Stanford, or NYU'.) So we propose that the conjunction fallacy arises when subjects rate answers for their *quality as guesses* rather than for their *probability of being true*. This is often a mistake; but it's a mistake that's easy to explain if guessing plays the central role in our cognitive lives that we argue it does. Thus our account puts pressure on certain thoroughly irrationalist interpretations of the conjunction fallacy, and instead helps situate it within a bounded-rationality picture of the human mind (§5).

That is part of a broader narrative. We focus on these three applications—belief, assertion, and the conjunction fallacy—not only because they are intrinsically interesting, but also because they help to paint the bigger picture of this paper. On that picture, guessing is a cognitively basic activity—one that we constantly engage in as we think, talk, and reason. Moreover, it's an activity that *makes sense* to engage in, for it's part of how computationally limited creatures like us manage cope with an intractably complex and uncertain world (§5).

[1]Our interest in guessing was sparked by recent work on its role in epistemology by Horowitz (2017) and Holguín (2020). Our approach will differ from each of theirs in arguing (in §3) that we should explain guessing using the tools of epistemic utility theory (Levi, 1967; Joyce, 1998; Pettigrew, 2016a), and in using this explanation to illuminate both the pragmatics of assertion (§4.2) and the conjunction fallacy (§4.3).

2 What We Guess

We start with a simple question: what sorts of guesses do we tend to make? The answer is both surprising, and surprisingly systematic. In this section we bring out some of these patterns, drawing on observations from Tversky and Kahneman (1983) and Holguín (2020). Along the way, we explain why the most obvious theories of guessing won't predict them, setting the stage for our own theory, which comes in the next section.

Recall the case of Latif, who's been accepted to four law schools. Your credences—that is, your degrees of confidence—in where Latif will go are as follows:

Yale	Harvard	Stanford	NYU
38%	30%	20%	12%

Now suppose you are asked to guess where Latif will go. It seems like there is a range of answers that could reasonably be your guess, given your credences:

(1) a. Yale. ✓
b. Yale or Harvard. ✓
c. Yale or Harvard or Stanford. ✓
d. Yale, Harvard, Stanford, or NYU. ✓

A few notes: first, these are all meant to be elliptical for 'He will go to Yale', 'He will go to Yale or he will go to Harvard', etc. We'll use shorthands like this throughout. Second, there may be some subtle effects concerning the order of disjuncts, but we'll ignore them, assuming throughout that a disjunction $p \vee q$ is permissible iff $q \vee p$ is. Third, we'll move freely between speaking of answers as propositions and as the corresponding sentences. Fourth, consider the answer 'Yale, Harvard, Stanford, or NYU'. Although *saying* this in response to a request to guess is weird, as we'll discuss in §4.2 below, we think this is for independent pragmatic reasons. Moreover, there is a sense in which it is fine for this to *be* your guess: that is just to decline to take a stance beyond what you are sure of. We are primarily interested not in the speech act of guessing but in the *cognitive state* of guessing—of something being your answer to a question—and in this sense it is clearly a permissible guess. Finally, there are a variety of ways of eliciting guesses: besides explicitly asking for a guess, we could ask questions like 'Where do you think Latif will go?, 'Where's he likely to go', and so on. We'll shift freely between formulations like this, and come back to this point in due course.

There is also a range of answers that are intuitively unacceptable; for example:

(2) a. Harvard. ✗
b. Stanford. ✗
c. NYU. ✗
d. Yale or Stanford. ✗
e. Yale or NYU. ✗
f. Harvard or Stanford. ✗
g. Not Yale. ✗
h. Harvard, Stanford, or NYU. ✗
i. Yale, and it's cold in London today. ✗
j. Yale, or he has a birthmark on his left toe. ✗

To be clear, we are not claiming that people never have guesses like those in (2). Our claim is rather normative: there is something peculiar—something irrational—about guesses like this. Our ultimate aim will be to give a theory which predicts that it's epistemically irrational for any of these answers to be your guess about the question 'Where will Latif go?', given your credences—and, insofar as people's practices of guessing (and saying their guess) track this normative structure, predicts that they will tend to favor answers like those in (1) over those in (2).

On to generalizing the patterns from (1) and (2). It'll be helpful to lay out a formal model of what a *question* is and what its *answers* are, drawing on standard formulations from the semantics and pragmatics literature (Hamblin, 1973; Karttunen, 1977; Groenendijk and Stokhof, 1984). Start with a set of possibilities (possible worlds) which comprise all and only the worlds compatible with the assumptions in a given context—i.e. the context's *context set* (Stalnaker, 1974, 1978). In the case of a single person thinking to themselves, these will be the set of worlds compatible with what the guesser is certain of. In the case of a conversation, it will be the set of worlds compatible with what the interlocutors are (in some sense) commonly certain of. To keep things simple, we'll focus on the case in which the context set and the guesser's certainties coincide.[2]

A question is a partition of the context set: that is, a set of mutually exclusive and jointly exhaustive subsets of the context set. The cells of the partition are the *complete answers* to the question. So, in our example, we can model our question 'Where will Latif go to law school?' as a partition of the options we leave open in this context, namely the set of propositions {*Latif will go to Yale, He'll go to Harvard, He'll go to Stanford, He'll go to NYU*}. We'll assume that at any given point in a conversation, there's a *question-under-discussion* (QUD) that guessers aim to address (Roberts, 2012). Sometimes this QUD will be set by explicit questions like, 'Where will Latif go?'; but other times it'll be gleaned from other parts of the context—or, in the case of an individual guesser, by the structure and goals of their inquiry.

Finally, we'll assume that the guesser has credences which can be modeled with a probability function P that is regular over the context set: for any (relevant) proposition p that has a nonempty intersection with the context set, $P(p) > 0$. We will, likewise, assume that questions are always finite partitions.[3]

With this formalism in hand, let's draw some generalizations from the observations above, by way of exploring and rejecting some obvious theories of guessing. The most obvious starting place is the constraint that you should guess p only if you think that p is more likely than not to obtain. But, as we've seen, this is wrong: while some of the acceptable guesses above have a greater than 50% chance ('Yale or Harvard', 'Yale, Harvard, or Stanford, ...), others, like 'Yale' (38%), do not. So *you don't always have to guess an answer that is more likely than not to obtain* (Kahneman and Tversky, 1982):

> **Improbable Guessing**: It's sometimes permissible to guess p even when $P(p) < 0.5$.

Note, second, that judgments about the reasonableness of guesses depend substantially

[2]The constraints that we discuss here can all be generalized, with minor tweaks, to the case where the context set and the guesser's certainties don't coincide.

[3]Infinite questions would require some infinity-friendly measure on the QUD. Notably, all of our constraints on guessing could be stated in terms of a comparative confidence ordering, though the model (in §3) that justifies them requires more structure. It's an open question whether some refinement of our model could explain these constraints appealing only to a comparative confidence ordering.

on what question is being answered. Suppose that your credences are as above, but instead of being asked where you think Latif will go, you're asked: 'Will Latif go to Yale?', i.e. the question {*Yale, not Yale*}. Recall that you think there's a 38% chance he'll go to Yale, and a 62% chance that he won't. Given that, when addressing *this* question it seems like 'Yale' is not a very natural guess, since it is the substantially less likely of the two complete answers. Thus, again following Kahneman and Tversky (1982):

> **Question Sensitivity**: Whether p is a permissible guess depends not just on the guesser's credence in p but also in what question is being answered.

This means that—holding your credences fixed—p can be a permissible guess relative to one question, but impermissible relative to another.

(Of course, if you are first asked 'Where do you think will Latif go?', and your guess is 'Yale', then if you are *subsequently* asked, 'So, do you think he'll go to Yale?' it is very natural to say 'yes'. This is naturally explained by positing some 'stickiness' in the contextual resolution of the question under discussion: the first QUD 'Where will Latif go?' may remain salient even if a new question 'Will Latif go to Yale?' is explicitly asked. In general, there can be small mismatches between questions explicitly asked and the QUD. For instance, when you're asked 'Do you think p?' we want to say that the QUD is usually p?, not the partition over mental states induced by 'Do you think p?')

A natural way to account for Improbable Guessing and Question Sensitivity says that, given a question Q, your answer should be the complete answer you have highest credence in. This theory predicts that the only acceptable answer is 'Yale', so it rules out all the bad responses in (2). But in doing so it overgeneralizes: a range of answers—all those in (1)—are permissible. In fact, it looks like for any number of complete answers, there is a permissible answer which comprises the union of that number of cells: you can give a one-cell answer ('Yale'), a two-cell answer ('Yale or Harvard'), a three-cell answer ('Yale, Harvard, or Stanford'), or a four-cell answer ('One of those four'). This leads to our next generalization (Holguín, 2020):

> **Optionality**: Given any question Q, for any $k : 1 \leq k \leq |Q|$, it's permissible for your guess about Q to be the union of exactly k cells of Q.[4]

How might we capture Optionality? A natural thought is that guessers may guess any answer that is *likely enough*, i.e. that is more likely than some threshold set by the context. This would be a natural generalization of the Lockean Thesis, which says that you should believe any proposition that you have high enough credence in (Foley, 1992; Sturgeon, 2008; Leitgeb, 2014; Dorst, 2019). This theory accounts for Optionality, provided the relevant threshold can be sufficiently low. But this cannot explain why all the answers in (2) are *im*permissible. In our case, since 'Yale' is permissible, we'd have to say the threshold is below 0.38—but then we'd predict that answers like 'Yale or Stanford', or 'Harvard or Stanford' are permissible. The Lockean account doesn't have the resources to rule out these answers.[5]

What answers *are* permissible, then? The following constraint—also from Holguín (2020)—accounts for the patterns above:

[4]Cases in which the probabilities of complete answers are strongly 'clustered' put pressure on Optionality, as we discuss in §3.3, where we show how our model can account for intuitions about clustering in a way that is compatible with Optionality.

[5]*Sophisticated* Lockeans could do so by making the thresholds proposition-sensitive (Easwaran, 2016; Dorst, 2019)—but then they'd need a story for what controls these thresholds.

Filtering: An answer is permissible only if it is *filtered*: if it includes a complete answer q, it must include all complete answers that are more probable than q.
Precisely: p is filtered iff for any $q, q' \in Q$: if $P(q') > P(q)$ and $q \subseteq p$, then $q' \subseteq p$.

In other words, your guess can't include a complete answer q while excluding a strictly more likely complete answer q'.[6]

Why 'Filtering'? Imagine a filter through which the complete answers are strained. The 'size' of a complete answer corresponds to its probability. Whether the mesh of a filter lets such an answer through depends only its size. An answer is *filtered* iff, for some such mesh (some threshold of probability), the answer is the union of the complete answers that survive this filtering process (that are more likely than that threshold).

Filtering rules out the answers in (2-a)–(2-h) above: each of those answers is non-filtered. For instance, *Harvard or NYU* is non-filtered because it includes *NYU* as a subset, but does not include every complete answer which is more likely than *NYU*—it is missing both *Yale* and *Stanford*. Likewise, *Harvard* is non-filtered because it includes *Harvard* as a subset, but does not include the more likely complete answer *Yale*.

Optionality and Filtering together predict the admissibility of the answers in (1), together with the inadmissibility of the answers (2-a)–(2-h). The latter are all non-filtered; the former are all filtered. In fact, for each k between 1 and 4 (the size of the QUD, $|Q|$), there is exactly one filtered answer which is the union of k complete answers, and these are the answers in (1). In general—apart from cases of ties in probability among complete answers—there will be exactly one filtered answer to a question Q which is the union of k complete answers, for any k between 1 and $|Q|$.

What about answers like (2-i)–(2-j), e.g. 'Yale, and it's cold in London today' or 'Yale, or he has a birthmark on his left toe'? Intuitively, such responses include *irrelevant* material. In particular, they crosscut complete answers: (2-i) and (2-j) cannot be derived as a union of complete answers to the question under discussion. In general:

Fit: If an answer crosscuts a complete answer, it's impermissible.
Precisely: p is impermissible if there is a $q \in Q$ such that $q \nsubseteq p$ and $p \cap q \neq \varnothing$.

Though Fit is familiar from the literature on pragmatics (a point we come back to in §4.2), this constraint applies just as much to the cognitive act of guessing: if you formulate a guess, to yourself, about where Latif will go, it's bizarre for your guess to be (2-i) or (2-j).

An important complication: some apparent violations of Fit can be felicitous enough—'Latif will go to Yale, and I'm sure he'll love it!'; 'Latif will go to Yale or Harvard, and if he goes to Yale, he'll learn a lot'.[7] Nonetheless, other violations seem robustly bad, like those in (2-i)–(2-j). The standard explanation the felicity of the former answers is that it is easy to

[6]An alternative formulation of Filtering would say that so long as q' is *at least* as probable a q, it must also be included. Which version you'll like depends on what you think about Optionality in the case of ties. We think that if we're about to toss a fair coin, it's permissible to guess that it'll land heads—despite the fact that tails is equally likely—so we endorse Optionality even in this case, and thus use the strict-inequality version of Filtering. The weak-inequality version could be derived from our theory below if we imposed the constraint that your guess must *uniquely* maximize expected answer-value, in which case we would validate Optionality only when there are no ties. These choices don't matter for our central points.

[7]Another class of responses that violate Fit are ones like 'I don't want to guess'. As Diego Feinmann has pointed out to us, these feel like ways *opting out* of answering the question (cf. Dorst, 2014); from the point of view of the cognitive attitude, rather than speech act, these do not seem like guesses at all.

accommodate more fine-grained questions that are in a similar vein to the QUD (e.g. 'Where will Latif go, and will he like it?'); relative to the finer-grained question, the answer satisfies Fit. In contrast, (2-i)–(2-j) are infelicitous because the finer-grained question which would need to be accommodated to satisfy Fit (e.g. 'Where will Latif go, and what is the weather in London?') seem to have nothing to do with the original QUD. Given the robustness of many intuitions about Fit, we're inclined to think this is the best way to make sense of the overall picture here.[8]

We'll draw out a couple more constraints on guesses in §3 below, but for now we'll focus on Improbable Guessing, Question-Sensitivity, Optionality, Filtering, and Fit. To strengthen the case for these constraints, consider another example, drawn from recent experience. Consider a moment in the 2020 Democratic presidential primary when the only remaining plausible candidates were Biden, Sanders, Warren, and Bloomberg. Following FiveThirtyEight's model, your credences in who'll win a plurality of delegates are as follows:

Biden	Sanders	Warren	Bloomberg
37%	30%	17%	16%

What's your guess about who will win? It seems that all and only the following guesses are permissible:

(3) a. Biden. ✓
b. Biden or Sanders. ✓
c. Biden, Sanders, or Warren. ✓
d. Biden, Sanders, Warren, or Bloomberg ✓

Other guesses, like the following, are intuitively not permissible:

(4) a. Sanders. ✗
b. Warren. ✗
c. Bloomberg. ✗
d. Biden or Warren. ✗
e. Biden or Bloomberg. ✗
f. Sanders or Warren. ✗
g. Biden or Sanders or Bloomberg. ✗
h. Sanders or Warren or Bloomberg. ✗
i. Biden, and it will rain tomorrow. ✗

Once again, we see evidence of Improbable Guessing and Optionality in the range of permissible answers in (3). Filtering accounts for the infelicity of the answers in (4-a)–(4-h) (since each of these contains some complete answer as a part while leaving out some strictly more likely complete answer), and Fit accounts for the weirdness of (4-i). Based on cases like this, we think the principles above hold robustly (for more examples, see Holguín 2020).

These observations—Improbable Guessing, Question Sensitivity, Optionality, Filtering, and Fit—bring out what guesses people tend to make, revealing that guessing shows surprising yet systematic patterns. The most obvious accounts fail to predict these patterns, and it's by no means obvious how to explain them. In the next section, we'll show how to do so.

[8]But see Feinmann (2020) for a different take.

3 How We Guess

We'll explain these generalizations about guessing by giving a model of *how* we guess: how we balance considerations in a way that leads us to decide to guess that Latif'll go to Yale, or that he'll go to Yale or Harvard, or something else.

The basic idea behind our approach is a familiar Jamesian (1956) thought. A good guess about a question Q is a good picture of how things stand, Q-wise. In trying to form such a picture, there's an inevitable tradeoff between two goals. On the one hand, we want our picture to be *accurate*; we want our guess to be true. But being true often doesn't cut it. After all, one way to guarantee that your guess is true is to have it say very little: when asked 'Where do you think Latif will go?', 'Somewhere' is sure to be true—but is in some obvious sense unhelpful. We also want to *take a stand* on things—to have an *informative* guess, one that helpfully narrows down the space of alternatives we're considering. These two goals compete. Typically, the more informative your guess is ('He'll go to Yale'), the less likely it is to be true; the more likely it is to be true ('He'll go somewhere'), the less informative it is. On this way of construing James's view, trying to form a picture of the world involves trading off the impulses toward *informativity* (to believe substantive truths) and *accuracy* (to avoid error). In this section, we'll develop this idea of an accuracy-informativity tradeoff to give an account of guessing.

To be clear: our model of guessing is intended to be a computational level explanation, in the sense of Marr (1982). Our question is, *What problem is the mind solving when it forms a guess?* and our answer is, *How to optimally trade off accuracy and informativity.* As always, this style of explanation is neutral on the precise algorithms through which the mind solves this problem, as well as on the question of whether the processes involved will be consciously accessible or not.

3.1 Jamesian guessing

Our approach will be to view guessing as a kind of epistemic decision problem.[9] First, we'll try to say what makes a guess *objectively valuable* (foreshadowing: true guesses are better than false ones; and among true guesses, the more informative the better). Then we'll propose that people aim to maximize this objective value by choosing a guess with the highest *estimated* value (given their credences).

So suppose you are trying to guess the answer to a question Q—say, 'Where do you think Latif will go?'. Your guess could be any proposition: 'Yale'; 'Stanford'; 'I like cheese'; etc. How good your guess is depends on how well it answers the question. That in turn depends on whether your guess is true or false, and how valuable that kind of guess would be in case it is true and in case it is false. Schematically, let $V_Q(p)$ be a function which yields

[9] We take inspiration from the *epistemic utility theory* literature that uses decision-theoretic tools to explain the constraints of epistemic rationality (Joyce, 1998; Pettigrew, 2016a). Though most of this literature focuses on degrees of belief, some of it focuses on qualitative states of 'outright belief' that are similar to guesses—see Hempel (1962); Levi (1967); Maher (1993); Easwaran (2016); Pettigrew (2016b); Dorst (2019). Levi (1967) is by far the closest precedent for our approach—see footnote 11 for a comparison. As mentioned in footnote 1, the work in (Sliwa and Horowitz, 2015; Horowitz, 2017; Builes et al., 2020) is closely related, but that work puts guessing to a rather different use—arguing that the relationship between credences and guessing can help obviate the need for the tools of epistemic utility theory. By contrast, we'll argue that it is precisely these tools that are needed to explain the relationship between credences and guessing.

the *answer-value* of guessing p in response to question Q. Whenever you're unsure whether p is true, you'll be unsure how much answer-value it has. Nevertheless, you can use your credences in the various possibilities to form an *estimate* about how much answer-value it has—p's *expected answer-value*, written $E_Q(p)$. Precisely, we'll assume we can model any permissible measure of answer-value with a real-valued function $V_Q(p)$, such that if p is true, guessing it yields answer-value $V_Q(p) = V_Q^+(p)$, and if it's false, guessing it yields answer-value $V_Q(p) = V_Q^-(p)$. Using our guesser's (probabilistic) credence function, we'll assume that expectations are defined in the standard way (assuming act-state independence, for simplicity). Thus the expected answer-value of p is a weighted average of the various possible values $V_Q(p)$ might take, with weights determined to how likely they are to obtain:

$$E_Q(p) \ := \ P(p) \cdot V_Q^+(p) \ + \ P(\overline{p}) \cdot V_Q^-(p)$$

The core of our theory says that you must make a guess that maximizes this quantity, relative to some epistemically permissible measure of answer-value:

> **Guessing as Maximizing:** A guess is epistemically permissible given a question iff it has maximal expected answer-value relative to that question, for some permissible measure of answer-value.

The crucial question: Which measures of answer-value are epistemically permissible? After offering an answer to that question in this subsection, we'll go on to show in the next two how it explains our observations about guessing.

V_Q is a measure of the value of your guess—of how good your guess is, given the QUD Q. True guesses are better than false ones, so any permissible V_Q must be *truth-directed*:

> V_Q is **truth-directed** iff any true guess has higher answer-value than any false guess. Precisely: for all p, r: $V_Q^+(p) > V_Q^-(r)$.

Truth matters. But—on our Jamesian picture—truth isn't *all* that matters. Answer-value also depends on informativity. The informativity of a guess depends on what question it's answering: 'Latif will go to Yale' is an informative answer to 'Where will Latif go?'—but an uninformative answer to 'What are we having for dinner tonight?'. So V_Q^+ and V_Q^- should be sensitive to the informativity of the answer, which, in turn, depends on Q.

In fact, holding fixed a guess's truth-value, Q is arguably the *only* thing V_Q should be sensitive to. Suppose you ask who will win the election. In *some* sense, 'Latif will go to Yale and my grandpa was bald' is more informative than 'Latif will go to Yale'—but this doesn't seem to be the sense of informativity that governs guesses. After all, if you wanted to know about my grandpa, you would've asked! Similarly, suppose that you guess that Latif will go to NYU. In *some* sense, this is more informative than 'Latif will go to Yale', insofar as learning that it's true would lead to a bigger change in your probabilities than would learning that Latif will go to Yale. But there is another, equally intuitive sense in which this is not a more informative guess than 'Yale': both guesses are maximally informative about the question asked.[10] Thus a natural thought is that how informative a guess is with respect to Q depends

[10]More generally, the natural alternative to question-based measures are *credal*-based ones, which measure accuracy by how well your guess promotes some desirable quality in your interlocutor's credence function. There are many such measures—(Shannon) information (Shannon, 1948), probability gain (Baron, 1985), accuracy

on only the number of answers to Q it rules out. Precisely, given a question Q and a guess p, let the **informativity** of p relative to Q be the proportion of complete answers to Q that p rules out: $Q_p := \frac{|\{q \in Q:\ p \cap q = \emptyset\}|}{|Q|}$. For example, if Q is 'Where will Latif go to law school?', then $Q_{Yale} = Q_{Stanford} = Q_{Harvard} = Q_{NYU} = \frac{3}{4}, Q_{Yale\ or\ Harvard} = \frac{1}{2}$, etc. If Q is 'Who will win a plurality of delegates in the Democratic primary?', then $Q_{Biden} = Q_{Sanders} = Q_{Warren} = Q_{Bloomberg} = \frac{3}{4}$, $Q_{Biden\ or\ Sanders} = Q_{Sanders\ or\ Warren} = \frac{1}{2}$, and so on. Given this, our second constraint is that, given the truth-value of p, $V_Q(p)$ should then be fully determined by p's informativity:

> V_Q is **question-based** iff for all p: $V_Q(p)$ is fully determined by p's informativity together with its truth-value.
> Precisely: for all p, r, if $Q_p = Q_r$, then $V_Q^+(p) = V_Q^+(r)$ and $V_Q^-(p) = V_Q^-(r)$.

Our first main addition to the Guessing as Maximizing account is this: *a measure of answer-value is (epistemically) permissible only if it is truth-directed and question-based.* Why? Truth-directedness is straightforward, but the requirement that V_Q be question-based is somewhat more surprising. Our primary argument for it is an inference to the best explanation: as we'll see in §3.2, any question-based measure will offer a simple explanation of two of our most distinctive constraints—Fit and Filtering—which is not available to non-question-based measures.

So suppose V_Q is truth-directed and question-based. Although this will establish Fit and Filtering, it doesn't yet say anything about our observed *permissions*—that sometimes permissible guesses can be less likely than not (Improbable Guessing), and that a variety of guesses are always permissible (Optionality). How can we account for these observations? A very permissive approach would be to say that *any* truth-directed, question-based measure of answer-value is permissible. It follows from our results below that this theory would yield all the observations about guessing mentioned in §2, so it is worth flagging this position as a natural, minimal version of our approach.

But we'll do more: we'll motivate a particular subclass of truth-directed, question-based measures as the epistemically permissible ones, which we call **Jamesian measures**. We do this for four reasons. First, giving a more specific class of measures helps bring out the basic idea behind these constraints in a more concrete way. Second, it is important to see that Improbable Guessing and Optionality will fall out of our approach even if we are not nearly so permissive about what measures are allowed (since these encode *permissions* rather than obligations, restricting the set of potential measures has the potential to invalidate them; we'll show that the class of Jamesian measures still validates them). Third, our preferred class of measures will yield an explanation of a further generalization about guessing that we'll bring out in §3.3. Finally, our account of the conjunction fallacy in §4.3 will rely on a

(Oddie, 1997; Carballo, 2018), etc.—but they are not well-suited for our purposes. They all involve quantifying how much *learning* (i.e. conditioning on) the answer to a question would improve a credence function. Yet our context involves *guessing* the answer—not learning it—and in general people shouldn't update their credence function by conditioning on guesses. (If we guess 'Latif will go to Yale, I think', then—even if you trust us—you shouldn't become certain he will!) In fact, in cases where the probabilities are common knowledge, often you shouldn't change your credences *at all* in response to a guess. If we all know this coin is 60% biased toward heads, then we know that you won't change your credences when we guess how it'll land. That means that if what we care about is the impact of our guess on your *credences*, then *any* guess is permissible (since none will have any effect) in this case. But that's wrong—'heads' is a permissible guess; 'tails' is not. This, in short, is why standard credal-based measures won't do for our purposes.

systematic theory how comparisons of expected answer-value will lead people to *rank* non-optimal guesses. We need a more specific account of answer-value in order to make concrete, empirically testable predictions about such rankings.[11]

However, the details of the model we will give are separable from many of the broader proposals of this paper. There are a variety of other sub-classes of the truth-directed, question-based measures that would predict (most of) our constraints on guessing—though none, we think, work quite as elegantly as the Jamesian ones we will explore presently.[12] There are obviously many different ways to work out the details of a model like this, and we hope that our discussion will invite the development and comparison of alternatives.

We also want to emphasize that the goal of these models is not to predict what particular people will guess in particular situations. Rather, the goal is twofold. First, the model aims to highlight the structural features that the practice of guessing is sensitive to, and thus *explain* why guesses tend to meet constraints like Filtering, Fit, Optionality, and so on. We think that seeing guesses as aiming to optimize a tradeoff between accuracy and informativity (in the sense above) sheds light on what might otherwise look like arbitrary patterns. Second, our model provides explanatory tools to predict a wide range of *tendencies* in guessing: how people, on average, tend to make guesses or rank alternative guesses.

We turn now to the exposition of Jamesian measures. We'll give away the ending here, and below explain the reasoning behind it for those interested. Jamesian measures are those for which, for all p: $V_Q^-(p) = 0$, and there is some $J \geq 1$ such that $V_Q^+(p) = J^{Q_p}$. The parameter

[11] This leads to the second of two core differences between our account of guessing and Levi's (1967) theory of belief—which also understands answer-value in terms of (truly) ruling out cells of a salient partition. The first difference is that Levi does not situate his account within the linguistic practices of guessing and asserting. As a result, he does not make use of the notion of a conversationally flexible question-under-discussion to generate the relevant partition, and therefore does observe or account for the constraints on guessing we've highlighted, such as Question-Sensitivity, Filtering, Fit, and Optionality. (In fact, his account is inconsistent with Optionality whenever there are ties in probability: if we're about to toss a fair coin, his approach would disallow you from guessing that it'll land heads.) The second difference is that Levi's approach focuses on the answers that *maximize* expected answer-value, and as a result gives implausible verdicts about *rankings* of expected answer-value. His formula for the expected answer-value of p reduces to $P(p) - \frac{[p]\cdot q}{n}$, where q is a 'boldness' parameter that can take any value between 0 and 1, n is the size of the relevant partition, and $[p]$ is the number of cells of the partition consistent with p. On this measure, the expected value of contradictions will always be 0, while that of contingent claims will often be *negative*. Example: if the question is 'Will Latif go to Yale?', the expected answer-value of 'Yes' is $0.38 - \frac{1\cdot q}{2}$, which is negative whenever $q > 0.76$. Thus Levi's approach predicts that when people rank guesses in terms of expected answer-value, they will sometimes rank 'He'll go to Yale, and he won't' as a better guess than 'He'll go to Yale'! As we'll argue in §4.3, such rankings of non-optimal guesses are crucial to offering a plausible account of the conjunction fallacy. Our account will generate rankings that predict a variety of empirical findings (§4.3), but examples like this show that Levi's account will not. We take this to show that Levi's (1985, 2004) suggestion that his specific account can explain the conjunction fallacy is incorrect—though we are obviously sympathetic to the more general idea.

[12] One subclass of measures that would work includes those that yield some positive constant c for truth, add informativity to that, and raise it to an exponent $t \geq 0$: $V_Q^+(p) = (c+Q_p)^t$ and $V_Q^-(p) = 0$. An interesting alternative is to shift the location of the variable parameter by moving to Rank Dependent Utility (RDU) theory (Quiggin, 1982; Buchak, 2013). RDU introduces a risk function r that is used to modify the weight of a given level of probability. Assuming that $V_Q^+(p) > V_Q^-(p)$, it sets your estimate for the value of guessing p to be $r(P(p)) \cdot V_Q^+(p) + (1 - r(P(p)) \cdot V_Q^-(p)$. We can let variations of r play roughly the role that J plays in our model: when r is convex, you're risk-seeking and care more about informative answers regardless of their low probability; when r is concave, you're risk-averse and care more about making sure your guess is true. If we treat any r function as epistemically permissible, we can use RDU to validate all our constraints and permissions on guessing with a simple question-based measure like $V_Q^+(p) = c + Q_p$, and $V_Q^-(p) = -b$ for positive constants c, b.

J represents the guesser's measure of the value of informativity, while Q_p is, again, the proportion of the QUD ruled out by p. This yields the following formula:

Jamesian Expected Answer-Value: $E_Q^J(p) \;=\; P(p)\cdot J^{Q_p} + P(\overline{p})\cdot 0 \;=\; P(p)\cdot J^{Q_p}$

Thus the expected answer-value of a guess is determined by two terms: its probability of being true—$P(p)$—and its answer-value-if-true—J^{Q_p}. When J is small, lowering the informativity Q_p of your guess will only lower J^{Q_p} a small amount—in the limiting case where $J = 1$, it won't lower it at all, and the way to maximize expected answer-value is to pick a maximally probable answer. Conversely, as J gets large, lowering the informativity of your guess will lower J^{Q_p} a large amount, and therefore the way to maximize expected answer-value is to pick an informative guess—in the limit, as $J \to \infty$, the way to do so is to pick a *maximally* informative (filtered) guess, regardless of how low its probability is.

Our formula $P(p)\cdot J^{Q_p}$ captures the Jamesian tradeoff between accuracy and informativity. Picking an uninformative but very probable guess ('He'll go somewhere') makes the right term (J^{Q_p}) small but the left term ($P(p)$) large; picking an informative but improbable guess ('He'll go to Yale') makes the right term large but the left term small. Making a good guess requires optimizing the tradeoff between these terms, in light of the probabilities P and your value of informativity J.

We can see this tradeoff graphically in the Latif case by plotting the expected answer-value of 'Yale' ('Y'), 'Yale or Harvard' ('Y or H'), etc. for various values of J (Figure 1). As J increases, the optimal tradeoff between accuracy and informativity shifts towards informativity: when $1.67 > J \geq 1$, the best guess is 'Yale, Harvard, Stanford, or NYU'; when $2.80 > J > 1.67$, it's 'Yale, Harvard, or Stanford'; when $10.25 > J > 2.80$, it's 'Yale or Harvard'; and when $J > 10.25$, it's 'Yale'.

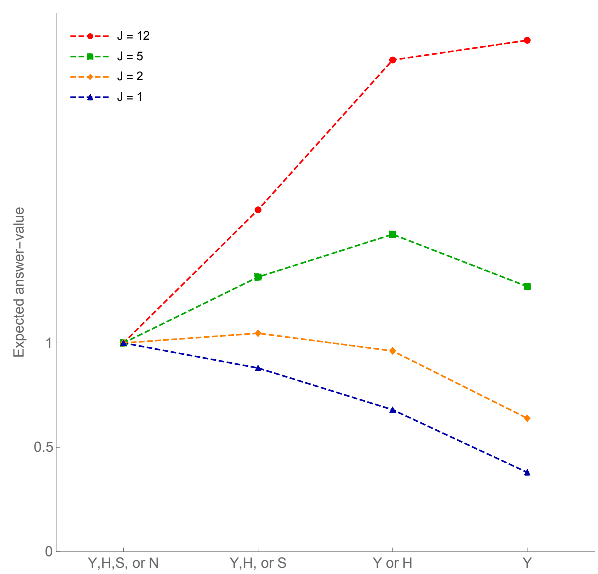

Figure 1: Expected answer-value of law-school guesses, varying the value of informativity (J)

This sums up our preferred way of measuring answer value. The rest of this subsection will explain the reasoning behind Jamesian measures; readers who are eager to see the applications of our model may wish to skip to the next subsection.

Begin with false guesses. How valuable is a false guess? Suppose we're asked where we think Latif will go; one of us says, 'Yale'; the other says 'Yale or Harvard'. Turns out, Latif goes to Stanford. Which of us was a more valuable guess, objectively speaking? Intuitively: neither! Both were maximally far from the truth—since both ruled out the true complete answer to the question. This intuition motivates setting $V_Q^-(p)$ equal to some constant low value, regardless of the informativity of p; we set this value to $V_Q^-(p) = 0$.[13] (This is not to say that all false guesses are the same: some false guesses will have higher *expected* answer-value than others, and thus will be better from that, subjective, perspective.)

What about true guesses? Since V_Q is question-based, $V_Q^+(p)$ will be a determined by the proportion of cells that p rules out, i.e. Q_p. How it's determined should depend on how much you value informativity. If you really value having a precise answer to Q, $V_Q^+(p)$ will increase very quickly with p's informativity; but if you don't particularly value such a precise answer, $V_Q^+(p)$ will increase much more slowly. By truth-directedness, any true answer—even an uninformative one—must have some minimal positive answer-value $t > 0$. This is the value of 'mere truth'. By what factor does guessing a *maximally* informative (true) answer improve on the value t of mere truth? That depends on how you value informativity. Let J be a real-valued parameter that measures this Jamesian value of (maximal) informativity, so that a maximally informative (true) answer yields value $J \cdot t$. Since a maximally informative (true) answer is at least as valuable as an uninformative one, $J \geq 1$.

How, exactly, should $V_Q^+(p)$ vary as both informativity (Q_p) and the value of informativity (J) change? Note that informativity, Q_p, has a minimum possible value of 0 (ruling out no complete answers), and has a least upper bound of 1, since a complete answer to Q has informativity $\frac{|Q|-1}{|Q|}$. If you don't care at all about informativity (if $J = 1$), then, no matter what Q_p is, $V_Q^+(p)$ should be the value of mere truth, i.e. t. Similarly, no matter how much you value informativity, if p is completely uninformative, then $V_Q^+(p)$ should again be t. Finally, as p becomes *maximally* informative, the value of truly guessing p should tend towards scaling the value of mere truth by a factor of J: as $Q_p \to 1$, we have $V_Q^+(p) \to J \cdot t$. A natural way to capture all these constraints is to raise J to the power of the (true) guess's informativity, and use the resulting value to scale the value of mere truth: $V_Q^+(p) = (J^{Q_p}) \cdot t$. When p has minimal informativity ($Q_p = 0$), $V_Q^+(p) = J^0 \cdot t = t$; likewise, if $J = 1$, then $V_Q^+(p) = 1^{Q_p} \cdot t = t$. And when informativity approaches its maximal value ($Q_p \to 1$), answer-value scales the value of mere truth by a factor of J: $V_Q^+(p) \to J^1 \cdot t = J \cdot t$. Generally: when J is small, answer-value rises slowly with increases in informativity; when J is large, it rises steeply; and as J gets arbitrarily large, increasing informativity dominates all other considerations (Figure 2).[14]

Summing up this discussion, we define our class of Jamesian measures of answer-value:

V_Q is **Jamesian** iff, for some $t > 0$ and $J \geq 1$:

$$V_Q = \begin{cases} V_Q^+(p) = J^{Q_p} \cdot t & \text{if } p \text{ is true} \\ V_Q^-(p) = 0 & \text{if } p \text{ is false} \end{cases}$$

[13] In some instances intuitions about verisimilitude may make some false guesses seem closer to the truth than others (Popper, 1963; Oddie, 2019; Schoenfield, 2019a); likewise, intutions about partial truth (Yablo, 2014); but we'll set such issues aside for our purposes.

[14] This, intuitively, is why we raise J to the power of Q_p rather than (say) multiplying them: we want high values of J to allow for small increases in informativity to matter more and more.

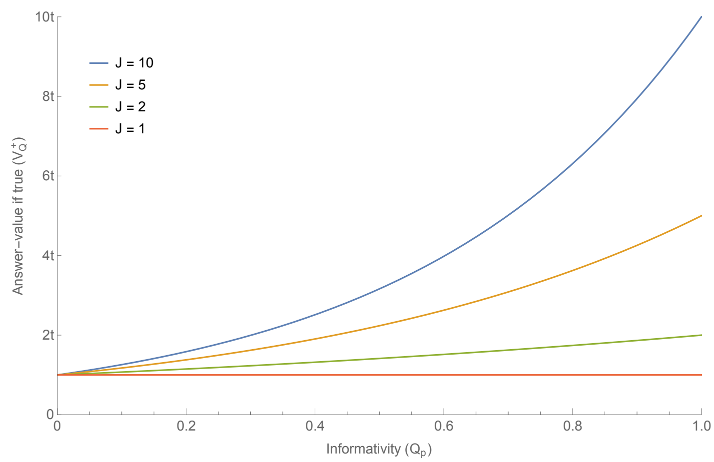

Figure 2: How answer-value varies with J and Q_p

Given a Jamesian measure, the expected answer-value of a guess reduces as follows:

$$\begin{aligned} E^J_Q(p) &= P(p) \cdot V^+_Q(p) \ + \ P(\overline{p}) \cdot V^-_Q(p) \\ &= P(p) \cdot (J^{Q_p} \cdot t) \ + \ P(\overline{p}) \cdot 0 \\ &= P(p) \cdot (J^{Q_p} \cdot t) \end{aligned}$$

Notably, comparisons of expected answer-value are insensitive to the value t of 'mere truth': for any p, r, $E_Q(p) > E_Q(r)$ iff $P(p) \cdot J^{Q_p} \cdot t > P(r) \cdot J^{Q_r} \cdot t$ iff $P(p) \cdot J^{Q_p} > P(r) \cdot J^{Q_r}$. So without loss of generality we can assume that $t = 1$, and simply say $V^+_Q(p) = J^{Q_p}$. Thus we arrive at the formula above for Jamesian expected answer value: $E^J_Q(p) = P(p) \cdot J^{Q_p}$

3.2 Deriving our constraints

We will thus adopt a theory on which p is epistemically permissible iff it has maximal expected answer-value relative to your credences and some Jamesian measure of answer-value V_Q—i.e. some setting of the value of informativity $J \geq 1$. In this section, we will show that the constraints and permissions observed in §2 all follow from this theory.[15]

Start with Fit—the claim that a guess is impermissible if it crosscuts a complete answer, like 'I think Latif will go to Yale, and it's cold in London today'. This follows easily. If p violates Fit, then there is some complete answer that it overlaps but does not fully include: there is a $q \in Q$ such that $q \cap p \neq \emptyset$ but $q \not\subseteq p$. Compare p to $p \cup q$. The two rule out exactly the same number of complete answers, and thus by question-basing, have the same answer-value if true and if false. But $p \cup q$ is more likely to be true than p (since we are assuming our probability measure is regular over the context set). Therefore the expected answer-value of

[15] To be clear, Filtering and Fit will follow for any truth-directed and question-based measures. We'll show that Improbable Guessing, Question Sensitivity, and Optionality also hold provided that all and only Jamesian measures are epistemically permissible—from which it follows that, if we treat a strictly larger class of measures as epistemically permissible, Optionality will still hold. Committing to Jamesian measures will yield helpful further predictions down the line, as we'll see.

$p \cup q$ is strictly higher than that of p. Any non-Fit answer will thus have lower expected answer-value than alternative some Fit answer, and thus will never be a permissible guess. (In this case, the expected answer-value of 'Yale' will always be higher than 'Yale, and it's cold in London today'.)

Next, Filtering: an answer is permissible only if it is filtered—if it includes a complete answer q, it must include all complete answers that are more probable than q. This follows from any truth-directed, question-based measure because swapping out the less-probable complete answer for the more-probable one maintains the same level of informativity, but increases the probability of your guess being true. (So a non-Filtered answer like 'Yale or Stanford' will always have lower expected answer-value than a filtered answer with the same size—in this case, 'Yale or Harvard'.)

In fact, Filtering is a special case of a more general constraint which is worth bringing out. If you are asked to *rank* the complete answers to a question, your ranking should follow the probabilities: 'Yale' is a better guess than 'Harvard', which is better than 'Stanford', etc. Likewise if you are asked to rank the two-cell answers: 'Yale or Harvard' is better than 'Yale or Stanford', and so on. Generally:

> **Filtered Rankings**: Equally informative answers should be ranked by probability. Precisely: if $Q_p = Q_r$, then $E_Q(p) > E_Q(r)$ iff $P(p) > P(r)$.

This follows from our model by the same reasoning.[16]

Consider next Improbable Guessing: sometimes it's permissible to guess p even if it's less likely than not to obtain. We've already seen that in our original Latif case, when $J \geq 10.25$, 'Yale' is the optimal guess—despite the fact that $P(\textit{Yale}) = 0.38$. The intuition is that high informativity can outweigh low probability, especially as J grows large.

The next observation was Question Sensitivity: whether a guess is permissible depends on what question it's addressing. This follows because Jamesian measures are based on a guess p's informativity Q_p, which in turn is determined by the question Q. For example, relative to the four-cell question, 'Where will Latif go?' (complete answers: $\{\textit{Yale, Harvard, Stanford, NYU}\}$), 'Yale' has informativity 3/4. But relative to the two-cell question 'Will Latif go to Yale?' (complete answers: $\{\textit{Yale}, \overline{\textit{Yale}}\}$), it has informativity 1/2. Thus 'Yale' can be a permissible answer to the former (as we have seen), but it can never be a permissible answer to the latter (because it's not filtered with respect to $\{\textit{Yale}, \overline{\textit{Yale}}\}$).

Our final observation was Optionality: given a question Q with $|Q|$ possible complete answers, for any $1 \leq k \leq |Q|$, it's permissible to give an answer that is a union of k-cells. In particular, the filtered answer that's a union of the the k most-probable cells is always permissible. Thus 'Yale', 'Yale or Harvard', 'Yale or Harvard or Stanford', and 'One of those four' are all permissible guesses. As we've seen above (Figure 1), each of these maximizes expected answer-value relative to certain values of J. The proof of Optionality requires some footwork (see the Appendix), but the basic idea is a straightforward generalization of this case. When J is low, being informative provides little additional value, so the best guess is an uninformative (but definitely true) guess. As J grows, being more informative gradually matters more and more such that—no matter your credences— you eventually start preferring

[16]This shows that our approach yields Horowitz's (2017) Lockean-like relationship between credences and guessing in a special case: *when p and r are equally informative*, then you should guess p over r if $P(p) > P(r)$. Thanks to Brian Hedden for pointing out this generalization of Filtering.

a $(|Q|-1)$-cell answer, then a $(|Q|-2)$-cell one, and so on until you prefer a 1-cell answer (see Figure 1, p. 12). We can thus rationalize guesses of different levels of informativity by ascribing to guessers different J-values: that is, different weights on informativity.

3.3 Setting J-values

This discussion raises a natural question: how are J-values set? And how do we know what subjects' J-values are?

We won't offer a definitive answer to either question, but we'll try to illustrate how the structure of our model allows us to say interesting things about this issue, and in so doing explain a further generalization about guessing.

First, it's worth re-emphasizing that we're focusing on the *cognitive attitude* of having something as your best guess about a question. We think this attitude is determined entirely by your J-values and credences, given a question. There can of course be all sorts of other (e.g. practical) values that affect the *speech act* of guessing. If we threaten to punch you unless you correctly guess the exact number of jellybeans in a jar, the rational response is obviously to pick some exact number—say, '457'—even if you have very little idea. Likewise: if we threaten to punch you if you make a mistake, then it makes sense to refrain from guessing—even if in fact you have an idea. But this doesn't imply that your J-value is high (/low), for in these cases the *speech act* of guessing plausibly comes apart from the corresponding cognitive attitude. This brings out that J-values aren't straightforwardly sensitive to practical stakes.

Instead, we take your J-value to be determined by your mental state in broadly the way credences, utilities, and (on some views) risk profiles are. But unlike standard interpretations of these states, we take J-values to be very flexible—able to adjust as desired as you switch amongst questions, or address the same question in different contexts. Our discussions later in the paper about the role of guessing (§4.1, §4.2, and §5) will give more of a picture for *why* J-values would need to be flexible like this. But we take it to be clear from our examples that they are: given the credences above, it's perfectly coherent to reply to 'Where do you think Latif will go?' with 'Yale, I think', and yet (perhaps in a different context) reply to 'Who do you think will win the primary?' with 'Biden, Sanders, or Warren'—despite the fact that these two guesses will require different J-values. In fact, this flexibility in J-values helps explain why, when people make guesses, their statements are often peppered with markers that flag various degrees of strength: 'He'll (definitely) go to Yale' vs. 'Yale, surely'; vs. 'I think Yale' vs. 'I'd guess Yale—but it's hard to say'. We think these markers are ways of flagging what J-value you're using, and therefore to what degree your guess is based on confidence in its truth vs. a desire to be informative.

Despite this flexibility, we *do* think there's something systematic to be said for how people tend to select their J-values. Consider a further observation about guessing. Although we think Optionality is true—any filtered guess is permissible—there are certain circumstances in which certain filtered guesses seem odd. In particular, as the probabilities of the various complete answers 'cluster' together more tightly, it becomes increasingly strange for your guess to crosscut these clusters—to include some but not all of the cells in a cluster. To see this, consider some variations on our law-school case in which your credences about where Latif are go are slightly different:

Scenario:	Yale	Harvard	Stanford	NYU
Original:	38%	30%	20%	12%
Close:	40%	35%	15%	10%
Near-Tie:	40%	39%	11%	10%
Tie:	40%	40%	10%	10%

Consider the guess 'He'll go to Yale, Harvard, or Stanford' (= 'not NYU'). This guess seems fine in the original case, a bit odd in the Close case, quite odd in the Near-Tie case, and pretty bizarre in the Tie case. We summarize this trend with the following generalization:

> **Clustering:** People tend to prefer guesses that don't crosscut clusters of complete answers with similar probabilities—i.e. guesses that include either all or none of them.

We state Clustering as a *tendency* because we don't think it imposes a hard constraint on permissible guesses. Two reasons. First, some guesses seem permissible even when they crosscut cells with the same probabilities: in response to 'How do you think this fair coin will land?', it seems perfectly permissible to guess 'Heads'. Even in this Tie case, it doesn't seem outright impermissible to guess 'Not NYU'. Second, Clustering needs to be a 'soft' constraint because its effects are *graded*: 'Yale, Harvard, or Stanford' gets progressively stranger as we move from the Original to the Close to the Near-Tie to the Tie case.

Our proposal is that Clustering reveals how people tend to select J-values—in particular, they tend to select a J-value that makes their guess *distinctive*: one that makes its expected answer-value not only maximal, but distinctively higher than that of alternative guesses. Note, for instance, that in the Tie case, 'Yale, Harvard, or Stanford' will always have the exact same expected value as 'Yale, Harvard, or NYU'—thus even if we pick a J-value that leads both of these to have maximal expected answer-value, neither can ever be *uniquely* maximal. Thus we think that guess is odd because there can be nothing to uniquely recommend it. (Similarly in the Near-Tie case, except that instead of the expected answer-values of 'Yale, Harvard, or Stanford' and 'Yale, Harvard, or NYU' being the *exact same*, they are merely very close—meaning neither can be very distinctive.) In contrast, for many values of J, in the Tie and Near-Tie cases, 'Yale or Harvard' has an expected answer-value that is substantially higher than any other potential answers. Thus our hypothesis is that Clustered guesses are natural because there's a way of valuing informativity that makes them distinctively best.

This notion of distinctiveness can be made precise as follows. Given credences P and a question Q, let the *J-distinctiveness* of a guess p, D_J^p, be the ratio of its expected answer-value to the highest expected answer-value of any other Fit guess (holding fixed J). That is, where F_p is the set of Fit answers to Q other than p, we have:

$$D_J^p := \frac{E_Q^J(p)}{\max\limits_{r \in F_p} \left(E_Q^J(r)\right)}.$$

And define the *distinctiveness* (period) of p, D^p, to be the maximal J-distinctiveness it can receive, for any value of J: $D^p := \sup\{D_J^p : J \geq 1\}$. So defined, we take the distinctiveness D^p of a guess to be a measure of how salient it is compared with the permissible alternatives. Our proposal is that there is a tendency (but not an obligation) to make guesses that are salient; and thus, *inter alia*, to make guesses with high distinctiveness; and thus to have J-values that allow guesses to have high J-distinctiveness.

This explains Clustering. To see why, note the following. $D^p \geq 1$ iff p is filtered.[17] If p includes complete answer q_1, excludes q_2, and $P(q_1) = P(q_2)+\epsilon$, then p's distinctiveness is no greater than $\frac{P(p)}{P(p)-\epsilon}$.[18] Thus if p crosscuts a cluster of equally-probable complete answers, its distinctiveness is the minimal value of 1. And when it crosscuts a cluster of *almost*-equally-probable answers, its distinctiveness will be only marginally greater than 1: ϵ will be small, so $D^p \leq \frac{P(p)}{P(p)-\epsilon} \approx 1$. Meanwhile, when the least-probable cell included in p is significantly more probable than the most-probable cell outside it (i.e. p doesn't crosscut any clusters), then D^p is substantially larger than 1. For example, in the Tie case, 'Yale or Harvard' has the highest distinctiveness of any guess—at 1.33—while that of 'Yale, Harvard, or Stanford' has the minimal distinctiveness of 1; in the Near-Tie case, the distinctiveness of the former is 1.32 while that of the latter is 1.01; in the Close case, the distinctiveness of the former is 1.25 while that of the latter is 1.04; and so on.[19] Thus a preference for distinctive guesses explains Clustering.[20]

Upshot: although J-values are flexible, the structure of Jamesian measures offers the resources to help explain how people select them, and in so doing explains the Clustering generalization. We take this result to both clarify and bolster the case for the Jamesian Guessing-as-Maximizing account.

4 When We Guess

So far we've offered an account of *what* we guess, and *how* we do so. We now turn to *when* we guess, exploring the role that guessing plays in our cognitive lives. We'll make the case that much of our cognitive lives involves trading off accuracy and informativity, i.e. making good guesses from limited information—and that, therefore, our theories of human cognition and conversation should give a privileged role to guesses. In particular, we'll argue that guessing plays a central role in *believing* (§4.1), *talking* (§4.2), and *reasoning* (§4.3).

Although we're going to make the strongest case we can for each of our applications, we want to emphasize their modularity. You might be convinced by our theory of guessing (§§1–3) and be unconvinced by our applications; you might be convinced by some but not others. Moreover, we want to flag that our applications come in reverse order of originality: §4.1 primarily refines existing ideas in the literature on belief; §4.2 proposes a revision to—and a

[17]If p is not filtered, it can never be maximal in expectation, so $D^p < 1$. And if p is filtered, then by Optionality, there is a setting of J on which it has maximal expectation, so $D^p \geq 1$.

[18]A relevant alternative to p is a guess p^* that swaps our q_2 for q_1. Since $Q_p = Q_{p^*}$, that means for any value of J, $D^p_J \leq \frac{E^J_Q(p)}{E^J_Q(p^*)} = \frac{P(p)\cdot J^{Q_p}}{P(p^*)\cdot J^{Q_p}} = \frac{P(p)}{P(p^*)} = \frac{P(p-q_1)+P(q_1)}{P(p-q_1)+P(q_2)} = \frac{P(p-q_1)+P(q_1)}{P(p-q_1)+P(q_1)-\epsilon} = \frac{P(p)}{P(p)-\epsilon}$.

[19]Supposing p is filtered, a rather complicated proof and calculation (which we omit) offers the following formula for D^p. Number the cells $q_1, ..., q_n$ of Q such that $P(q_1) \geq ... \geq P(q_n)$ (it doesn't matter how ties are ordered), and define $Q^k := q_1 \cup ... \cup q_k$. If p includes k cells, then for all J, $E^J_Q(p) = E^J_Q(Q^k)$. Let $p' := (Q^k - q_k) \cup q_{k+1}$ (where $q_{k+1} = \emptyset$ if $k+1 > n$). Then if $k = 1$ or $k = n$, $D^p = \frac{P(p)}{P(p')}$; and if $1 < k < n$, then $D^p = \min\left\{\frac{P(p)}{P(p')}, \frac{P(p)}{\sqrt{P(Q^{k-1})P(Q^{k+1})}}\right\}$.

[20]An alternative approach to clustering would be based on the fact that picking a non-clustered answer often requires having a J-value in a very specific range. It may be natural to think of people's J-values as standardly *imprecise*—that is, as comprising a range of values rather than a specific value; and it may be that non-clustered answers require the agent to have an unduly narrow (and therefore arbitrary-seeming) range of values. Thanks to Jason Konek for this suggestion.

new explanation of—the standard pragmatics of assertion; §4.3 develops a new theory of the conjunction fallacy. Regardless of which of these applications you find plausible or exciting, we hope to convince you that understanding guessing in terms of expected answer-value helps us pose and address a variety of fruitful questions in epistemology, philosophy of language, and cognitive science.

4.1 Guess when you believe?

Start with belief. Following the recent literature, we'll assume that believing that p is thinking that p, and move freely between the two.[21] We want to call attention to two relevant threads in the recent literature: the *weakness* and *question-sensitivity* of belief.

Start with weakness. At least in the sense of 'belief' referred to by the natural-language term, believing p doesn't require having a particularly strong attitude toward p. It doesn't require knowing or being sure, for it is perfectly coherent to say, 'I $\left\{ \begin{smallmatrix} \text{don't know} \\ \text{am not sure} \end{smallmatrix} \right\}$ if it'll rain, but I $\left\{ \begin{smallmatrix} \text{think} \\ \text{believe} \end{smallmatrix} \right\}$ it will' (Hawthorne et al., 2016). Nor does it require having non-statistical evidence, since it's perfectly sensible to say 'I think your lottery ticket will lose'. In fact, believing that p doesn't even seem to require believing that p is more likely than not! For in response to the question, 'Where do you think Latif will (likely) go?', it's reasonable to reply, 'I $\left\{ \begin{smallmatrix} \text{think} \\ \text{believe} \end{smallmatrix} \right\}$ he'll (likely) go to Yale' (Kahneman and Tversky, 1982; Hawthorne et al., 2016; Dorst, 2019; Moss, 2019; Rothschild, 2019). In fact, as Holguín (2020) brings out, this seems true *no matter how unlikely Latif is to go to Yale*, so long as it's the most likely option (cp. Windschitl and Wells, 1998).[22]

However, you can think Latif will go to Yale when it's less likely than not only if you're addressing the right question—this is where question-sensitivity enters the picture.[23] Although it's fine to say that you think that Latif will go to Yale in response to the question 'Where do you think he'll go?', if you're instead asked 'Do you think Latif will go to Yale, or not?' it's much more natural to that say that you do not. A simple explanation of this contrast is that belief is not simply a relation to a proposition, but rather a relation to a proposition, relative to a question. Thus *beliefs are answers to questions.*

Drawing these two threads together, Holguín (2020) proposes that to believe p relative to question Q is for p to be entailed by your best guess about Q. The key motivation is that best guesses, Holguín observes, obey (normatively speaking)z the distinctive features of both Filtering and Optionality—*and so do beliefs.* For example, with respect to the question 'Where do you think Latif will go?', it's permissible for the strongest thing you think (believe) to be that he'll go to Yale, or that he'll go to Yale or Harvard—but it's not permissible for it to be that he'll go to Harvard, nor that he'll go to Harvard, Stanford, or NYU. Thus both guessing and believing are weak and question-sensitive in the same ways—so it's natural

[21] See Hawthorne et al. (2016); Dorst (2019); Rothschild (2019); Holguín (2020) for extensive discussion.

[22] Some have replied to this 'belief is weak' picture by arguing that the intended reading of the philosophical term 'belief' is much stronger—something like *being sure* (Clarke, 2013; Greco, 2015; Friedman, 2019; Moss, 2019; Williamson, 2020). Although we'll say more about this in §4.2 and §5, for now we focus on the natural-language, weak reading of 'belief'.

[23] E.g. Levi 1967; Kahneman and Tversky 1982; Thomason 1986; Yalcin 2011, 2018; Drucker 2020; Hoek 2020a; Holguín 2020; cf. also Schaffer 2005, 2007; Schaffer and Knobe 2012; Schaffer and Szabó 2014; Gerken and Beebe 2016 for closely related discussion. The theory of belief in Leitgeb 2017 is also partition-sensitive, but in very different ways (for example, it requires beliefs to be more probable than not).

to hypothesize that your beliefs *just are* your best guesses. (For other arguments, see the references in footnote 23.)

Our account of guessing combines nicely with Holguin's account of belief. While his account explains the weakness and question-sensitivity of belief in terms of guessing, our account of guessing explains *why* guesses (and hence beliefs) have these features: in forming your best guess—that is, what you think—about Q, you must maximize expected answer-value relative to some Jamesian measure.[24] Upshot: if we follow Holguín in analyzing believing in terms of guessing, and add our account of guessing, we get an explanatory account of both the weakness of beliefs, and the particular ways in which they are question-sensitive—including Filtering, Optionality, and (we add) Fit, and Clustering.[25]

4.2 Guess when you talk?

Turn now from thinking to talking. We suspect that guessing plays a key role in ordinary exchanges of information. *Explicit* requests for guesses are not all that common in ordinary conversation—but it is *very* common to ask or report what someone *thinks* or *believes* or *thinks is likely* about some question. A natural thing to ask about Latif is *where you think* he'll go to law school; and a natural reply is that *you think* he'll go to Yale, even if you're not sure. Regardless of whether we're right that you think p iff it's entailed by your guess (§4.1), we take our examples to have shown that these are natural ways to get your interlocutor to take a guess —after all if we ask you 'Where do you think Latif will go?', we'll be unphased if you give an improbable but filtered answer ('Yale'), yet puzzled if you give a probable but unfiltered one ('Harvard, Stanford, or NYU').[26] Since these common questions elicit guesses, it seems that guessing *does* show up frequently in ordinary conversation.

But we want to propose that guessing plays an even deeper role in communication:

> **Say Your Guess:** When the question under discussion is Q, you should communicate your best guess about Q.

In particular, we propose that there are two different ways to communicate your guess, with two different conversational effects. Suppose we ask, 'Where will Latif go to law school?' One way to reply is to *merely* guess—e.g. 'I think he'll go to Yale'. This sort of reply registers your commitment and conveys your opinion, but doesn't update the common ground (Stalnaker, 1978). (You could follow up with: 'But of course he might go somewhere else'.) But there's

[24] An open question is the post-semantic issue of where the J-value comes into the semantic calculations of attitude ascriptions: is it supplied by the subject, or by the context of assertion or evaluation? The former answer fits naturally with the picture her, and we think standard arguments against subject-sensitive invariantism are not compelling in the case of 'believe' or 'think': 'If more was riding on it, I wouldn't think that the bank's open—I'd suspend judgment until I double-checked' seems coherent. But our view is also consistent with contextualist or relativist treatments of J-values.

[25] A final epistemological note: our account also helps with a classic problem in Jamesian, 'believe-truths-avoid-falsehoods' epistemology (Elgin, 2017; Carballo, 2018)— namely, *which* truths and *which* falsehoods? Our approach offers an answer, since the epistemic value of a guess is determined by the question under discussion. Given an account of how these questions should be determined—say, by conversational dynamics or by the structure of inquiry—our account explains why we should care about some truths and not others.

[26] Even if the account in §4.1 is right—so you believe anything that is entailed by your guess to Q—a residual question remains: when asked what you thinks about Q, why isn't any answer entailed by your guess (e.g. 'Yale, and I like pancakes') permissible? We take it this is naturally explained by some version of Grice's Maxim of Quantity, which requires you to say the *strongest* thing you believe about Q—i.e. your guess.

another way to communicate your guess: you can use it as a proposal to update the common ground; you can *assert* your guess—e.g. ‘He’ll go to Yale, Harvard, or Stanford’. Following the standard idea that knowledge is closely linked to assertability (Williamson, 2000; Brown, 2008), we propose that you should only use your guess to update the common ground if you *know* your guess (or—for an internalist version—only if you *think* you know your guess).

Why think that Say Your Guess is correct? We think it should be uncontroversial that there are these two ways of responding to a question—one that contains explicit hedges (‘I think p’), and one that does not (‘p’). These are both clearly ways of addressing the question.[27] Moreover, we think the data we’ve presented so far makes it clear that the former speech act is an expression of a guess in our sense—after all, we’ve seen that such hedged (‘I think...’) responses must conform to Filtering, Fit, Optionality, etc. The controversial idea here is that an *assertion* is also the expression of a kind of guess—in particular, a guess which you (think that you) know to be true—and thus that asserting and merely guessing are two ways of conforming to a broader norm. Why accept this more controversial part of Say Your Guess?

Compare it to the standard pragmatic story about assertion, which has two components:

> **Standard Pragmatics:** When the question under discussion is Q, you should give an answer that (1) you are certain of, and (2) is a *partial answer* to Q—a union of cells of Q that rules out at least one such cell (Grice, 1975; Stalnaker, 1978; Roberts, 2012).

In a slogan: *assert the strongest partial answer you’re sure of*. We have two arguments for Say Your Guess. The first is that *if* Standard Pragmatics is true, then Say Your Guess explains why it’s true. The second is that there is actually quite a bit of difficult data for part (1) of Standard Pragmatics—and if we take this data seriously, Say Your Guess offers a natural generalization of Standard Pragmatics.

Suppose that the first part is true: assertions require certainty, perhaps because asserting requires knowledge which, in turn, requires certainty. But why is the second part true? We could simply stipulate it, as most are happy to do. But we can do better, for Say Your Guess *explains* it. If asserting is the special case of saying your best guess that you are certain of (i.e., guessing with a minimal J-value of 1), then assertions will have to be Fit—since, as we’ve seen, any permissible guess in our framework is Fit. Moreover, the Gricean Maxim of Quantity plausibly entails that you shouldn’t assert a contextual tautology.[28] But a guess that is Fit but that also rules out some worlds from the context set is just a partial answer! So Say Your Guess (together with the Maxim of Quantity) explains a fundamental feature of Standard Pragmatics—a feature that is typically just stipulated.

But *is* Standard Pragmatics true? We think part (2) is right—you should only assert partial answers to the QUD, and Say Your Guess explains that. But is part (1) correct? Should you only assert things you’re certain of?

Once you look for it, a variety of linguistic data tells against this; people often say things (flat-out) when they are clearly not certain of them. It’s not unusual to overhear exchanges like, ‘What’s going to happen in the primary?’, ‘It’ll be Biden or Bernie’; ‘Where’s Latif going

[27] And—perhaps surprisingly—it turns out that people are roughly equally inclined to retract answers of the form ‘I think p’ as those of the form ‘p’ when they learn that p is false (though they are much more inclined to judge the latter assertions to be false). See Phillips and Mandelkern (2019).

[28] We should also say your assertion cannot have a *part* which is a contextual tautology. This rules out assertions like like ‘Latif will go to Yale, and I have two hands’. See Mayr and Romoli 2016 for an overview of recent developments of local implementations of the Maxim of Quantity.

to go?', 'He'll end up at Yale'; 'Is teaching going to be in person in the fall?', 'No way—it's going to be online'; 'It looks like it's going to rain; will the concert be cancelled?', 'Nah, it'll happen'; 'Will Bernie win in South Carolina?', 'No way—that's Biden's state'; etc. These assertions are clearly not certain. What's more, there is reason to think that they are reports of guesses. First, it is fully natural (but not obligatory) to follow them up by remarking as such: 'It'll be Biden or Bernie. That's my guess, anyways.' Despite flags like this, it's felicitous to report the speaker has having *said p*, rather than having said that they think or guess that p ('Jim said it'll be Biden or Bernie.') Thus it looks like, at least in some contexts, assertions can be *Improbable*, like guesses. But—again, like guesses—not just anything goes. While 'Latif will end up at Yale' can be acceptable, #'Latif will go to Harvard—but it's more likely he'll go to Yale' is nonsense; likewise (given your credences), 'He'll go to Harvard or Stanford' is decidedly odd. Thus it seems that assertions must be Filtered. They must also satisfy Fit, of course—'Latif will go to Yale, and it is cold in London' is unacceptable. And (thus) they are obviously Question Sensitive. In short, the day-to-day practice of information transfer seems to include a lot of exchanges that are structurally just like guesses.

This is something everyone needs to explain. A natural explanation is that assertion just *is* a particular kind of guessing—it is making a guess, and proposing to use that guess to update the common ground because you (take yourself to) know it. This is one way of elaborating Say Your Guess, and it seems promising to us.

It requires that you can *know* your guesses even when they are improbable. But it seems that you can, for there is a notion of (weak) knowledge that, we think, goes naturally with a theory of weak belief. Recall the unconfident examinee: 'When did William the Conqueror land in England?'; 'Um... 1066?'; 'Correct'; 'I knew it!' (Radford, 1966). Here and elsewhere, when your guess turns out to be correct, it's natural to conclude that you knew the answer—for example, 'He went to Yale!'; 'I knew it' (see Holguín 2019 for a wide variety of cases along these lines). Of course, it's easy to move into a more demanding context where we are less likely to ascribe knowledge ('Do you really know that?'). But that is a standard observation, not a special problem for a theory of weak knowledge (DeRose, 1992; Lewis, 1996).

Why say that asserting involves knowledge at all (even the weak kind)? The key motivation for this is the infelicity of Moorean sentences like #'Biden will win, but I don't know if he will', and the contrast with 'I think Biden will win, but I don't know if he will'. So long as asserting requires (thinking that you have) knowledge—even if knowledge can be very easy to come by—we have straightforward explanations for this contrast. Things get trickier with sentences like #'Biden will win, but I'm not $\left\{ \substack{\text{sure} \\ \text{certain}} \right\}$ he will', but there are a variety of options for explaining their infelicity. One is to point out that people do seem to use 'sure' and 'certain' in a way that doesn't correspond to credence 1: 'Who will win?', 'Biden—I'm sure of it'; 'Where's Latif going to go?', 'Surely he won't go to NYU'; etc. Perhaps, then, 'sure' and 'certain' in natural language express markers for what is (or should be) taken to be common ground—like 'might' and 'must'—rather than having the mental state of credence 1. This would explain the infelicity of the sentences in question. Moreover, this pragmatic story is buttressed by the fact that there are *practical* Moore sentences (Mandelkern, 2019): #'You must close the door, but I'm not sure you will' is just as infelicitous as #'You'll close the door, but I'm not sure you will'—yet it's not at all plausible that the explanation of the former is that you can give a command only if you're sure it'll be fulfilled. Perhaps, as Mandelkern suggests, both commands and assertions involve a kind of *pretense* of authority that is pragmatically incompatible with expressions of doubt.

Summing up: Say Your Guess *explains* a key feature of Standard Pragmatics, namely, why assertions must be Fit—i.e. must not crosscut cells of the QUD—and thus, in concert with the Maxim of Quantity, why they must be partial answers. According to Say Your Guess, you can respond to a question by either merely guessing your answer, or—when you (think you) know your guess—by asserting it. As we've seen, there are ways to fill in the details of this story that include or omit the constraint that you must be certain of what you assert. Clearly more work is needed to decide between these approaches; we think there is an intriguing case to be made for the latter approach, but the choice between them is independent of our larger goal here. That goal is to establish that guessing plays an important role in the exchange of communication, and is intimately related to the practice of assertion.

4.3 Guess when you reason?

We turn finally from the role of guesses thinking and talking to its role in *reasoning.* In particular, we'll argue that our theory helps explain a puzzling finding in the psychological literature on reasoning known as the *conjunction fallacy.* Recall that this is the observation that people sometimes rank a conjunction as more probable than one of its conjuncts. We think that this is a particularly interesting application for two reasons. First, the patterns that we have tried to explain above involve intuitive patterns around guessing, belief, and assertion. All of those patterns could be tested experimentally, but—to our knowledge—have not yet been. We think it is significant, then, that our account can also explain surprising and intricate patterns of judgments that *have* been explored experimentally. Second, the conjunction fallacy is a cornerstone of a fairly standard case in psychology that humans are fundamentally not very good at reasoning under uncertainty.[29] Our account gives a different diagnosis of what is going on here, consistent with a different picture of human reasoning. This is philosophically interesting in its own right; and, we think, provides further support for the thesis that guessing in general—and something along the lines of our theory in particular—has an important role to play in the theory of human cognition.

To begin, consider the most famous case from the literature on the conjunction fallacy (Tversky and Kahneman, 1983). Subjects were first given the following vignette:

> Linda is 31 years old, single, outspoken and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations.

Subjects were then asked which of two alternatives was more probable:

- · Linda is a bank teller.
- · Linda is a bank teller and is active in the feminist movement.

85% of the subjects chose the second option over the first, contrary to the laws of probability.

[29] For classic and modern statements of this kind of picture, see Tversky and Kahneman (1974); Kahneman and Tversky (1982, 1996); Fine (2005); Ariely (2008); Kahneman (2011); Lewis (2016). For resistance to it, including the rise of the 'rational analysis' program in cognitive science, see Anderson (1990); Gigerenzer (1991); Gigerenzer and Goldstein (1996); Oaksford and Chater (1994, 1998); Kelly (2004, 2008); Oaksford and Chater (2007); Tenenbaum and Griffiths (2006); Hahn and Oaksford (2007); Tenenbaum et al. (2011); Griffiths et al. (2012); Hahn and Harris (2014); Hedden (2018); Miller and Sanjurjo (2018).

To see how our account of guessing might help explain this result, recall a central lesson of our discussion of guessing above: it is acceptable to guess a less likely answer when it is correspondingly more informative—if asked where Latif is likely to go, it can be acceptable to guess 'Yale' over 'Yale or Harvard'. In particular, when we measure informativity in terms of the proportion of cells to a salient question under discussion that are ruled out, a conjunction will be more informative than its conjuncts, when both conjuncts address the QUD—so although the conjunction may be less probable, it may nevertheless be a better guess. Our hypothesis is that this accuracy-informativity tradeoff can explain the conjunction fallacy:

> **The Answer-Value Account:** People commit the conjunction fallacy because they rank outcomes according to their expected answer-value, rather than in terms of how probable they are.

A key motivation for this account is that evaluating expected answer-value—that is, comparing potential guesses—is something we all do all the time. As we've seen: often when we ask people what they think is likely, we are asking for them to guess.

Let's see in more detail how this could work. Take the Linda case, where Tversky and Kahneman found that a large proportion of subjects rated 'Linda is a feminist bank teller' (FT) as more likely than 'Linda is a bank teller' (T). First, we assume that the vignette and the alternatives offered (T vs. FT) generate a QUD based on the relevant characteristics in play, namely, the four-cell QUD obtained from combining the questions of whether Linda is a feminist or not (F or $\overline{F}$), and whether she is a bank teller or not (T or $\overline{T}$): $Q = \{FT, F\overline{T}, \overline{F}T, \overline{F}\overline{T}\}$.[30] Then, given credences P and $J \geq 1$, the expected answer-values of 'feminist bank teller' and 'bank teller' are as follows:

$$E^J_Q(FT) = P(FT) \cdot (J^{3/4})$$
$$E^J_Q(T) = P(T) \cdot (J^{1/2})$$

Thus the expected answer-value of FT is greater than that of T iff:

$$P(FT) \cdot (J^{3/4}) > P(T) \cdot (J^{1/2})$$
$$\Leftrightarrow \quad \frac{P(FT)}{P(T)} > \frac{J^{1/2}}{J^{3/4}}$$
$$\Leftrightarrow \quad P(F|T) > \frac{1}{J^{1/4}}$$

When the value of informativity is minimal ($J = 1$), the latter fraction equals $\frac{1}{1^{1/4}} = 1$, and the expected answer-value of FT is never higher than that of T (since $P(F|T) \leq 1$). But as J grows, this fraction shrinks. Thus the expected answer-value of 'feminist bank teller' is higher than that of 'bank teller' iff the conditional probability of Linda being a feminist, given that she's a bank teller, is sufficiently high—where what counts as 'sufficient' is determined by the Jamesian value of informativity J.

[30]Importantly, our theory's predictions don't depend on the details of the QUD selected: one that draws more distinctions about what Linda is like would yield the same (expected) answer-value scores for T, FT, etc., provided those distinctions are all (contextually) orthogonal to F and T.

Why does the conditional probability $P(F|T)$ matter, on our account? Because although the conjunction FT always has a lower probability than the conjunct T, the *degree* to which it's lower is determined by how likely F is given T, since $P(FT) = P(T) \cdot P(F|T)$. Thus when the conditional probability $P(F|T)$ is high, FT will be only slightly less likely than T—which speaks in favor of trading the (slightly) more probable but less informative guess, T, for the (slightly) less probable but more informative one, FT.

To make this more concrete, suppose (as seems reasonable to us) that you judge the probability that Linda is a feminist to be 0.8, and to be independent of whether she's a bank teller, so that $P(F|T) = P(F) = 0.8$. Then you should guess 'feminist bank teller' over 'bank teller' iff $P(F|T) = 0.8 > \frac{1}{J^{1/4}}$, which in turn holds iff $J > 2.44$, i.e. iff the value of a (true) maximally informative answer is a bit more than twice that of a (true) completely uninformative answer. While we don't expect to have direct intuitions about J-values, recall that in our original law-school case, 'Yale' was the best guess iff $J > 10.25$—so the value of informativity required to yield the conjunction fallacy is quite modest.[31]

To be clear: we're not claiming that subjects have the judgments they do in conjunction fallacy cases because they are actually *guessing*. In fact, while the guess FT plausibly has higher expected answer-value than T, neither guess is filtered given intuitive probability assignments—'feminist and *not* a bank teller' ($F\overline{T}$) is more probable than 'feminist bank teller' (FT), for example. So our claim is not that FT is a *good guess*, but rather that it is a *better* guess than T in many cases.

This brings out something interesting about the present application of our theory. We motivated our theory of guessing mainly via binary observations about which guesses—and beliefs and assertions—are *permissible*. To account for these judgments, we gave a theory which actually gives us something more general: *rankings* of all possible answers, according to their expected answer-value. While this may have at first seemed over-committal, we hope to have shown how these rankings naturally arise out of an intuitive picture (§3)—and now we are claiming they can be put to important work.

With the basic idea of our theory of the conjunction fallacy in hand, we want to briefly highlight some key predictions of our account. The conjunction fallacy is an intricate phenomenon which has generated an enormous amount of literature; for reasons of space, our discussion here must be limited. In future work, we plan to offer a more detailed examination of both the empirical literature and the ways our theory stacks up against other approaches. Here, we'll simply gesture at what we take to be the key selling points of our account.

First:

Prediction 1: Ranking AB over B will be more common as $P(A|B)$ goes up.

[31] One strategy for making concrete predictions here is to use the distinctiveness measure from §3.3 to predict how people will set J-values. We have some hesitancy about this, since that measure is best motivated as a way of seeing what the salient answers are, yet a conjunction fallacy case is one in which, by design, the options for answers are artificially restricted (e.g. 'feminist non-bank-teller' is not an option). But setting this hesitancy aside, here's how the approach would go. Suppose for illustration the probability of Linda being a bank teller is 0.1. Then the probabilities of the various complete answers are: $F\overline{T}$: 0.72; $\overline{F}\,\overline{T}$: 0.18; FT: 0.08; $\overline{F}T$: 0.02. This makes the distinctiveness of 'feminist non-teller' $\frac{0.72}{0.18} = 4$, while the next highest value is 1.071; thus we predict that there'll be a strong preference for setting J to a value that makes $F\overline{T}$ substantially higher than its alternatives. $F\overline{T}$ becomes more distinctive as J grows, thus predicting that people will have a high J-value and thus will likely commit the conjunction fallacy.

We have already seen why this follows on our view: when the conditional probability, say, $P(F|T)$ is high, then FT will be only slightly less likely than T, which speaks in favor of trading the (slightly) more probable but less informative guess T for the (slightly) less probable but more informative one FT.

This intuitive prediction is confirmed by a variety of empirical studies (e.g. Gavanski and Roskos-Ewoldsen 1991; Fantino et al. 1997; Costello 2009a,b and especially Tentori and Crupi 2012; though see Tentori et al. 2013 for a challenge). For instance, Tentori and Crupi (2012) asked subjects about two claims about a character Mark and a 100-ticket lottery: 'Mark is a scientist' (S) or 'Mark is a scientist and will win the lottery' (WS). Subjects were given different information about how many lottery tickets Mark has (either none, 1, 20, 50, 80, or all). The rates of conjunction fallacy (i.e. the rate of ranking 'Mark is a scientist and will win the lottery', WS, as more probable than 'Mark is a scientist', S) increased strictly with the number of lottery tickets Mark had—see Figure 3—which means, in turn, that they increased strictly with the conditional probability of 'Mark will win the lottery' on 'Mark is a scientist' (since these are probabilistically independent, $P(W|S) = P(W)$).

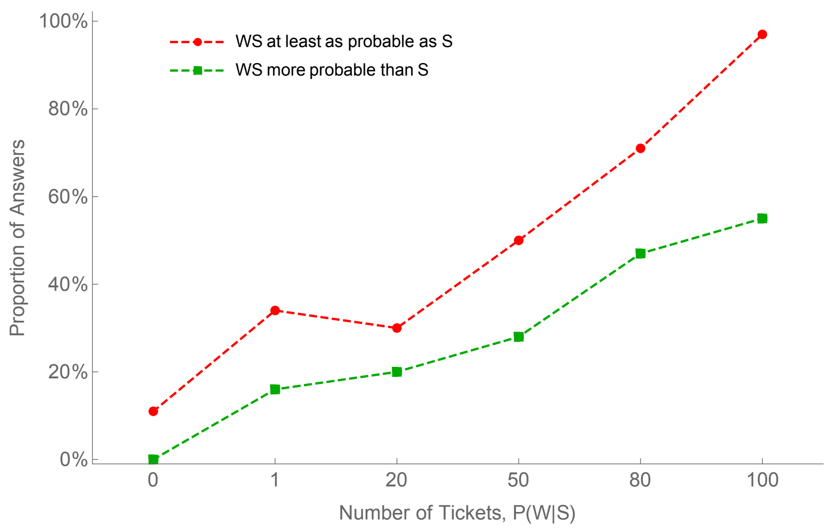

Figure 3: Conjunction-fallacy rates, varying $P(A|B)$ (Tentori and Crupi, 2012)

> **Prediction 2:** Ranking AB over B will not generally depend on the content of A and B, but instead on their (conditional) probabilities.

This prediction follows because the answer-value account of the conjunction fallacy is only sensitive to (i) how much more informative AB is than B, and (ii) how much less probable it is. Thus it is not generally sensitive to what A and B are about.[32] Perhaps surprisingly, this is empirically confirmed: the conjunction fallacy happens for unrelated conjunctions of events.

[32] We include the 'generally' rider because sometimes it will be near impossible for one of the conjuncts to play a role in a plausible QUD. In comparing 'feminist bank teller' to 'bank teller', a sensible overall question is 'What are Linda's social and political positions?' But if we ask you to compare 'Linda is a bank teller' to 'Linda is a bank teller and has at least seven eyelashes', you'll be hard-pressed to come up with a sensible overall question in which the second conjunct could play a part. We suspect this observation may help account for the data found in Tentori et al. (2013).

For example, Yates and Carlson (1986) found that with two probable but completely unrelated events—namely, 'Governor Blanchard will succeed in raising the Michigan state income tax' and 'Bo Derek will win an Academy Award for the movie she is currently making'—people committed the conjunction fallacy 56% of the time. Similarly for Costello (2009a) with unrelated weather events.

> **Prediction 3:** When $P(A|B)$ and $P(B|A)$ are *both* high, 'double'-conjunction fallacies will be common: people will rank $AB \succ A, B$. Meanwhile, when $P(A|B)$ is high but $P(B|A)$ is low, 'single'-conjunction fallacies will be common: $A \succ AB \succ B$.

These predictions follow because our account is symmetric: the expected answer-value of AB is higher than that of B iff $P(A|B) > \frac{1}{J^{1/4}}$; and it is higher than that of A iff $P(B|A) > \frac{1}{J^{1/4}}$.[33] And indeed, in cases where both conditional probabilities are high, people standardly rank the conjunction as more probable than each conjunct (Crupi et al., 2018), as in this famous case from Tversky and Kahneman (1983):

> A young college runner, Peter, has already run the mile in 4:06. Please rank the following for probability:
>
> (a) Peter will complete the mile under 4 min.
> (b) Peter will run the second half-mile under 1:55 and will complete the mile under 4 min.
> (c) Peter will run the second half-mile under 1:55

48% of subjects ranked (b) as the most probable of (a)–(c). We predict this because $P(\text{a}|\text{c})$ and $P(\text{c}|\text{a})$ are both relatively high.

> **Prediction 4:** Ranking AB over B will still occur regardless of how exactly the conjunction AB and conjunct B are phrased.

In particular, nothing about our account requires that the relevant expressions are literally conjunctions and conjuncts—what matters is simply their informativity and probability. As a result, different ways of expressing pairs of claims, one of which is more informative (but less probable than) the other, will result in the same effect. Thus the account predicts that the effect can occur when one of the claims is a disjunct and the other is a disjunction, or when one is a broad category and the other is a narrow one ('humanities' vs. 'literature'), etc. This is in line with findings in Bar-Hillel and Neter (1993); Costello (2009a). Moreover, the account predicts the effect to occur even when the claims are carefully phrased to avoid various implicatures, e.g. the reading that 'Linda is a bank teller' implicates that she's *not* a feminist. This is empirically confirmed.[34]

> **Prediction 5:** AB will often be ranked over B regardless of whether any evidence relevant to A or B is provided.

[33] In the Linda case people standardly rank rank $F \succ FT \succ T$ (Tversky and Kahneman, 1983). This is in line with our predictions, for in this case $P(F|T)$ is high (so likely above the threshold $\frac{1}{J^{1/4}}$), while $P(T|F)$ is low (so likely below it).

[34] E.g. Tversky and Kahneman (1983); Adler (1984); Agnoli and Krantz (1989); Macdonald and Gilhooly (1990); Politzer and Noveck (1991); Dulany and Hilton (1991); Gigerenzer (1991); Messer and Griggs (1993); Hertwig and Gigerenzer (1999); Mosconi and Macchi (2001); Tentori et al. (2004); Hertwig et al. (2008). Moro (2009) gives a helpful overview.

That is, we predict that whether AB will be ranked over B depends on how high $P(A|B)$ is, which in turn may be so regardless of whether subjects have received any confirming evidence for A (or B) as part of the experimental setup. Such evidence *is* part of the experimental setup in the Linda case—the vignette provides evidence that she is a feminist—and that has motivated 'confirmation-theoretic' accounts of the conjunction fallacy, based on the idea that people may say 'feminist bank teller' is more likely because it's *more confirmed* than 'bank teller' by the relevant evidence. But the conjunction fallacy is also observed in scenarios wherein subjects are provided with no relevant evidence by the experimenters, meaning neither answer is confirmed (Tversky and Kahneman, 1983; Yates and Carlson, 1986; Costello, 2009a). For example, Tversky and Kahneman (1983) asked some subjects to evaluate the probability of (a), and others to evaluate the probability of (b):

(a) There will be a massive flood somewhere in North America in 1983, in which more than 1000 people drown.

(b) There will be an earthquake in California sometime in 1983, causing a flood in which more than 1000 people drown.

The estimates for (b) were higher than for (a). Our account (unlike confirmation-theoretic accounts) generalizes immediately to this version of the conjunction fallacy, since (b) is more informative than (a) relative to a salient QUD. [35]

> **Prediction 6:** Since informativity relative to the QUD drives the effect, we expect that corresponding effects will diminish in cases involving estimates.

For example, tell subjects that 100 individuals fit Linda's description, and ask them to estimate the proportion of them that are ___s, where the blank is filled in by 'bank teller' or 'feminist bank teller.' Here the QUD is 'What proportion of people have property ___?', so regardless of what fills in the blank, each answer is equally informative in our sense (saying that more people are FT than T is not more informative). This prediction—that conjunction-fallacy rates diminish in estimate settings like these—is empirically confirmed (Tversky and Kahneman, 1983; Gigerenzer, 1991; Costello, 2009a; Moro, 2009).

In sum: our account makes a variety of subtle predictions which generally match existing findings in the literature. What about new predictions? To our knowledge, all the constraints on guessing brought out in §2 and §3.3 are new to the conjunction fallacy literature. As such, the answer-value account predicts that given standard conjunction-fallacy stimuli, people will tend to give answers that fit these constraints. Moreover, insofar as what matters for our account is an accuracy-informativity tradeoff, we predict that strategies for emphasizing the

[35] Confirmation accounts were to our knowledge first proposed in Sides et al. (2002) but developed most extensively in Crupi et al. (2008); Tentori and Crupi (2012); Tentori et al. (2013); Crupi et al. (2018). An additional advantage of our account over the confirmation one is that it avoids a version of the 'problem of irrelevant conjunctions' for confirmation theory: namely, if B is independent of E given A, then E confirms A iff E confirms AB. (Suppose $P(E|AB) = P(E|A)$. Then $P(A|E) > P(A)$ iff $P(E|A) > P(E)$ iff $P(E|AB) > P(E)$ iff $P(AB|E) > P(AB)$.) So, for instance, the confirmation account seems to predict that subjects will rate 'Linda is a feminist bank teller born on July 30, 1992, with a birthmark on her left toe' as more likely than 'Linda is a bank teller', since the former is confirmed by the vignette while the latter is not. This strikes us as counterintuitive, though it should be investigated experimentally. Tentori et al. (2013) argue in favor of confirmation accounts over accounts like ours which predict rates of conjunction fallacy to pattern with conditional probabilities; we hope to explore their arguments in future work (see footnote 32).

importance of accuracy (and de-emphasizing that of informativity) may mitigate occurrence of the conjunction fallacy.[36] We have not yet tested these predictions, but we think they are plausible and offer an interesting route for future research.

Having explained the basic contours of the Answer-Value Account of the conjunction fallacy, we want to briefly reflect more broadly on the account. The general idea that an accuracy-informative tradeoff is behind the conjunction fallacy has been discussed before, most prominently by Tversky and Kahneman (1983).[37] Their discussion is brief, concluding that 'it is unlikely that our respondents interpret the request to rank statements by their probability as a request to rank them by their expected (informational) value' (312). The worry seems to be that an account like ours is simply undermotivated (compare Moro, 2009, 18–19): *why* would people who are asked about probabilities respond using an accuracy-informativity tradeoff?

We think that the work we've done in this paper helps put this worry to rest. We've argued that assessing expected answer-value is a cognitively basic practice that plays a central role in guessing, believing, and talking. If this is right, assessing expected answer-value is a natural *default* mode of evaluating potential answers to questions. In other words, we think our discussion changes the dialectic: rather than introducing a new apparatus to explain the conjunction fallacy, we are showing how it arises naturally from a mechanism that arguably plays a central role in our cognitive lives.

Of course, there are undoubtedly many dynamics behind the conjunction fallacy; we don't claim that the Answer-Value Account is the whole story. In particular, we leave it open that other proposed factors—like confirmation, similarity, implicature, and noise—also influence the effect; but we maintain that assessments of expected answer-value play a central role in explaining the core phenomenon.

Zooming out a bit more: we are talking about the conjunction fallacy not just because it's a promising application, but because our account of it opens up philosophically interesting avenues. The conjunction fallacy is often held up as part of the core evidence that humans are fundamentally quite bad at dealing with uncertainty (Kahneman and Tversky, 1996; Kahneman, 2011). While our account is of course consistent with this picture, it's also consistent with a very different one.

Here's what we mean. Everyone should agree that people are bad at *conscious* probabilistic reasoning. This is demonstrated by the conjunction fallacy, and is made especially clear when people choose to *bet* on conjunctions rather than their conjuncts—as they sometimes do (Tversky and Kahneman, 1983; Bar-Hillel and Neter, 1993; Sides et al., 2002; Bonini et al., 2004). What should we conclude from this fact? In particular, does it show that the way people form judgments under uncertainty is fundamentally non-probabilistic—that they make do with other (worse) ways of managing uncertainty? Though popular[38], that conclusion sits

[36] Though how exactly to manipulate the 'stakes' of being accurate effectively is itself a tricky empirical question; see e.g. Buckwalter and Schaffer (2015).

[37] Again, a framework similar to ours was developed by Levi (1967), who suggested it could be applied to the conjunction fallacy—but we don't think the details work (see footnote 11). The idea that we trade off accuracy and informativity is also present in Yaniv and Foster (1995), but they do not give a general framework for evaluating informativity. Similar ideas are taken up Adler (1984) and Cross (2010); and Moro (2009) gives a helpful discussion of the idea. For the general idea that question sensitivity plays a central role in human reasoning, spelled out in a different framework, see Koralus and Mascarenhas (2013, 2018).

[38] See, for example, Kahneman et al. (1982); Tversky and Kahneman (1983); Gigerenzer and Goldstein (1996); Kahneman and Frederick (2002); Hastie and Dawes (2009); Kahneman (2011); Thaler (2015); Tetlock

uneasily with the burgeoning literature on Bayesian cognitive science, which uses probabilistic models to help explain the remarkable feats of human learning and inference (e.g. Anderson, 1990; Gopnik, 1996; Tenenbaum et al., 2011; Lake et al., 2016). That literature suggests that people are very *good* are (implicit) probabilistic reasoning.

Our account of the conjunction fallacy may help to reconcile these pictures. On our view, conscious "probabilistic" reasoning is in fact governed by the outputs of calculating expected answer-value, which in turn is governed by (implicit) probabilistic calculations. Even if those implicit calculations are broadly Bayesian—as our account allows, and popular heuristics-and-biases accounts deny—people's conscious probabilistic reasoning will be poor because expected answer-value doesn't conform to the rules of probability, and its the outputs of expected answer-value assessments that people have access to. In other words: people may be bad at conscious probabilistic reasoning because they are bad at *pulling apart* judgments about pure probability from judgments about expected answer-value.

Of course, probability is a *component* of expected answer-value. So should it be surprising, on our view, that people are good at assessing the latter and bad at assessing the former?[39] We don't think so—for two reasons. First, the *language* we use to talk about probabilities is very close (in fact, often identical) to the language we use to elicit guesses. Questions like 'What's most likely?', or 'What would you bet will happen?', are naturally used to elicit guesses, not probability judgments. Because of this, we think most people simply don't have much practice distinguishing these two types of judgments. [40] This may help make sense of why they sometimes have stubborn responses to criticisms of their answers when they commit the conjunction fallacy: they feel like they are being tricked by the question. For example, Michael Lewis recounts the following illustrative anecdote. Kahneman tried to convince a group of his students that the conjunction fallacy was an error, asking: 'Do you realize you have violated a fundamental rule of logic?' Lewis recounts: "So what!' a young woman shouted from the back of the room. 'You just asked for my opinion!" (Lewis, 2016, 325). If opinions are beliefs, and hence guesses, then this woman is protesting in exactly the same way we would expect—she thought she was being asked for her guess.

Second, and more generally, it's normal for components of core cognitive competences to be required for—but hard to separate from—those competences. You can effortlessly recognize a face, but would struggle to articulate any of its distinctive features. You can easily press the brake just hard enough to avoid a collision, but would be at a loss to articulate the rates

and Gardner (2016).

[39]Thanks to Josh Knobe for helpful discussion on this point.

[40]A question we want to remain neutral on here: do words like 'likely' and 'probably' have a *meaning* according to which they mean 'has high expected answer value'? This is not by any means outlandish; for instance Yalcin 2010 argues on the basis of cases like those we have focused on (following Windschitl and Wells 1998) that 'probably' is assessed relative to a salient QUD. He leaves open the exact form of QUD-sensitivity; one could naturally incorporate our account of guessing into a story about the meaning of 'probably'. This would lay the foundation for a very hard line on the rationality of the conjunction fallacy: if 'probably'/'likely' literally have a *meaning* on which they are measures of expected answer value rather than probability, then there is no mistake at all in conjunction fallacy judgments in response to questions about what is probable/likely. A softer line would say that the literal meaning of these words is about probability, but for reasons of pragmatics, questions about probability are naturally (mis/over)interpreted as questions about expected answer value. This softer line is still consistent with a broadly rationalist line on the conjunction fallacy (compare implicatures: one might deny that 'John had some cookies' literally *means* that he didn't have all of them, without thus thinking that subjects make an *error* if they conclude that he didn't). The choice between these approaches involves interesting methodological issues and deserves careful exploration.

of change of your car's speed and the distance between it and the one ahead, or why those rates will reach zero at the same point. Likewise: you can seamlessly respond to 'Where do you think Latif will go?' with 'Yale or Harvard', without having any conscious access to the calculations that went into this—first, of what question was being asked, and second, of how you chose among the options by weighing up probabilities and informativity. Thus, on our picture, assessing probabilities is a crucial step in assessing something we assess all the time (the quality of a guess), but one that can be consciously separated from it only with much practice (Tetlock and Gardner, 2016).

In sum, we think the existence of the conjunction fallacy supports the broader argument of this paper: that guessing is a core cognitive competence that plays a central role in how we think, talk, and reason about an uncertain world.

5 Why We Guess

We've covered a lot of ground. What guesses do people make? The answer is subtle but surprisingly systematic (§2; §3.3). How do people make guesses? By optimizing a tradeoff between accuracy and informativity (§3). When do people make guesses? All the time: they make guesses whenever they form beliefs, (§4.1), answer questions, (§4.2), or reason under uncertainty (§4.3).

But *why*? We've argued that guessing aims at both accuracy and informativity. We've given this hypothesis a simple exposition, and argued that it helps to explain a variety of patterns. We think the abductive case for it is strong.

Yet it may—and probably should—retain an air of mystery. When asked about guessing, the first things that come to mind are quiz shows and country fairs: 'Guess the exact number of jellybeans in this jar'; 'Try to guess which cup I hid the prize under—you get three tries'; etc. These guessing games have their own idiosyncratic rules: often only maximally informative guesses are allowed; sometimes multiple guesses are permitted; etc. Considering examples like this, it may seem that the practice of guessing itself will have no intrinsic rules or standards. Yet we've argued at length that it *does*: that the cognitive attitude of guessing the answer to a question—of figuring out what you think—always aims at accuracy, and is sensitive to informativity in a systematic way. If we are correct that the practice of guessing plays a central role in our cognitive lives, there must be some explanation of why it involves *these* rules. What is that explanation?

Here we can only speculate. Start with a more general question: why would you want to form a guess at all? If you already have credences—which, after all, are an *input* to our theory of guessing—why not just use them in making your way through the world? A natural answer, in the spirit of theories of bounded rationality, is that your credences can at best be extremely partial given your limited computational powers and the intractability of general probabilistic inference.[41] Thus you can only form credences over a relatively small set of propositions—for instance, those generated by the question under discussion. Those credences can be used to form a guess, which highlights a particular region of possibilities—the region that you take most seriously, given the question at issue and your contextual priorities.

What, then, do you do with this region? Our proposal is that you use it to guide further

[41] E.g. Simon (1957); Cherniak (1986); Dagum and Luby (1993); Weirich (2004); Bradley (2017).

investments of cognitive resources. You *reason within your guess*: supposing it to be true, you can form (and discuss) plans, preferences, or opinions about what to do, want, or think, if so. Having a small region like this can greatly simplify thought and talk about these activities. If we ask, 'Where do you think Latif will go?' and you reply, 'I think he'll go to Yale or Harvard', then it's incredibly natural to follow up with plans ('So we'll be able to visit him on the weekends'), preferences ('So I'll want to put him in touch with Jane'), or more fine-grained credences ('So it's likely he'll need a car'). You can do all of this without losing sight of the fact that the region is just your guess—possibilities inconsistent with your guess remain in your peripheral vision, so to speak, and are what prevent you from betting the farm on your guess the moment you make it. But by being able to highlight certain regions of possibilities for further thought and talk, you can reach a more fine-grained assessment of the various routes to action *in the scenario you guess will occur*.[42]

If something like this were right, it'd make sense of why guessing has the profile we've argued it does. Guesses should be accurate, since if your guess turns out to be false, any contingency-planning you've done within it will be wasted. But guesses should also be informative: in choosing a region of possibilities to highlight for further investigation, it pays to have a region that offers a specific answer to the live question since this cuts down on the number of distinctions you need to account for. An informative guess allows you to make fine-tuned plans even when you don't have the resources to plan for every contingency. For example: if you guess that Latif will go to Yale, you can focus on apartment listings in New Haven; if you guess that he'll go to Yale or Harvard, you can at least look at flight prices to the Northeast; but if you guess that he'll go to Yale, Harvard, or Stanford, you're plans within this guess can't be nearly as specific. Thus the 'reason within your guess' picture may have the resources to explain why guesses are subject to an accuracy-informativity tradeoff of the kind we've spelled out.

Moreover, there's empirical evidence that people *do* tend to reason within their guesses. First, a common claim in the literature on confirmation bias in psychology—and on theory-choice in philosophy of science—is that people have a tendency latch onto a specific, favored hypothesis, and expend most of their cognitive effort using its predictions to guide their investigations.[43] Second, a common mistake in poker is to 'put someone on a hand'—guess what they have, and use that guess to guide your betting. The mistake is not in guessing per se, but in having a guess that's overly specific (and hence improbable); *good* poker players put their opponents on a *range* of hands.[44] Third, studies of doctors' reasoning shows that they tend to commit the conjunction fallacy—that is, we think, to *guess*—when proposing diagnoses, which presumably has a direct impact on which procedures they go on to perform (Tversky and Kahneman, 1983; Rao, 2009; Crupi et al., 2018). Similarly, as mentioned above, people have some tendency to commit the conjunction fallacy when selecting bets—indicating that they're using their guess to frame and guide their actions.

As these examples illustrate, reasoning within guesses can lead to mistakes. But, again,

[42]This picture is related to (but, we think, interestingly distinct from) views on which subjects choose a set of possibilities to treat *as certain* in a given context—for discussion, see Harsanyi (1985); Lance (1995); Lin and Kelly (2012); Lin (2013); Clarke (2013); Tang (2015); Greco (2015); Leitgeb (2017), and Staffel (2019). We suspect it may interact in interesting ways with the "inquisitive decision theory" of Hoek (2020b).

[43]See, for example, Wason (1960); Kuhn (1962); Koriat et al. (1980); Klayman and Ha (1987); Kuhn (1989); Maher (1993); Gopnik (1996); Nickerson (1998).

[44]Thanks to Ben Holguín for the example.

we think situating them within a broader, bounded-rationality theory of guessing suggests that they are very different *kinds* of mistakes than is standardly thought. For the same mechanism that leads the human mind into conjunction-fallacy betting also—perhaps—helps it consistently outperform computers in novel situations of intractable complexity (Tenenbaum et al., 2011; Huang and Luo, 2015; Lake et al., 2016). There are over 10,000 known human diseases, over 2.5 million different poker hands, and always infinitely many empirically adequate scientific theories. Yet doctors, poker players, and scientists don't simply freeze up; instead, they make guesses that allow them to reason within such complexity.

We can illustrate this point close to home, for we've reached the stage where we speculate about future directions for our theory—that is, the stage where we guess. Obviously we can't formulate detailed opinions or plans about all the directions a theory like this could go. What we *can* do is highlight a region of possibilities for future investigation. That region should be small enough that we can see—albeit dimly—how such investigations might proceed; but it should be large enough that it is likely to contain promising directions. That is: it should be a good guess.

Our guess, then, is that this approach could be usefully applied to a range of topics in both cognitive science and philosophy.

In cognitive science: expected answer-value may help to explain other peculiar patterns in human judgments, like sub-additivity effects.[45] The way people generate and then reason within guesses may help to explain or refine the data surrounding confirmation bias—such as the infamous Wason selection task (Wason, 1966; Klayman and Ha, 1987; Nickerson, 1998). And, of course, our predictions about the conjunction fallacy may lead to new discoveries about it and related phenomena.

In philosophy: the accuracy-informativity tradeoff of guessing may help refine our theories of both conversational implicature (Grice, 1975) and prediction (Ninan, 2019; Cariani, 2020). The connection between guessing, (weak) belief, and action suggests that previous authors may have been too quick to treat weak belief as a linguistic phenomenon that plays no role in decision theory.[46] And our account epistemic value and its relation to new constraints on guesses and beliefs may open up new territory for the tools of epistemic utility theory (Joyce, 1998; Pettigrew, 2016a,b; Horowitz, 2018; Schoenfield, 2019b)—for instance, although we've focused on what make for good *guesses*, our account may also shed light on what make for good *questions* (Carballo, 2018).

Finally, at the intersection of philosophy and cognitive science: the accuracy-informativity tradeoff of guessing may contribute to both the debate about human (ir)rationality, and to our understanding of the scope and purpose of question-sensitivity in our cognitive lives.

[45]Given a proposition q and partition Q of q, people report probability judgments $P(\cdot)$ such that $\sum_{x \in Q} P(x) > P(q)$ (see Tversky and Koehler 1994; Redelmeier et al. 1995; Rottenstreich and Tversky 1997). Notably, (Jamesian) expected answer-value is subadditive: for any partition Q of q, $\sum_{x \in Q} E^J_Q(x) \geq E^J_Q(q)$, with equality only if $Q = \{q\}$ or $J = 1$.

[46]E.g. Christensen (2004); Križ (2015); Dorst (2019); Friedman (2019); Moss (2019); Williamson (2020).

Appendix: A Proof of Optionality

Theorem (Optionality). If P is regular over a question Q with $|Q| = n$, then for any $1 \leq k \leq n$, there is some $J \geq 1$ such that any filtered k-cell answer maximizes $E^J_Q(\cdot)$.

Proof. Order the cells q_i of the QUD by probability, so that $P(q_1) \geq P(q_2) \geq ... \geq P(q_n)$, and let $Q^i := q_1 \cup ... \cup q_i$ be the union of the first i cells. The only guesses that can maximize expected answer-value, for any value of J, are the filtered ones. And if p is filtered, then there is a Q^i that it is equivalent to in expectations—in particular, if p is a union of k cells and p is filtered, then $P(p) = P(Q^k)$ and $Q_p = Q_{Q^k}$, so—recalling that $E^J_Q(r) := P(r) \cdot J^{Q_r}$—we have that for all J, $E^J_Q(p) = E^J_Q(Q^k)$. Thus it suffices to show that for any $1 \leq k \leq n$, there is a $J \geq 1$ such that $E^J_Q(Q^k) > E^J_Q(Q^i)$ for all $i \neq k$.

We begin with several observations about the probabilities and expected values of the Q^i.

Lemma 1. The pairwise conditional probabilities are ordered: $P(Q^1|Q^2) < P(Q^2|Q^3) < \cdots < P(Q^{n-1}|Q^n)$.

Proof. Take an arbitrary $1 \leq i \leq n-2$ and consider $P(Q^i)$, $P(Q^{i+1})$, and $P(Q^{i+2})$, relabelling them p_0, p_1, p_2 respectively. Note that $P(Q^i|Q^{i+1}) = \frac{P(Q^i)}{P(Q^{i+1})} = \frac{p_0}{p_1}$, and similarly $P(Q^{i+1}|Q^{i+2}) = \frac{p_1}{p_2}$. Thus to establish Lemma 1 it suffices to show that $\frac{p_0}{p_1} < \frac{p_1}{p_2}$.

Note that by construction, $0 < p_0 < p_1 < p_2$, and moreover that $p_1 - p_0 = P(Q^{i+1}) - P(Q^i) = P(q_{i+1})$ and similarly $p_2 - p_1 = P(q_{i+2})$. Since we know $P(q_{i+1}) \geq P(q_{i+2})$, it follows that $p_1 - p_0 \geq p_2 - p_1$. Thus we have that $2p_1 \geq p_0 + p_2$ and so $\frac{p_0+p_2}{2} \leq p_1$. Note that what we want to show is that $\frac{p_0}{p_1} < \frac{p_1}{p_2}$, which holds iff $p_0 p_2 < (p_1)^2$; by the above inequality it suffices to show that $p_0 p_2 < (\frac{p_0+p_2}{2})^2$. This holds iff

$$\begin{aligned} p_0 p_2 &< \frac{p_0^2 + 2p_0p_2 + p_2^2}{4} \\ \Leftrightarrow\ 4p_0p_2 &< p_0^2 + 2p_0p_2 + p_2^2 \\ \Leftrightarrow\ 0 &< p_0^2 - 2p_0p_2 + p_2^2 \\ \Leftrightarrow\ 0 &< (p_0 - p_2)^2 \end{aligned}$$

which of course is true. It follows that $P(Q^i|Q^{i+1}) = \frac{p_0}{p_1} < \frac{p_1}{p_2} = P(Q^{i+1}|Q^{i+2})$, and since i was arbitrary, Lemma 1 follows in turn. □

Lemma 2. For any $1 \leq i \leq n-1$: $E^J_Q(Q^i) \gtreqless E^J_Q(Q^{i+1})$ iff $P(Q^i|Q^{i+1}) \gtreqless \frac{1}{J^{1/n}}$.

Proof. Noting that the proportion of Q ruled out by Q^i is $\frac{n-i}{n}$, we have that $E^J_Q(Q^i) \gtreqless E^J_Q(Q^{i+1})$ iff

$$\begin{aligned} P(Q^i) \cdot J^{\frac{n-i}{n}} &\gtreqless P(Q^{i+1}) \cdot J^{\frac{n-(i+1)}{n}} \\ \Leftrightarrow \frac{P(Q^i)}{P(Q^{i+1})} &\gtreqless \frac{J^{\frac{n-i-1}{n}}}{J^{\frac{n-i}{n}}} \\ \Leftrightarrow P(Q^i|Q^{i+1}) &\gtreqless J^{\frac{n-i-1}{n} - \frac{n-i}{n}} = J^{\frac{n-i-1-n+i}{n}} = J^{\frac{-1}{n}} = \frac{1}{J^{1/n}} \end{aligned}$$

as desired. □

From here we establish that for any J, the expectations of the Q^i are 'single-peaked':

Lemma 3 (Single-Peaked Expectations). For any $1 < i < n$: if $E^J_Q(Q^{i-1}) > E^J_Q(Q^i)$, then $E^J_Q(Q^i) > E^J_Q(Q^{i+1})$; and if $E^J_Q(Q^i) < E^J_Q(Q^{i+1})$, then $E^J_Q(Q^{i-1}) < E^J_Q(Q^i)$.

Proof. Suppose $E_Q^J(Q^{i-1}) > E_Q^J(Q^i)$. By Lemma 2 this implies that $P(Q^{i-1}|Q^i) > \frac{1}{J^{1/n}}$. By Lemma 1, we know that $P(Q^i|Q^{i+1}) > P(Q^{i-1}|Q^i)$. Stringing these inequalities together yields:

$$P(Q^i|Q^{i+1}) \ > \ P(Q^{i-1}|Q^i) \ > \ \frac{1}{J^{1/n}}$$

But Lemma 2 again tells us that since $P(Q^i|Q^{i+1}) > \frac{1}{J^{1/n}}$, we have $E_Q^J(Q^i) > E_Q^J(Q^{i+1})$, as desired.

If $E_Q^J(Q^i) < E_Q^J(Q^{i+1})$, parallel reasoning establishes that $E_Q^J(Q^{i-1}) < E_Q^J(Q^i)$. □

We're now in a position to complete the proof of Optionality. Given an arbitrary k such that $1 \leq k \leq n$, we find a $J \geq 1$ for which $E_Q^J(Q^k)$ is maximal amongst the Q^i.

If $k = n$, then, by setting $J = 1$, Lemma 2 implies that $E_Q^J(Q^{n-1}) < E_Q^J(Q^n)$ iff $P(Q^{n-1}|Q^n) < \frac{1}{1^{1/n}} = 1$, which (by regularity) holds. Lemma 3 then implies that $E_Q^J(Q^1) < \cdots < E_Q^J(Q^n)$, as desired. Meanwhile, if $k = 1$, then sending $J \to \infty$ suffices, since this sends $\frac{1}{J^{1/n}} \to 0$, and once $P(Q^1|Q^2) > \frac{1}{J^{1/n}}$, Lemma 2 implies the $E_Q^J(Q^1) > E_Q^J(Q^2)$, and then Lemma 3 implies that $E_Q^J(Q^1) > \cdots > E_Q^J(Q^n)$, as desired.

Now consider the case when $1 < k < n$. Given the single-peaked expectations from Lemma 3, it suffices to show that there is a $J \geq 1$ such that $E_Q^J(Q^{k-1}) < E_Q^J(Q^k) > E_Q^J(Q^{k+1})$. By Lemma 2, this holds iff both $P(Q^{k-1}|Q^k) < \frac{1}{J^{1/n}}$ and $P(Q^k|Q^{k+1}) > \frac{1}{J^{1/n}}$. By Lemma 1 we know that there are $t, \epsilon > 0$ such that $P(Q^{k-1}|Q^k) = t < t+\epsilon = P(Q^k|Q^{k+1})$. Thus it suffices to show that there is a $J \geq 1$ such that $\frac{1}{J^{1/n}}$ is strictly between these thresholds—say, $\frac{1}{J^{1/n}} = t + \frac{\epsilon}{2}$. We know that $t + \frac{\epsilon}{2} \in (0, 1)$. Note that when $J = 1$, $\frac{1}{J^{1/n}} = 1 > t + \frac{\epsilon}{2}$; when $J \to \infty$, $\frac{1}{J^{1/n}} \to 0 < t + \frac{\epsilon}{2}$; and $\frac{1}{J^{1/n}}$ is continuous in J. Thus by the intermediate value theorem, there is some $\mathcal{J} > 1$ such that $\frac{1}{\mathcal{J}^{1/n}} = t + \frac{\epsilon}{2}$. This establishes that $E_Q^{\mathcal{J}}(Q^{k-1}) < E_Q^{\mathcal{J}}(Q^k) > E_Q^{\mathcal{J}}(Q^{k+1})$, completing the proof. □

References

Adler, J. E. (1984). Abstraction is uncooperative. *Journal for the Theory of Social Behaviour*.

Agnoli, F. and Krantz, D. (1989). Suppressing natural heuristic by formal instruction: The case of the conjunction fallacy. *Cognitive Psychology*, 21(515-550).

Anderson, J. R. (1990). *The Adaptive Character of Thought*. Erlbaum Associates.

Ariely, D. (2008). *Predictably irrational*. Harper Audio.

Bar-Hillel, M. and Neter, E. (1993). How alike is it versus how likely is it: A disjunction fallacy in probability judgments. *Journal of Personality and Social Psychology*, 65(6):1119.

Baron, J. (1985). *Rationality and intelligence*. Cambridge University Press, New York, NY, US.

Bonini, N., Tentori, K., and Osherson, D. (2004). A different conjunction fallacy. *Mind & Language*, 19(2):199–210.

Bradley, R. (2017). *Decision Theory with a Human Face*. Cambridge University Press.

Brown, J. (2008). The Knowledge Norm for Assertion. *Philosophical Issues*, 18(1):89–103.

Buchak, L. (2013). *Risk and Rationality*. Oxford University Press, Oxford.

Buckwalter, W. and Schaffer, J. (2015). Knowledge, stakes, and mistakes. *Nous*, 49(2):201–234.

Builes, D., Horowitz, S., and Schoenfield, M. (2020). Dilating and contracting arbitrarilly. *Noûs*.

Carballo, A. P. (2018). Good Questions. *Epistemic Consequentialism*, page 123.

Cariani, F. (2020). On predicting. *Ergo*.

Cherniak, C. (1986). *Minimal Rationality*. Bradford Books, Cambridge, MA.

Christensen, D. (2004). *Putting logic in its place: Formal constraints on rational belief*. Oxford University Press on Demand.

Clarke, R. (2013). Belief is Credence One (In Context). *Philosophers' Imprint*, 13(11).

Costello, F. J. (2009a). Fallacies in probability judgments for conjunctions and disjunctions of everyday events. *Journal of Behavioral Decision Making*, 22:235–251.

Costello, F. J. (2009b). How Probability Theory Explains the Conjunction Fallacy. *Journal of Behavioral Decision Making*, 22:213–234.

Cross, T. (2010). Skeptical success. *Oxford Studies in Epistemology*.

Crupi, V., Elia, F., Apra, F., and Tentori, K. (2018). Double conjunction fallacies in physicians' probability judgment. *Medical Decision Making*, 38(6):756–760.

Crupi, V., Fitelson, B., and Tentori, K. (2008). Probability, confirmation, and the conjunction fallacy. *Thinking & Reasoning*, 14(2):182–199.

Dagum, P. and Luby, M. (1993). Approximating probabilistic inference in Bayesian belief networks is NP-hard. *Artificial intelligence*, 60(1):141–153.

DeRose, K. (1992). Contextualism and knowledge attributions. *Philosophy and Phenomenological Research*, 52(4):913–29.

Dorst, K. (2014). Can the Knowledge Norm CoOpt the Opt Out? *Thought: A Journal of Philosophy*, 3(4):273–282.

Dorst, K. (2019). Lockeans Maximize Expected Accuracy. *Mind*, 128(509):175–211.

Drucker, D. (2020). The attitudes we can have. *The Philosophical Review*.

Dulany, D. E. and Hilton, D. J. (1991). Conversational implicature, conscious representation, and the conjunction fallacy. *Social Cognition*, 9(1):85–110.

Easwaran, K. (2016). Dr. Truthlove Or: How I Learned to Stop Worrying and Love Bayesian Probabilities. *Noûs*, 50(4):816–853.

Elgin, C. Z. (2017). *True enough*. MIT Press, Cambridge, MA.

Fantino, E., Kulik, J., Stolarz-fantino, S., and Wright, W. (1997). The conjunction fallacy: A test of averaging hypotheses. *Psychonomic Bulletin & Review*, 4(1):96–101.

Feinmann, D. (2020). *Lexical Illusions, (Non-)Maximality, and Invisible Gaps*. PhD thesis, University of Sheffield, Sheffield, UK.

Fine, C. (2005). *A Mind of its Own: How Your Brain Distorts and Deceives*. W. W. Norton & Company.

Foley, R. (1992). The Epistemology of Belief and the Epistemology of Degrees of Belief. *American Philosophical Quarterly*, 29(2):111–124.

Friedman, J. (2019). Inquiry and Belief. *Nous*, 53(2):296–315.

Gavanski, I. and Roskos-Ewoldsen, D. (1991). Representativeness and conjoint probability. *Journal of Personality and Social Psychology*, 61(2).

Gerken, M. and Beebe, J. R. (2016). Knowledge in and out of contrast. *Noûs*, 50(1):133–164.

Gigerenzer, G. (1991). How to make cognitive illusions disappear: Beyond "heuristics and biases". *European review of social psychology*, 2(1):83–115.

Gigerenzer, G. and Goldstein, D. G. (1996). Reasoning the fast and frugal way: models of bounded rationality. *Psychological review*, 103(4):650.

Gopnik, A. (1996). The Scientist as Child. *Philosophy of Science*, 63(December):485–514.

Greco, D. (2015). How I learned to stop worrying and love probability 1. *Philosophical Perspectives*, 29(1):179–201.

Grice, H. P. (1975). *Logic and Conversation*. Harvard.

Griffiths, T. L., Chater, N., Norris, D., and Pouget, A. (2012). How the bayesians got their beliefs (and what those beliefs actually are): Comment on bowers and davis (2012). *Psychological Bulletin*, 138(3):415–422.

Groenendijk, J. and Stokhof, M. (1984). *Studies on the Semantics of Questions and the Pragmatics of Answers*. PhD thesis, University of Amsterdam.

Hahn, U. and Harris, A. J. (2014). What does it mean to be biased: Motivated reasoning and rationality. In *Psychology of Learning and Motivation - Advances in Research and Theory*, volume 61, pages 41–102. Elsevier.

Hahn, U. and Oaksford, M. (2007). The rationality of informal argumentation: a Bayesian approach to reasoning fallacies. *Psychological review*, 114(3):704.

Hamblin, C. L. (1973). Questions in Montague English. *Foundations of Language*, 10:41–53.

Harsanyi, J. C. (1985). Acceptance of empirical statements: A Bayesian theory without cognitive utilities. *Theory and decision*, 18(1):1–30.

Hastie, R. and Dawes, R. M. (2009). *Rational choice in an uncertain world: The psychology of judgment and decision making*. Sage Publications.

Hawthorne, J., Rothschild, D., and Spectre, L. (2016). Belief is weak. *Philosophical Studies*, 173(5):1393–1404.

Hedden, B. (2018). Hindsight Bias is not a Bias. *Analysis*, To appear.

Hempel, C. G. (1962). Deductive-nomological vs. statistical explanation. *Minnesota Studies in the Philosophy of Science*, 3:98–169.

Hertwig, R., Benz, B., and Krauss, S. (2008). The conjunction fallacy and the many meanings of and.

Cognition, 108(3):740–753.
Hertwig, R. and Gigerenzer, G. (1999). The 'conjunction fallacy'revisited: How intelligent inferences look like reasoning errors. *Journal of behavioral decision making*, 12(4):275–305.
Hoek, D. (2020a). Minimal rationality and the web of questions. In Elswyk, P. V. and Egan, A., editors, *Unstructured Content*. Oxford University Press.
Hoek, D. (2020b). Questions in Action. Manuscript, Virginia Tech.
Holguín, B. (2019). Knowledge by constraint. Manuscript, NYU.
Holguín, B. (2020). Thinking, guessing, and believing. Manuscript, NYU.
Horowitz, S. (2017). Accuracy and Educated Guesses. In *Oxford Studies in Epistemology*. Oxford University Press.
Horowitz, S. (2018). Epistemic value and the Jamesian goals. *Epistemic Consequentialism*, pages 269–289.
Huang, Z. J. and Luo, L. (2015). It takes the world to understand the brain: International brain projects discuss how to coordinate efforts. *Science*, 350(6256):42–44.
James, W. (1956). The will to believe. In *The Will to Believe and Other Essays in Popular Philosophy*, pages 1–31. Dover, New York.
Joyce, J. (1998). A nonpragmatic vindication of probabilism. *Philosophy of Science*, 65(4):575–603.
Kahneman, D. (2011). *Thinking Fast and Slow*. Farrar, Straus, and Giroux.
Kahneman, D. and Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment. *Heuristics and biases: The psychology of intuitive judgment*, 49:81.
Kahneman, D., Slovic, P., and Tversky, A., editors (1982). *Judgment Under Uncertainty: Heuristics and Biases*. Cambridge University Press.
Kahneman, D. and Tversky, A. (1982). Variants of uncertainty. *Cognition*, 11:143–157.
Kahneman, D. and Tversky, A. (1996). On the reality of cognitive illusions. *Psychological Review*, 103(3):582–591.
Karttunen, L. (1977). Syntax and semantics of questions. *Linguistics and Philosophy*, 1:3–44.
Kelly, T. (2004). Sunk costs, rationality, and acting for the sake of the past. *Nous*, 38(1):60–85.
Kelly, T. (2008). Disagreement, Dogmatism, and Belief Polarization. *The Journal of Philosophy*, 105(10):611–633.
Klayman, J. and Ha, Y.-w. (1987). Confirmation, disconfirmation, and information in hypothesis testing. *Psychological Review*, 94(2):211–228.
Koralus, P. and Mascarenhas, S. (2013). The erotetic theory of reasoning: bridges between formal semantics and the psychology of deductive inference. *Philosophical Perspectives*, 27:312–365.
Koralus, P. and Mascarenhas, S. (2018). Illusory inferences in a question-based theory of reasoning. In *Pragmatics, Truth and Underspecification*, pages 300–322. Brill.
Koriat, A., Lichtenstein, S., and Fischhoff, B. (1980). Journal of Experimental Psychology : Human Learning and Memory Reasons for Confidence. *Journal of Experimental Psychology: Human Learning and Memory*, 6(2):107–118.
Križ, M. (2015). *Aspects of Homogeneity in the Semantics of Natural Language*. PhD thesis, University of Vienna.
Kuhn, D. (1989). Children and adults as intuitive scientists. *Psychological review*, 96(4):674.
Kuhn, T. (1962). *The Structure of Scientific Revolutions*. The University of Chicago Press.
Lackey, J. (2007). Norms of Assertion. *Noûs*, 41(4):594–626.
Lake, B. M., Ullman, T. D., Tenenbaum, J. B., and Gershman, S. J. (2016). Building Machines That Learn and Think Like People. *Behavioral and Brain Sciences*, 40:1–101.
Lance, M. N. (1995). Subjective probability and acceptance. *Philosophical Studies*, 77:147–179.
Leitgeb, H. (2014). The Stability Theory of Belief. *Philosophical Review*, 123(2):131–171.
Leitgeb, H. (2017). *The Stability of Belief. How Rational Belief Coheres with Probability*. Oxford University Press.
Levi, I. (1967). *Gambling with truth: An essay on induction and the aims of science*. MIT Press, Cambridge, MA.
Levi, I. (1985). Review: Illusions about Uncertainty. *The British Journal for the Philosophy of Science*, 36(3):331–340.
Levi, I. (2004). Jaakko Hintikka. *Synthese*, 140(1):37–41.
Lewis, D. (1996). Elusive knowledge. *Australasian Journal of Philosophy*, 74(4):549–567.
Lewis, M. (2016). *The undoing project: A friendship that changed the world*. Penguin UK.
Lin, H. (2013). Foundations of everyday practical reasoning. *Journal of Philosophical Logic*, 42(6):831–862.

Lin, H. and Kelly, K. (2012). Propositional reasoning that tracks probabilistic reasoning. *Journal of Philosophical Logic*, 41(6):957–981.

Macdonald, R. and Gilhooly, K. (1990). More about Linda or conjunctions in context. *The European Journal of Cognitive Psychology*, 2:57–70.

Maher, P. (1993). *Betting on Theories*. Cambridge University Press.

Mandelkern, M. (2019). Practical Moore sentences. *Noûs*.

Marr, D. (1982). *Vision*. MIT Press.

Mayr, C. and Romoli, J. (2016). A puzzle for theories of redundancy: Exhaustification, incrementality, and the notion of local context. *Semantics and Pragmatics*, 9(7):1–48.

Messer, W. and Griggs, R. (1993). Another look at Linda. *Bulletin of the Pyschonomic Society*, 31:193–196.

Miller, J. B. and Sanjurjo, A. (2018). Surprised by the Hot Hands Fallacy? A Truth in the Law of Small Numbers. *Econometrica*, 86(6):2019–2047.

Moro, R. (2009). On the nature of the conjunction fallacy. *Synthese*, 171(1):1–24.

Mosconi, G. and Macchi, L. (2001). The role of pragmatic rules in the conjunction fallacy. *Mind & Society*, 2(1):31–57.

Moss, S. (2019). Full belief and loose speech. *Philosophical Review*, 128(3):255–91.

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, 2(2):175–220.

Ninan, D. (2019). Easy foreknowledge. Manuscript, Tufts University.

Oaksford, M. and Chater, N. (1994). A Rational Analysis of the Selection Task as Optimal Data Selection. *Psychological Review*, 101(4):608–631.

Oaksford, M. and Chater, N. (1998). *Rational models of cognition*. Oxford University Press Oxford.

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

Oddie, G. (1997). Conditionalization, Cogency, and Cognitive Value. *The British Journal for the Philosophy of Science*, 48(4):533–541.

Oddie, G. (2019). What accuracy could not be. *The British Journal for the Philosophy of Science*, 70(2):551–580.

Pettigrew, R. (2016a). *Accuracy and the Laws of Credence*. Oxford University Press.

Pettigrew, R. (2016b). Jamesian Epistemology Formalized: An Explication of 'The Will to Believe'. *Episteme*, 13(3):253–268.

Phillips, J. and Mandelkern, M. (2019). Eavesdropping: What is it good for? Manuscript, Harvard and All Souls College.

Politzer, G. and Noveck, I. A. (1991). Are conjunction rule violations the result of conversational rule violations? *Journal of psycholinguistic research*, 20(2):83–103.

Popper, K. (1963). *Conjectures and Refutations*. Routledge and Kegan Paul, London.

Quiggin, J. (1982). A theory of anticipated utility. *Journal of Economic Behavior and Organization*, 3(4):323–343.

Radford, C. (1966). Knowledge: By Examples. *Analysis*, 27(1):1–11.

Rao, G. (2009). Probability error in diagnosis: the conjunction fallacy among beginning medical students. *Family Medicine*, 41:262–5.

Redelmeier, D. A., Koehler, D. J., Liberman, V., and Tversky, A. (1995). Probability judgment in medicine: Discounting unspecified alternatives. *Medical Decision Making*, 15:227–231.

Roberts, C. (2012). Information structure in discourse: Towards an integrated formal theory of pragmatics. *Semantics and Pragmatics*, 5(6):1–69.

Rothschild, D. (2019). What it takes to believe. *Philosophical Studies*.

Rottenstreich, Y. and Tversky, A. (1997). Unpacking, repacking, and anchoring: Advances in support theory. *Psychological Review*, 104:406–415.

Schaffer, J. (2005). Contrastive knowledge. In Gendler, T. S. and Hawthorne, J., editors, *Oxford Studies in Epistemology*. Oxford University Press.

Schaffer, J. (2007). Knowing the answer. *Philosophy and Phenomenological Research*, 75:383–403.

Schaffer, J. and Knobe, J. (2012). Contrastive knowledge surveyed. *Noûs*, 46(4):675–708.

Schaffer, J. and Szabó, Z. G. (2014). Epistemic comparativism: A contextualist semantics for knowledge ascriptions. *Philosophical Studies*, 168:491–543.

Schoenfield, M. (2019a). Accuracy and Verisimilitude: The Good, The Bad, and the Ugly. *The British Journal for the Philosophy of Science*, To Appear.

Schoenfield, M. (2019b). Permissivism and the value of rationality: A challenge to the uniqueness thesis. *Philosophy and phenomenological research*, 99(2):286–297.

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, 27(3):379–423.

Sides, A., Osherson, D., Bonini, N., and Viale, R. (2002). On the reality of the conjunction fallacy. *Memory & Cognition*, 30:191–198.

Simon, H. (1957). *Models of man, social and rational: Mathematical essays on rational human behavior in a social setting*. Wiley, New York.

Sliwa, P. and Horowitz, S. (2015). Respecting *all* the evidence. *Philosophical Studies*, 172(11):2835–2858.

Staffel, J. (2019). How do Beliefs Simplify Reasoning? *Noûs*, 53(4):937–962.

Stalnaker, R. (1974). Pragmatic presuppositions. In Munitz, M. K. and Unger, P., editors, *Semantics and Philosophy*, pages 197–213. New York University Press, New York.

Stalnaker, R. (1978). Assertion. In Cole, P., editor, *Syntax and semantics*, volume 9, pages 315–322. New York: Academic Press.

Sturgeon, S. (2008). Reason and the Grain of Belief. *Noûs*, 42(1):139–165.

Tang, W. H. (2015). The tale of bella and creda. *Philosophers' Imprint*, 15(31):1–9.

Tenenbaum, J. B. and Griffiths, T. L. (2006). Optimal Predictions in Everyday Cognition. *Psychological Science*, 17(9):767–773.

Tenenbaum, J. B., Kemp, C., Griffiths, T. L., and Goodman, N. D. (2011). How to grow a mind: Statistics, structure, and abstraction. *science*, 331(6022):1279–1285.

Tentori, K., Bonini, N., and Osherson, D. (2004). The conjunction fallacy: A misunderstanding about conjunction? *Cognitive Science*, 28(3):467–477.

Tentori, K. and Crupi, V. (2012). How the conjunction fallacy is tied to probabilistic confirmation: Some remarks on Schupbach (2009). *Synthese*, 184(1):3–12.

Tentori, K., Crupi, V., and Russo, S. (2013). On the determinants of the conjunction fallacy: Probability versus inductive confirmation. *Journal of Experimental Psychology: General*, 142(1):235.

Tetlock, P. E. and Gardner, D. (2016). *Superforecasting: The art and science of prediction*. Random House.

Thaler, R. H. (2015). *Misbehaving: The Making of Behavioural Economics*. Penguin.

Thomason, R. H. (1986). The context-sensitivity of belief and desire. In Georgeff, M. P. and Lansky, A. L., editors, *Reasoning about Actions and Plans*, pages 341–360. Morgan Kaufmann, Los Altos, CA.

Tversky, A. and Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185(4157):1124–1131.

Tversky, A. and Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review*, 90(4):293–315.

Tversky, A. and Koehler, D. J. (1994). Support theory: A nonextensional representation of subjective probability. *Psychological Review*, 101:547–567.

Wason, P. C. (1960). On the failure to eliminate hypotheses in a conceptual task. *Quarterly journal of experimental psychology*, 12(3):129–140.

Wason, P. C. (1966). Reasoning. In Foss, B., editor, *New Horizons in Psychology*, pages 135–151. Penguin Books.

Weirich, P. (2004). *Realistic decision theory: Rules for non-ideal agents in non-ideal circumstances*. Oxford University Press, Oxford.

Williamson, T. (2000). *Knowledge and its Limits*. Oxford University Press.

Williamson, T. (2020). Knowledge, Credence, and the Strength of Belief. In Flowerree, A. and Reed, B., editors, *Expansive Epistemology: Norms, Action, and the Social World*, page To appear. Routledge.

Windschitl, P. D. and Wells, G. L. (1998). The alternative-outcomes effect. *Journal of Personality and Social Psychology*, 75(6):1411–1423.

Yablo, S. (2014). *Aboutness*. Princeton University Press.

Yalcin, S. (2010). Probability operators. *Philosophy Compass*, 5(11):916–937.

Yalcin, S. (2011). Nonfactualism about epistemic modality. In Egan, A. and Weatherson, B., editors, *Epistemic Modality*, pages 295–332. Oxford University Press.

Yalcin, S. (2018). Belief as question-sensitive. *Philosophy and Phenomenological Research*, XCVII(1):23–47.

Yaniv, I. and Foster, D. P. (1995). Graininess of judgment under uncertainty: An accuracy-informativeness trade-off. *Journal of Experimental Psychology: General*, 124(4):424.

Yates, F. J. and Carlson, B. W. (1986). Conjunction errors: Evidence for multiple judgment procedures, including "signed summation". *Organizational Behavior and Human Decision Processes*, 37(2):230–253.