Better permissible guesses

Niels Linnemann*†1 and Feraz Azhar*‡2

[1]*Department of Philosophy, University of Geneva, Geneva, Switzerland*
[2]*Department of Philosophy, University of Notre Dame, Notre Dame, Indiana, U.S.A.*

This is a pre-peer reviewed version of an article that has been accepted for publication in *Philosophy and Phenomenological Research*.

Dated: February 13, 2024

It has recently become popular to analyze scenarios in which we *guess*, in terms of a trade-off between the accuracy of our guess (namely, its credence) and its specificity (namely, how many answers it rules out). Dorst and Mandelkern describe an account of guessing, based on epistemic utility theory (EUT), in which permissible guesses vary depending on how one weighs accuracy against specificity. We provide a minimal formal account of guessing that: (i) does not employ EUT, but rests on how such trade-offs are treated in the sciences; (ii) is relatively parsimonious; and (iii) is consistent with a variety of more specific models that describe what an agent is doing when they (rationally) guess. Our account also recovers patterns of guessing and predictions about typical outcomes of guessing, as identified by Dorst and Mandelkern. Furthermore, we focus on how permissible guesses can be improved upon, via changes in an agent's credence distribution. Such *better permissible guesses* can be generated in solving *Fermi problems*—guessing problems of a type that has received almost no attention in the philosophical literature—which we also analyze. Our account strengthens the case for understanding guessing (now, very broadly considered) in terms of accuracy-specificity trade-offs.

1 Introduction

How many penguins can you fit into your kitchen? Take a guess. It might not be entirely clear where to start, but upon further reflection you figure that you could: (i) guess (perhaps very well) the number of penguins you could fit into a shoebox; (ii) guess (perhaps fairly well) the number of shoeboxes you could fit into a large moving box; and (iii) guess (also fairly well) the number of large moving boxes you could fit into your kitchen. This would provide you with an approximate answer to the original question: an approximate overall guess that you would obtain by multiplying your three guesses in (i), (ii), and (iii). There are surprisingly many guessing problems that are like this. You might initially only have a vague sense of the possible answers and relatively unopinionated credences (more formally, perhaps a uniform distribution) over possible answers; but you can proceed by decomposing the problem into subproblems,

*The authors contributed equally to this paper.
†niels.linnemann@unige.ch
‡fazhar@nd.edu

providing guesses for the subproblems that are then combined to provide an overall guess. If, furthermore, the decomposition is a multiplicative one, that is, if the guesses to the subproblems ought to be multiplied to yield an overall guess, then the guessing problem is a *Fermi problem*. Many quantitative guessing problems are Fermi problems. And, interestingly, Fermi problems perhaps extend beyond explicit quantitative guessing problems: for, as described by Von Baeyer (1993, p. 6): "Fermi problems are reminiscent of the ordinary [problems we] encounter every day".[1]

The types of guessing scenarios analyzed in the recent philosophical literature do not involve Fermi problems. Instead, the standard guessing problem is one in which possible (complete) answers, together with (fixed) credences for those answers, are assumed. It is for such guessing problems that Dorst and Mandelkern (2022) [see also Skipper (2023), as well as Levi (1973) for a precursor of this work] characterize *permissible* guesses—using a specific model in the context of epistemic utility theory—in terms of a trade-off between two qualities of guesses. These qualities have been termed the 'accuracy' of a guess (namely, its credence)[2] and the 'specificity' of a guess (namely, how many complete answers it rules out).[3] Notably, the theory advanced by Dorst and Mandelkern (2022) recovers plausible patterns of guessing gleaned from the practice of guessing; and makes predictions about typical outcomes of guessing. In this paper, we present a different account (what we call a 'minimal formal account') of guessing understood in terms of a trade-off between accuracy and specificity; one that we deploy in the analysis of guessing problems that go beyond the standard guessing problem as described above—including an explication of Fermi problems. Our account (as well as its applications), we believe, significantly strengthens the case for understanding guessing (now, very broadly considered) as a trade-off between accuracy and specificity.

In Sec. 2, we present our minimal formal account of guessing. This account does not employ epistemic utility theory, is parsimonious (compared with accounts that do employ such a theory), and treats the trade-off in a manner that is consistent with how it is often treated in economics and in the sciences. Our account is plausibly consistent with a variety of more specific models (including those understood in the context of epistemic utility theory)—that describe what an agent is doing when they (rationally) guess. Our account recovers the patterns of guessing identified by Dorst and Mandelkern, as well as their predictions about typical outcomes of guessing. Furthermore, in Sec. 3, we focus on how permissible guesses can be improved upon, that is, on *better permissible guesses*—as a result of changes in an agent's credence distribution. We describe, in Sec. 3.1, how to characterize better permissible guesses in the context of our minimal formal account. And then, in Sec. 3.2, we describe how to generate better permissible guesses by analyzing Fermi problems in the context of our minimal formal account (with some further details in an appendix). We conclude with a summary in Sec. 4.

[1] Note that part of our characterization of Fermi problems includes how one goes about solving them. They are named after physicist Enrico Fermi who highlighted such problems and famously estimated the strength of the first atomic bomb test based on the scattering of bits of paper (Von Baeyer, 1993).

[2] The way in which 'accuracy' is used, here, is not consistent with how it is used in the sciences—where it captures how close a possible outcome is to the true outcome. A less ambiguous term would be the "credence" or "probability" of a guess, which we indeed prefer but won't adopt for the sake of continuity with the recent philosophical literature.

[3] Dorst and Mandelkern (2022) [and Skipper (2023)] also refer to 'permissible guesses' as 'good guesses'. However, the former phrase strikes as most apt given their goal of identifying guesses that reflect aspects of the practice of guessing. Furthermore, what they call 'good guesses' can still be (relatively) poor guesses when one allows for credences to change—as we'll describe in Sec. 3.

2 What makes for a permissible guess?

Thinking about the practice of guessing reveals regularities in the guesses made by agents. Such regularities, when suitably formalized, serve as constraints that theories of guessing ought to satisfy. Dorst and Mandelkern (2022) develop these regularities, partly drawing from the work of Kahneman and Tversky (1982) and Holguín (2022). We'll recap these regularities (as we'll reanalyze them in the context of our minimal formal account), after first setting up some notation and terminology.

2.1 The theory of Dorst and Mandelkern

In the standard guessing scenario, as it has recently been developed in the philosophical literature, an agent is asked to provide an answer to a specific question for which the answers, and credences for those answers, are specified. So, adapting (and summarizing) an example from Skipper (2023), let the question be about an upcoming horse race: "*Which horse is going to win?*". This question is associated with a set of *complete answers*, denoted by $Q \equiv \{Ajax, Benji, Cody, Dusty, Ember\}$. The *size* of the question, $|Q|$, is just the total number of complete answers: so that in this example, $|Q| = 5$. Let the credence function of the agent, denoted (generally) by P, take on values as shown in the following table.

Ajax	*Benji*	*Cody*	*Dusty*	*Ember*
0.4	0.3	0.25	0.04	0.01

Now, there are all sorts of answers that the agent could provide. One answer might simply be "*Ajax*" (short for "Ajax will win"—indeed one of the complete answers). This answer is somewhat accurate (given that it has a credence of 0.4), but it is not the most accurate answer one could give. The answer "*Ajax* or *Benji*" has a higher credence (of 0.7) and so is even more accurate. But "*Ajax* or *Benji*" it is less specific than "*Ajax*", for the former leaves out just three of the complete answers whilst the latter leaves out four. If you value accuracy over specificity, you might guess "*Ajax* or *Benji*", whereas if you value specificity over accuracy, you might guess "*Ajax*". Nonetheless, both of these guesses might seem like reasonable guesses (whereas "Ember", for instance, might strike you as unreasonable).

Dorst and Mandelkern summarize regularities in (rational) guessing practice in five principles of guessing [see Dorst and Mandelkern (2022, pp. 585–587)].[4]

> **Improbable Guessing**: It's sometimes permissible to answer p even when $P(p) < 0.5$. (Indeed, "*Ajax*" is one such permissible answer.)

> **Question Sensitivity**: Whether p is a permissible answer depends not just on the guesser's credence in p but also on what question is being answered. [If the question is changed from "*Which horse is going to win?*" to "*Will Ajax win or not?*", then "*Ajax*", with a credence of 0.4, doesn't seem permissible—for the only other available answer "*not-Ajax*" (short for "Ajax will not win") has a higher credence of 0.6.]

[4] We will slightly change the wording and/or the precise presentation of Dorst and Mandelkern's statements of some of the principles, partly to better accord with our notation [see also Skipper (2023) who presents some of the principles along similar lines to us].

Optionality: Given any question Q, for any k with $1 \leq k \leq |Q|$, it's permissible for your guess about Q to be the disjunction of exactly k complete answers to Q. [That is, it seems reasonable to accept the following guesses as permissible: "Ajax" (where $k = 1$); "*Ajax* or *Benji*" ($k = 2$); "*Ajax* or *Benji* or *Cody*" ($k = 3$); "*Ajax* or *Benji* or *Cody* or *Dusty*" ($k = 4$); "*Ajax* or *Benji* or *Cody* or *Dusty* or *Ember*" ($k = 5$).]

Filtering: A guess about Q 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. (So, if a guess includes "*Cody*", then it must also include both "*Ajax*" and "*Benji*": "*Cody*" as a guess on its own seems unreasonable.)

Fit: p is a permissible guess only if there are $q_1, q_2, \ldots, q_k \in Q$ such that $p = q_1$ or q_2 or $\ldots$ or q_k. (Thus, "*Ajax* or *It won't rain tomorrow*" is an impermissible guess to the question "*Which horse is going to win?*", when this question is associated with the complete answers $Q \equiv \{Ajax, Benji, Cody, Dusty, Ember\}$.)

These five principles serve as constraints on theories of guessing. The theory of Dorst and Mandelkern (2022) characterizes the act of arriving at a permissible guess as one that optimizes a particular function: one that expresses a trade-off relation between accuracy and specificity. As a result, the permissible guesses that one can generate satisfy all the constraints above. More formally, a guess, p, is permissible if and only if it maximizes the *expected answer-value* function, $E(p; J)$, where

$$E(p; J) = P(p) \times J^{S(p)}. \tag{1}$$

The first term on the right-hand side, $P(p)$, is the agent's credence in p—this term accounts for the accuracy of the guess. The second term, $J^{S(p)}$, is a parameter-based weighting of the specificity $S(p)$, in terms of a parameter $J \geq 1$, which is unbounded from above. For Dorst and Mandelkern, the specificity of a guess is defined by

$$S(p) \equiv 1 - \frac{c_p}{|Q|}, \tag{2}$$

where c_p is the number of complete guesses that are compatible with the guess p. If p leaves out no complete guesses, it is not specific at all: indeed, $c_p = |Q|$, and the specificity takes on its minimal value of $S(p) = 0$. If p is compatible with just one complete answer then the guess is maximally specific: $c_p = 1$ and the specificity takes on its maximal value of $S(p) = 1 - 1/|Q|$.

The agent is free to specify a value of J depending on how much they might favor specificity over accuracy (indeed there is no objective value it ought to take in any particular guessing scenario). When $J = 1$ the objective function is characterized purely by the first term (so the maximization routine will yield a guess with the highest probability). As J assumes larger values, the specificity becomes more important. And, indeed, different choices for J will yield different permissible guesses. In the horse-race example, above, setting $J = 1$ will yield, as the permissible guess, the full disjunction over complete answers: "*Ajax* or *Benji* or *Cody* or *Dusty* or *Ember*". For sufficiently large J, only "*Ajax*" will be permissible.

In this way, the account of Dorst and Mandelkern (2022) provides a quantitative characterization of how to arrive at a permissible guess, in a setting in which complete guesses and credences associated with those complete guesses are known to the guesser. What is clearly captured by this account is that when confronted with a question about whose answer we are uncertain, there is not just one reasonable guess we can adopt: reasonable guesses can vary depending on how we weigh accuracy against specificity. Another key merit of their account is that it reproduces plausible regularities that arise when one analyzes the practice of guessing.[5]

Despite these merits, the account is developed for (standard) guessing problems in which the (complete) answers are unambiguously identified and the credences over those answers are known and *fixed*. But these aren't the only types of guessing problems we encounter. In what follows, we present a wholly different account of guessing that: (i) not only recovers, in a more parsimonious manner, the constraints described above (as well as other predictions of guessing); but (ii) accounts for improvements in guessing scenarios and (iii) provides a window into a rather ubiquitous type of guessing problem (a *Fermi problem*) that goes beyond the standard problem as explicated above.

2.2 A minimal formal account of permissible guesses

Our minimal formal account of guessing identifies permissible guesses in terms of a trade-off between the qualities of accuracy and specificity. But we identify such guesses in a general fashion (without invoking epistemic utility theory). In our case, as we will show in this section, permissible guesses are those that lie along a 'Pareto front': a collection of points (or more generally a surface) that simultaneously optimize multiple qualities of interest (in our case, accuracy and specificity). Such fronts arise very broadly in multiobjective optimization problems in economics and in the sciences.

First, a brief outline of how we'll identify permissible guesses on our account. We will: (1) define a set of 'fit guesses', denoted by G_Q, by adopting a version of **Fit** from Sec. 2.1; (2) define, using minimal formal machinery, the accuracy and specificity of any guess in G_Q; and finally (3) define how one guess in G_Q can (weakly) dominate another. This will allow us to identify a Pareto front in an accuracy-specificity plot of all fit guesses. Fit guesses that lie along the Pareto front will be identified as the *permissible guesses*, with the remaining fit guesses being impermissible guesses. Let's now turn to the details.

For a question whose complete answers form a set labeled by Q, we adopt

fit: p is a *fit guess* if and only if there are $q_1, q_2, \ldots, q_k \in Q$ such that $p = q_1$ or q_2 or $\ldots$ or q_k.

This principle is only slightly different from **Fit** as defined in Sec. 2.1, and so we have denoted it by a lowercase f. Let the set of fit guesses, G_Q, be the set of all possible disjunctions among the complete answers:

$$G_Q \equiv \text{Set of all fit guesses derived from } Q \text{ by forming all disjunctions.} \tag{3}$$

(We'll say more about various features of this set later.)

Each (fit) guess $p \in G_Q$ can be assigned an accuracy and a specificity. The accuracy of p is the sum of the individual credences for each disjunct (each complete answer) that makes up the guess. Thus the accuracy of a guess is defined in the same way as in Dorst and Mandelkern

[5] Skipper (2023) also broadly endorses a trade-off conception of permissible guesses and employs epistemic utility theory. The key difference between their accounts is in how Skipper defines the specificity of a guess. (Indeed, Skipper's paper is broadly a reply to that of Dorst and Mandelkern.) For Skipper, $S(p) \equiv \log\left(|Q|/c_p\right)$. This definition is consistent with a different (but overlapping) set of constraints on guessing than that endorsed by Dorst and Mandelkern. We'll return to these differences and how they relate to our work toward the end of Sec. 2.3.

(2022). More formally, we'll denote the accuracy of a guess p by $A(p)$ and we can think of this as a function that maps a guess to its credence:

$$\begin{aligned} A : G_Q &\to [0,1] \\ p &\mapsto A(p) \equiv P(p), \end{aligned}$$

where $P(p)$ is the credence of p. To define the specificity of p we adopt the general idea [expressed by Dorst and Mandelkern (2022) and by Skipper (2023)] that the number of complete answers that are *left out* by p provides a measure of how specific is the guess. Then, defining

$$c_p \equiv \text{Total number of complete answers that comprise the fit guess } p,$$

any monotonically increasing function of $|Q|/c_p$ will have the following desirable property. If all but one complete answer is left out of p, then p has the highest possible specificity; and the specificity gets smaller as more complete answers are included in p (that is, as c_p gets larger). Thus the specificity of a guess will generally be denoted by $S(p) = f(|Q|/c_p)$, where f is any monotonically increasing function.[6] Note that the smallest value of the specificity of a guess will be obtained by the guess that includes all complete answers: $c_p = |Q|$, in which case $S(p) = f(1)$. And the largest value of the specificity of a guess will be obtained for a guess that includes just one complete answer: $c_p = 1$, in which case $S(p) = f(|Q|)$. More formally,

$$\begin{aligned} S : G_Q &\to [f(1), f(|Q|)] \\ p &\mapsto S(p) \equiv f(|Q|/c_p). \end{aligned}$$

Each fit guess can be associated with a pair of numbers $(S(p), A(p))$ that can be represented by a point on a two-dimensional plane, which we'll call 'objective space', shown in Fig. 1. To keep notation as simple as possible (and unambiguous), we'll denote the pair of numbers associated with a fit guess p by $\hat{p} \equiv (S(p), A(p))$.

We now introduce definitions for when a fit guess can *weakly dominate* or *dominate* another fit guess [see Giagkiozis and Fleming (2014) for some background].

> A fit guess p_1 will *weakly dominate* (WD) a fit guess p_2 if and only if $S(p_1) \geq S(p_2)$ and $A(p_1) \geq A(p_2)$ and at least one of these inequalities is a strict inequality.

Thus weak dominance is characterized by better performance in at least one of the two desired qualities we aim for in guesses (namely, the qualities of specificity and accuracy). The relation can be made vivid in objective space, as illustrated in Fig. 2 (left). Next consider the following definition.

> A fit guess p_1 will *dominate* (D) a fit guess p_2 if and only if $S(p_1) > S(p_2)$ and $A(p_1) > A(p_2)$.

[6] Although we leave the function that defines the specificity, f, arbitrary (but monotonically increasing), a simplifying assumption, which will suffice for our purposes, is that it is fixed for all agents across all questions. (Indeed, this is the assumption effectively employed by Dorst and Mandelkern and by Skipper.) One can then assign something like an objective quality to the values assumed by f, so that we can compare the specificity of permissible guesses across questions—just like it makes some sense to compare the accuracy (that is, the probability) of permissible guesses across questions. Furthermore, for any specific f, we will assume that the value of the specificity obtained for a fit guess formed from some k disjunctions will be deemed (by the agent) to be significantly *different* from the value obtained for a fit guess formed from $k+1$ disjunctions—indeed the latter value will signal (to the agent) a significantly less specific guess.

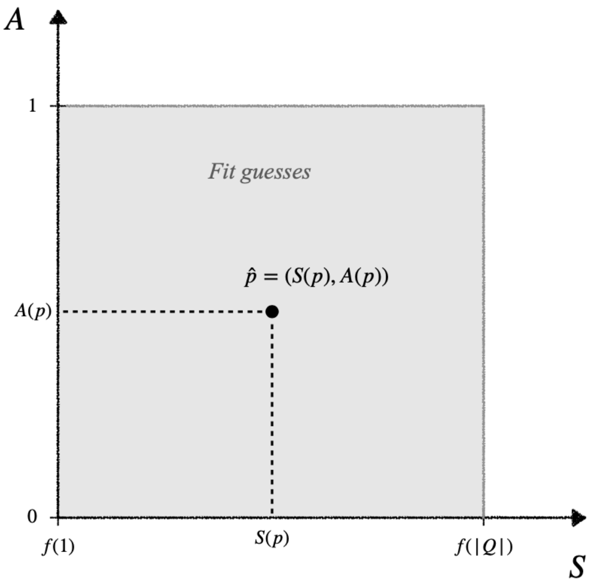

Figure 1: Objective space: the two-dimensional space in which all fit guesses can be located. A denotes accuracy, S denotes specificity. f is an arbitrary monotonically increasing function. A single fit guess p is represented by $\hat{p}$.

Domination is characterized by better performance in both dimensions of the two desired qualities we aim for in guesses—see Fig. 2 (right). Note that if a guess p_1 dominates another guess p_2, then p_1 weakly dominates p_2. Thus we arrive at the following definition.

> A fit guess p is *Pareto optimal* if and only if there is no other fit guess that weakly dominates p. (And so, by the above discussion, there is also no other fit guess that dominates p.)

Simply, a fit guess is Pareto optimal if there is no other fit guess that does better on either (or both) of the two desired qualities we aim for in guesses. Fig. 3 illustrates this pictorially. This leads us to our key definition, namely that of a permissible guess:

A *permissible guess* is a fit guess that is Pareto optimal.

The points in objective space that correspond to the specificity-accuracy pair for all Pareto optimal (that is, permissible) guesses denotes the *Pareto front*; and it is the Pareto front for a given Q that represents the relevant trade-off 'curve' for our analysis.

Before we work through an example that will help to tie together the various definitions above, we'll highlight a straightforward algorithm (not the only such algorithm) for generating the Pareto front, given a question and a credence function over complete answers for that question. To describe the algorithm, we'll need to say a few more things about the set of fit guesses G_Q.

The set G_Q can be straightforwardly partitioned into cells—that is, broken up into disjoint subsets. Let subset k of the partition, denoted by g_k, contain all fit guesses that are derived from Q by forming the disjunction of exactly k complete guesses. [So, subset 1, denoted by g_1 is simply Q. Subset 2, denoted by g_2, will contain guesses of the form $p = q_1$ or q_2 where

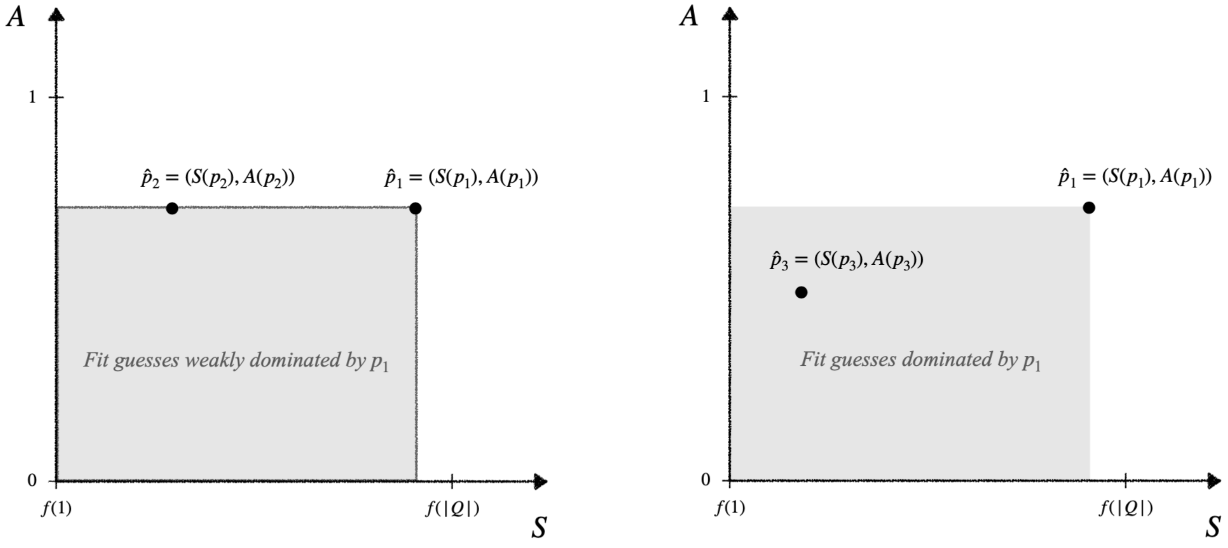
A
1
$\hat{p}_2 = (S(p_2), A(p_2))$
$\hat{p}_1 = (S(p_1), A(p_1))$
Fit guesses weakly dominated by p_1
0
$f(1)$
$f(|Q|)$
S
A
1
$\hat{p}_1 = (S(p_1), A(p_1))$
$\hat{p}_3 = (S(p_3), A(p_3))$
Fit guesses dominated by p_1
0
$f(1)$
$f(|Q|)$
S

Figure 2: Illustrating weak dominance (left) and dominance (right). In the left panel, guess p_1 weakly dominates all fit guesses that fall in the shaded region including the boundary lines (but not including the point $\hat{p}_1$). For example, p_1 weakly dominates p_2. In the right panel, guess p_1 dominates all fit guesses that fall in the shaded region (not including the boundary lines). Thus p_1 dominates p_3.

$q_1, q_2 \in Q$.] The size of g_k is given by the number of ways of choosing exactly k complete answers (out of which one forms disjunctions) from $|Q|$ complete answers. This number is given by the *binomial coefficient* $|g_k| = \binom{|Q|}{k} \equiv |Q|!/(k!(|Q|-k)!)$. [Recall that $n! \equiv n \times (n-1) \cdots \times 1$. If $|Q| = 5$, as for the horse-race example earlier, then g_2 will contain $\binom{5}{2} \equiv 5!/(2!(5-2)!) = 10$ elements.] Letting k range over $1, 2, \ldots, |Q|$, thus generates the various subsets of the partition. And so we may write: $G_Q = g_1 \cup g_2 \cup \cdots \cup g_{|Q|}$.[7]

Now we can describe the algorithm that generates the Pareto front.

For each of $k = 1, 2, \ldots, |Q|$:

1. Construct the subset g_k. (Note that for each element of a given g_k, the number of complete answers that comprise that element, c_p, is just $c_p = k$.)
2. For each element p of g_k:

   2a. Compute the specificity of p, S(p). (For each $p \in g_k$, this is given by $S(p) \equiv f(|Q|/c_p) = f(|Q|/k)$.

[7] The size of the set of all fit guesses, $|G_Q|$, is $2^{|Q|} - 1$. We can see this by noting that the size of G_Q is just the sum of the sizes of the subsets in the partition: $|G_Q| = |g_1| + |g_2| + \cdots + |g_{|Q|}|$. Given that $|g_k| = \binom{|Q|}{k}$, we have

$$|G_Q| = \binom{|Q|}{1} + \binom{|Q|}{2} + \ldots + \binom{|Q|}{|Q|} = 2^{|Q|} - 1. \tag{4}$$

The expression on the right-hand side of Eq. (4) can be found from the binomial theorem, which states that $(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}$. (Simply set $x = 1 = y$, $n = |Q|$, and then rearrange the terms to obtain the claimed result.)

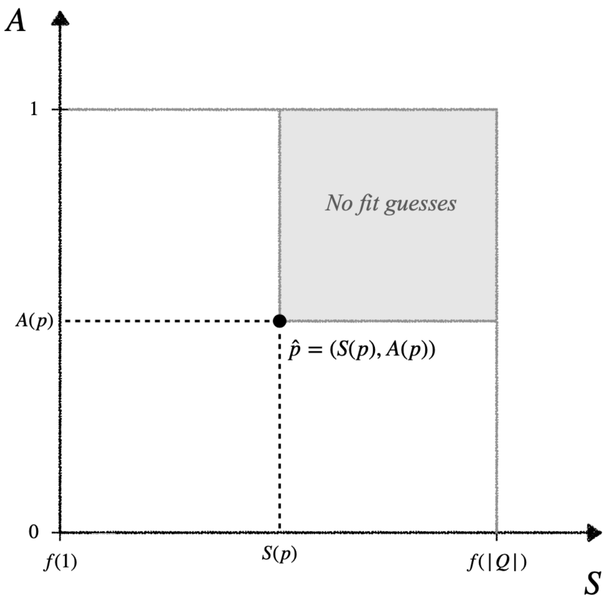

Figure 3: Pareto Optimality. Fit guess p is not weakly dominated by any other fit guess if there are no fit guesses that are mapped into the shaded region (including the boundaries of that region), but not including the point $\hat{p}$.

2b. Compute the accuracy of p, $A(p) \equiv P(p)$. (This is given by the sum of the credences of the k complete answers that comprise p.)

2c. Plot $\hat{p} = (S(p), A(p))$ in objective space.

The Pareto front corresponds to the set of points that, for each subset (or, equivalently, each possible specificity value), has maximal accuracy. The type of plot that arises as a result of this algorithm is illustrated in Fig. 4. A line that interpolates between points on the Pareto front will be referred to as a 'trade-off line'.

Let's now briefly tie together some of the definitions, above, by providing an explicit example of the permissible guesses one obtains via the above construction. Consider again the horse race example from Sec. 2.1, where $Q = \{Ajax, Benji, Cody, Dusty, Ember\}$. Let the credence function over the complete guesses be as described in the table there [that is, $P(Ajax) = 0.4$, $P(Benji) = 0.3$, $P(Cody) = 0.25$, $P(Dusty) = 0.04$, $P(Ember) = 0.01$]. The subset g_1 is just Q. The accuracy of each guess p in g_1 is $A(p) \equiv P(p)$ and this is the distribution of credences just mentioned. The specificity of each guess in g_1 is $f(|Q|/k) = f(5)$. The guess that lies on the Pareto front from g_1 will be the guess with the highest accuracy and this is "*Ajax*", with an accuracy of 0.4. The subset g_2 is given by all pairwise disjunctions between elements in Q. The specificity of each pair is the same and is given by $f(|Q|/k) = f(5/2) = f(2.5)$. The guess that lies on the Pareto front from g_2 will be the guess with the highest accuracy and this just "*Ajax* or *Benji*", with an accuracy of 0.7. One can repeat this construction for each subset g_3, g_4, and g_5. The remaining guesses on the Pareto front are (respectively) "*Ajax* or *Benji* or *Cody*" $[(S(p), A(p)) = (f(5/3), 0.95)]$, "*Ajax* or *Benji* or *Cody* or *Dusty*" $[(S(p), A(p)) = (f(5/4), 0.99)]$, and '*Ajax* or *Benji* or *Cody* or *Dusty* or *Ember*" $[(S(p), A(p)) = (f(1), 1)]$.[8]

[8] Note that if there are ties in credences, then (unproblematically) there can be different guesses that lie on the

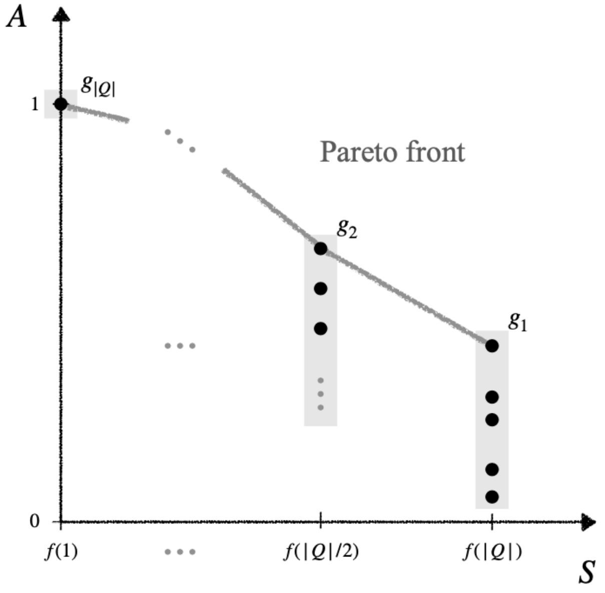

Figure 4: An illustration of the Pareto front, as it arises via the algorithm described in the main text. A solid black dot corresponds to a specificity-accuracy pair $\hat{p} = (S(p), A(p))$ for a fit guess p. Such pairs appear in bands (highlighted in light gray) depending on whether p arises from the disjunction of one complete answer ($p \in g_1$) or from two complete answers ($p \in g_2$), etc. The chalk-gray line—what we dub the 'trade-off line'—is an interpolation between those fit guesses that define the Pareto front.

Note that our account doesn't recommend a single "best" guess on the Pareto front. For such a recommendation, one requires extra theoretical structure. Nonetheless, we'll highlight one (general) way in which such structure can be included. Borrowing terminology from economics, there is an *opportunity cost* associated with choosing guesses with a higher specificity, namely, the cost of lower accuracy. If this drop in accuracy is deemed to be 'too high' then a guesser ought not to accept the more specific guess. Purely formally, "too high" can be cashed out in terms of the (magnitude of the) slope of the trade-off line. (So, if this magnitude is too large, the opportunity cost is too high.) Given a criterion for assessing when a cost is too high (we'll come back to how one can determine such a criterion), one can thus use the slope of the trade-off line to identify *preferred* 'regions' of the Pareto front. In particular, guessers who ultimately care more about accuracy than specificity, may select guesses in regions on the Pareto front that are on the high-accuracy side of a trade-off line that exhibits an unacceptably steep decline in accuracy.

How, then, does one determine a criterion for assessing when a cost is too high. In general, such criteria are determined by considering some external value relevant to the setting of interest—such as 'profit' or, perhaps, the happiness of an agent. Such external values provide ways to determine criteria of the type that are indeed in play in the guessing scenarios analyzed by Dorst and Mandelkern and by Skipper. For them, the relevant external value is *expected utility*

Pareto front. For instance, if $P(Ajax) = 0.6$, $P(Benji) = 0.15$, $P(Cody) = 0.1$, $P(Dusty) = 0.1$, $P(Ember) = 0.05$, then the fit guesses $p_1 \equiv$ "*Ajax* or *Benji* or *Cody*" and $p_2 \equiv$ "*Ajax* or *Benji* or *Dusty*" both lie on the Pareto front.

(and maximizing expected utility allows for the selection of a single best guess). Our account, as a minimal account, remains agnostic as to what such values ought to be (and so is agnostic about the selection of preferred regions on the Pareto front).[9] We think it is plausible that such values are sensitive to the guessing context and to the guesser.

2.3 Accounting for constraints on guessing

We now address whether and in what way our minimal formal account can capture the various patterns on guessing described by Dorst and Mandelkern (2022, § 2).

> **Improbable Guessing**: It's sometimes permissible to answer p even when $P(p) < 0.5$.
>
> If there is just one complete answer that has the highest credence assigned to any complete answer and that highest credence is less than 0.5, then this guess will be a permissible guess. It will be the only guess from the subset g_1 that lies on the Pareto front. (If there are ties in credences then there may be more than one such permissible guess from g_1.) If the two highest credences assigned to complete answers sum to less than 0.5, then this constraint will also be satisfied by a guess from g_2. (Such a pattern may extend to other subsets g_k.)
>
> **Question Sensitivity**: Whether p is a permissible answer depends not just on the guesser's credence in p but also on what question is being answered.
>
> Consider a guess q that lies on the Pareto front. It is therefore, according to our account, permissible. If the credence of q is less than 0.5, then one can rephrase the question to obtain a new question so as to ensure that q does not lie on the (new) Pareto front. The guess q will thereby be rendered impermissible. Consider again the horse-race example: if *Ajax* has a credence of winning of 0.25 and this is the highest among all the horses, then "*Ajax*" (an element of g_1) will lie on the Pareto front associated with the question "*Which horse is going to win?*"; but it won't lie on the Pareto front associated with the question "*Will Ajax win or not?*". With respect to the latter question, the guess from g_1 that will lie on the Pareto front will be "*not-Ajax*" (and this answer will weakly dominate "*Ajax*" because it has the same specificity as "*Ajax*" but a higher accuracy).
>
> **Optionality**: Given any question Q, for any k: $1 \leq k \leq |Q|$, it's permissible for your guess about Q to be the disjunction of exactly k complete answers to Q.
>
> The Pareto front is defined via one point from each of the subsets $g_1, ..., g_{|Q|}$ and so the set of all permissible guesses will exhaust the ways in which disjunctions may be formed. The only case in which this won't arise is if there are complete answers with a credence of zero. As a limiting case of such a scenario, consider the case in which one complete answer has a credence of 1, and the remaining complete answers (indeed however many of them there are) have a credence of zero. Then the Pareto front will correspond to just one point (in the far upper-right of objective space). But, this is a special case (as are cases where there is at least one complete answer with a credence of zero) and we exclude such exceptional cases from consideration.

[9] Note that one can roughly translate between the two approaches: the effect of maximizing expected utility after choosing a J on their accounts corresponds to the effect of choosing a point on the Pareto front.

Filtering: A guess about Q 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.

Again, this is guaranteed by construction. For any subset g_k the point on the Pareto front from g_k will correspond to a filtered guess. All non-filtered guesses have lower accuracy than the filtered guess and aren't on the Pareto front. Note also that Skipper (2023) claims that **Filtering** can be equivalently formulated as

> **No accuracy-dominance**: A permissible guess (p') is never accuracy dominated, where p accuracy dominates p' iff $c_p \leq c_{p'}$ & $P(p) > P(p')$.

This is also straightforward to see on our account. Let p accuracy dominate p'. Then since $P(p) > P(p')$, p will have higher accuracy than p'. And since $c_p \leq c_{p'}$, the specificity of p, $S(p)$, will be greater than or equal to the specificity of p'.[10] In objective space, these two facts mean that p will be represented by a point ($\hat{p}$) above and possibly to the right of the point representing p' ($\hat{p}'$). In which case, p would weakly dominate p' and p' would not appear on the Pareto front.[11]

Fit: p is a permissible guess only if there are $q_1, q_2, \ldots, q_k \in Q$ such that $p = q_1$ or q_2 or $\ldots$ or q_k.

Note that permissible guesses on our account are all guesses that satisfy **fit**. That is, they are guesses that are formed from disjunctions of complete answers in Q. (Indeed permissible guesses on our account are generally a proper subset of guesses that satisfy fit.) In this way permissible guesses on our account straightforwardly satisfy **Fit** as well.

In this way, our minimal formal account captures the various constraints on guessing described by Dorst and Mandelkern. But note that these aren't the only constraints that have been discussed in the recent literature. In particular, Skipper (2023) describes a theory of guessing that satisfies two further potential constraints not satisfied by the theory of Dorst and Mandelkern [what he calls **Neutrality** and **Independence of irrelevant alternatives (for guessing)**]. Furthermore (and as it turns out, more importantly for us) Skipper's account does not generally satisfy **Optionality**. The two further potential constraints need not detain us, however. The problems that Skipper highlights for the theory of Dorst and Mandelkern arise specifically in the context of epistemic utility theory—and our minimal formal account does not include the extra theoretical structure that exposes a difference in their accounts as regards guessing constraints that are satisfied. This difference, to be clear, arises as a result of (i) the specific functional form adopted for the specificity function [$S(p)$] and (ii) technicalities that relate to the trade-off parameter J (described in Sec. 2.1). Our minimal formal account is agnostic about the specific functional form for specificity and we do not involve any such trade-off parameter explicitly. In this way, one could look to further develop our account along either of the two (or indeed other) lines—and we see this as a virtue of our account.

[10] This claim about the relative sizes of the specificity is straightforward to establish. Note that $c_p \leq c_{p'} \iff |Q|/c_p \geq |Q|/c_{p'} \iff S(p) \equiv f(|Q|/c_p) \geq S(p') \equiv f(|Q|/c_{p'})$, where, recall, f is monotonically increasing.

[11] Skipper's **No specificity-dominance** [also satisfied by the account of Dorst and Mandelkern (2022)] is also satisfied by construction. This constraint reads as follows: A permissible guess p' is never specificity-dominated, where p specificity-dominates p' iff $c_p < c'_p$ & $P(p) \geq P(p')$. By a similar argument to that presented in the main text, in the case where p specificity-dominates p', p would weakly dominate p' and p' would not appear on the Pareto front.

As regards **Optionality**, the situation is perhaps more subtle. As we have described above, our account by construction does satisfy this regularity of guessing (and Skipper's doesn't). This regularity won't be satisfied when it's not the case that a permissible guess will come from each of $g_1, g_2, \ldots, g_{|Q|}$ (in our notation): that is, there will be at least one value of k for which there won't be a permissible guess formed from the disjunction of exactly k complete answers to Q. To make sense of the difference between our accounts, it will help to analyze a counterexample to **Optionality** that Skipper develops. He considers credences for a three-way horse race that take the following values:

Ajax	*Benji*	*Cody*
0.5	0.26	0.24

Notably the credences for "*Benji*" and "*Cody*" aren't too far apart. On our account, the Pareto front will be defined by the pair of values in objective space assumed by the following three guesses: one for each of g_1, g_2, and g_3, respectively: "*Ajax*" (from g_1); "*Ajax* or *Benji*" (from g_2); and "*Ajax* or *Benji* or *Cody*" (from g_3). Any three of these guesses are permissible, according to our minimal formal account.

The only permissible guesses on Skipper's account are "*Ajax*", or else "*Ajax* or *Benji* or *Cody*". The counter-example arises because increasing J (the degree to which one values specificity over accuracy) smoothly transitions the guess with the highest expected answer value from "*Ajax* or *Benji* or *Cody*" (the guess with the highest expected answer value when, say, $J = 1$—where all that matters is accuracy) to "*Ajax*" (when the value of J is sufficiently high). His account renders as impermissible the selection of just one of *Benji* or *Cody* to pair with *Ajax*, when credences for the two choices one would need to cut across to effect such a pairing are sufficiently close together. (Dorst and Mandelkern allow such cross-cutting but find it odd).

We see therefore that our account includes as a permissible guess one that is impermissible according to Skipper's account (but that is permissible on the account of Dorst and Mandelkern). In this way, our minimal formal account will render guesses as permissible that might be rendered impermissible, if one were to add further theoretical structure. But which guesses will be rendered impermissible is not settled: the structure added by Dorst and Mandelkern renders the guess "*Ajax* or *Benji*" permissible (but perhaps odd); that added by Skipper renders that guess impermissible. Adjudicating between this dispute lies outside the scope of our remarks, here.

2.4 Accounting for predictions about guessing

An interesting application of the account developed by Dorst and Mandelkern is to the *conjunction fallacy*, for which they provide an explanation in terms of their theory. In this section, we'll exhibit a different explanation in terms of the concept, introduced above, of weak dominance. We'll furthermore exhibit explanations for key empirical predictions described by Dorst and Mandelkern for guesses that agents make, but where we account for such predictions in terms of weak dominance.

The conjunction fallacy arises when an agent judges the conjunction of two propositions as *more probable* than either of its conjuncts. This judgement contradicts a standard result in probability theory in which the probability of a conjunction is strictly less than or equal to the probability of either of its conjuncts.[12] There is a large literature on this fallacy but perhaps the most famous example is that provided by Tversky and Kahneman (1983), sometimes known as the *Linda problem*.

> *Linda problem*: 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 (Tversky and Kahneman, 1983, p. 297).
>
> Which of the following two possibilities do you take to be more probable:
>
> (i) *Linda is a bank teller* (T);
>
> (ii) *Linda is a bank teller* and *Linda is active in the feminist movement* ($T\&F$)?

Tversky and Kahneman found that 85% of agents judged that the conjunction $T\&F$ was *more probable* than the conjunct T.

Dorst and Mandelkern account for the conjunction fallacy by providing

> **The Answer-Value Account**: People commit the conjunction fallacy because they rank outcomes according to their expected answer-value, rather than their probability (Dorst and Mandelkern, 2022, p. 604).

On their account, although people may not be correctly answering the question as phrased (namely, they aren't correctly judging which of the two outcomes is more probable), they *are* accurately ordering the outcomes according to *some* rational criterion: that of having the highest expected answer-value. Dorst and Mandelkern exhibit the answer-value account of the conjunction fallacy in the context of the *Linda problem* in the following way. Let the question be partitioned as follows: $Q = \{T\&F, T\&\overline{F}, \overline{T}\&F, \overline{T}\&\overline{F}\}$. (Here an overbar corresponds to a negation of the underlying proposition: for example, $\overline{F} \equiv$ 'Linda is not active in the feminist movement'.) They then show that there are values of $J \equiv J^*$ such that the expected answer value of $T\&F$ is higher than that of T: $E(T\&F; J^*) > E(T; J^*)$ [using notation introduced in Eq. (1)]. That is, although the question was originally about which of T and $T\&F$ has a higher *probability*, there are cases (for agents with a sufficiently high $J = J^*$) where $T\&F$ has a higher expected answer-value than T; and agents who answer "$T\&F$" are reporting a ranking based on the latter criterion. (Note that this holds for a "sufficiently high $J = J^*$", because $T\&F$ is more specific than T.)

Here we provide an alternative explanation that, like Dorst and Mandelkern's explanation, crucially relies on an accuracy-specificity trade-off, but where we do not employ epistemic utility theory. In particular, the fallacy can be explained via an appeal to the notion of weak dominance. We'll do this specifically in the context of the *Linda problem*, but our remarks will easily generalize.

In the *Linda problem*, as construed above, where complete answers are presented as $Q = \{T\&F, T\&\overline{F}, \overline{T}\&F, \overline{T}\&\overline{F}\}$, there are no explicit credences specified. Certainly, in the way the problem is presented, the agent is not given, for example, frequencies that might guide rational credences. Thus, insofar as the agent who is tackling the problem is doing so via some sort of an accuracy-specificity trade-off (which we are indeed interested in exploring, here), there is plausibly some uncertainty about the credences that one might adopt for each of the complete answers.

[12] For instance, if A and B are the propositions of interest then, using the definition of conditional probability, $P(A\&B) = P(A|B)P(B)$. Since $P(A|B)$ is a probability distribution, $P(A|B) \leq 1$. Thus we must have that $P(A\&B) \leq P(B)$ [a similar argument establishes that $P(A\&B) \leq P(A)$].

In this way, the *Linda problem* is rather different from the examples first employed in the papers by Dorst and Mandelkern (and by Skipper) that help to explicate their theories of guessing. In those examples (as for the horse-race example above), one has crisp credences that are given by past frequencies, for instance, or that are at least explicitly specified at the outset. In the *Linda problem* there are no such explicit credences.

Now, insofar as an agent is indeed using an accuracy-specificity trade-off, credences must enter the picture (for accuracy is defined in terms of credences)—this much we will assume. But that such credences are sharp (that is, that they take specific numbers) is defeasible. It is perhaps more plausible that upon hearing the question, the agent will adopt credences that come with room for error: for instance, credences for the two propositions being compared will come with a tolerance of, say, 10% for each.[13] Thus the agent finds themselves in an epistemic situation in which there is uncertainty about the credences of the two propositions of interest [$T \equiv (T\&F) \vee (T\&\overline{F})$ and $T\&F$], but where one of those propositions ($T\&F$) is a good deal more specific than the other (T). If the credences of the two propositions of interest are judged by the agent to lie within the tolerance of the other, then it is plausible that the agent judges them to be roughly equal $P(T) \approx P(T\&F)$. In which case the (unambiguously) greater specificity of one of the alternatives ($T\&F$) means that it will (approximately) weakly dominate the other.[14] This motivates why the agent may respond that "*Linda is a bank teller* and *Linda is active in the feminist movement*" is more 'probable' than "*Linda is a bank teller*".

So here is our more general response to the conjunction fallacy,

> **The Weak-Dominance Account**: People commit the conjunction fallacy because, under uncertainty about credences of two options under consideration, they rank the outcomes according to whether one option (approximately) weakly dominates the other, rather than the probability of the options.

As for Dorst and Mandelkern, guesses such as $T\&F$ aren't necessarily permissible guesses—in our case this amounts to the claim that they don't necessarily lie on the relevant Pareto front. But by virtue of the concept of weak dominance (when crisp credences can be assigned) and approximate weak dominance (when there are uncertainties about values of credences) one can order answers to problems by how well they do in terms of an accuracy-specificity trade-off. One loses some amount of precision compared with the account provided by Dorst and Mandelkern in that there isn't a single objective function (the expected answer-value in their case) that can be used to order answers to a question. Nonetheless without this extra theoretical structure, it is clear that much can be recovered by an account based on weak dominance. Also, the putative precision gained in applying tools from epistemic utility theory runs into issues that remain unsettled—such as, whether the extra structure is apt and precisely what extra structure ought to be introduced.[15]

[13] Note that one could look to be more (mathematically) precise about how such tolerances are specified. Since one may decompose T into mutually inconsistent disjuncts $T \equiv (T\&F) \vee (T\&\overline{F})$, one has that $P(T) = P(T\&F) + P(T\&\overline{F})$. Thus, uncertainties about each of $P(T\&F)$ and $P(T\&\overline{F})$ can be formally propagated to yield uncertainties in $P(T)$. This extra mathematical precision we believe obscures the main point, which is captured by there being *some* tolerance in the credence of each proposition that is being compared.

[14] Approximate weak dominance in this instance only arises when the credences are within specified tolerances of each other. Recall we assume that distinct values of the specificity (for answers with different numbers of disjuncts) will always be judged to be significantly different (see fn. 6).

[15] Note that there is a way in which to develop a response to the conjunction fallacy based on considerations of slope of the trade-off line that is complementary to **The Weak-Dominance Account**. As discussed toward the end

Note that we aren't necessarily arguing here that a rational mind isn't using expected answer value-type accounts to solve such problems. If they are, the account presented above is broadly consistent with it. But if they aren't then the above account provides a way to think of the type of computation that occurs when an agent is deciding among possible answers and is engaged in optimizing a trade-off between accuracy and specificity.

The scope of the concept of weak dominance is broad. Dorst and Mandelkern describe a number of predictions related to **The Answer-Value Account** that can also be accounted for by **The Weak-Dominance Account**. Here (due to limitations of space) we'll highlight what we take to be two of their more salient predictions.[16] (We leave a more empirically grounded assessment of the scope of the **The Weak-Dominance Account** for future work.)

> **Prediction 1**: Ranking $A\&B$ over B will be more common as $P(A|B)$ goes up.
>
> By the definition of conditional probability, $P(A\&B) = P(A|B)P(B)$. Of course, this means that $P(A\&B) \leq P(B)$. Increasing $P(A|B)$ will draw $P(A\&B)$ closer to $P(B)$. For an agent with some tolerance about credences, eventually, $P(A|B)$ will be large enough so that $P(A\&B)$ will be judged to be so close to $P(B)$ that $A\&B$ will be judged to (approximately) weakly dominate B; thus the agent will indeed rank $A\&B$ over B.
>
> **Prediction 3**: When $P(A|B)$ and $P(B|A)$ are *both* high, 'double'-conjunction fallacies will be common: people will rank $A\&B \succ A, B$. Meanwhile, when $P(A|B)$ is high but $P(B|A)$ is low, 'single'-conjunction fallacies will be common: people will rank $A \succ A\&B \succ B$.
>
> As just described, a high value of $P(A|B)$ can lead to a situation where $A\&B$ will be judged to (approximately) weakly dominate B. Similarly, a high value of $P(B|A)$ can lead to a situation where $A\&B$ will be judged to (approximately) weakly dominate A. And thus a 'double'-conjunction fallacy may arise. But if $P(A|B)$ is high and $P(B|A)$ is low then it is possible that: (i) owing to a small value of $P(B|A)$, $P(A\&B) \ll P(A)$, so that the usual conjunction fallacy as it might otherwise relate to $(A\&B)$ v. A is averted; but (ii) owing to a large value of $P(A|B)$, $A\&B$ will be judged to (approximately) weakly dominate B. And thus the usual 'single'-conjunction fallacy may arise.

3 What makes for a better permissible guess?

As should hopefully be evident from our above discussion, we agree with a particular strain in the recent guessing literature in which guessing is characterized via a trade-off between accuracy

of Sec. 2.2, one can look to identify regions where the magnitude of the slope of the trade-off line is not too large. In such an instance, a "natural guess" would be one that lies to the right of a region where the slope is deemed to not be too large. (For then, the opportunity cost of increased specificity—in terms of a drop in accuracy—would not be deemed to be prohibitive.) Given this, here is how a response to the conjunction fallacy based on **The Weak-Dominance Account** and one based on natural guesses are complementary. On the **The Weak-Dominance Account**, $T\&F$ can be judged to outrank T because the former is judged to weakly dominate the latter—owing to probabilities that are presumably within some tolerance. But given that their probabilities are relatively close, $T\&F$ will also then be a natural guess (that is generally impermissible). On some account based on natural guesses, when the magnitude of the slope of the line joining T to $T\&F$ is small enough, $T\&F$ becomes a natural guess. But then $T\&F$ also weakly dominates T.

[16] The predictions will correspond to what Dorst and Mandelkern denote as **Prediction 1** and **Prediction 3**; their other predictions also follow straightforwardly in our account.

and specificity (Carter, 2020; Dorst and Mandelkern, 2022; Skipper, 2023). In the latter two papers, a guessing scenario is represented via a set of complete answers to a question, together with a probability distribution over the complete answers that (typically) corresponds to credences of an agent. In this section we analyze the phenomenon of the improvement (or deterioration) of a guessing situation for an agent. In particular, we'll focus on the issue of how permissible guesses can be improved upon, that is, on *better permissible guesses*. We will describe how to: (i) characterize better permissible guesses in the context of our minimal formal account of guessing; and (ii) generate such guesses, by introducing and analyzing *Fermi problems*—which extend the types of guessing scenarios that have been in focus in the literature thus far. We'll furthermore show how the analysis in (ii) relates to our minimal formal account.

3.1 Characterizing better permissible guesses

Generally, a guessing scenario—understood as a set of complete answers, Q, together with a credence distribution, P, specified over those answers—can improve when the credence distribution over those answers 'improves'. And this is achieved in general when that distribution becomes more 'peaked': in that a greater weight of probability is borne by fewer complete answers.

Consider a relatively extreme situation in which almost all of the probability is concentrated on a single complete answer, with the remaining probability being distributed among the remaining complete answers. The table below illustrates one such instance for the horse-race example earlier (where the question is "*Which horse is going to win?*").

Ajax	*Benji*	*Cody*	*Dusty*	*Ember*
0.9	0.04	0.03	0.02	0.01

The guessing scenario here seems rather good. The complete answer with the highest probability ("*Ajax*") has both high accuracy and high (indeed maximal) specificity. Since in principle we are interested in maximizing both of these qualities, there's a good chance we may, to our satisfaction, achieve these aims in this scenario. In the language of our minimal formal account, the Pareto front will consist of a set of points that lie near the line $A = 1$ in objective space (so, a set of points almost parallel to the x-axis in objective space).

Now, consider a different extreme case (in the table below) in which one is almost indifferent over the complete answers so that the credence distribution over those answers is almost uniform.

Ajax	*Benji*	*Cody*	*Dusty*	*Ember*
0.19	0.21	0.21	0.19	0.2

In this second scenario, it seems as though guessing won't be as easy as in the first. That is, doing as good a job at maximizing both qualities (of accuracy and specificity) now seems more difficult. The Pareto front for this second scenario will consist of a set of points that are almost equally spaced as measured along the y-axis (the accuracy-axis) of objective space. And, generally, the Pareto front in the first case and the Pareto front in the second case will (roughly) bound possible Pareto fronts obtained using difference credence distributions over the various complete answers.

These remarks motivate our conception of better guessing scenarios and of better permissible guesses. In what follows, let a guessing scenario, G, be identified by the set of complete answers,

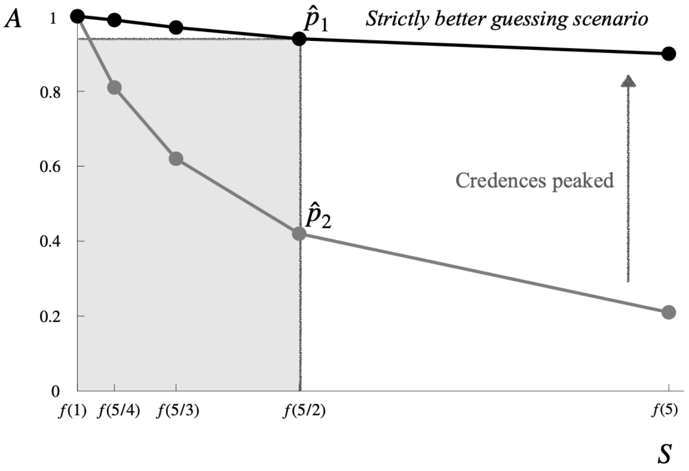

Figure 5: Pareto fronts associated with two different guessing scenarios. (These scenarios correspond to the horse-race examples presented in the two tables at the outset of this section.) The (permissible) guess corresponding to $\hat{p}_1$ ("*Ajax* or *Benji*") weakly dominates the (permissible) guess corresponding to $\hat{p}_2$ ("*Benji* or *Cody*"). Thus the former is a better permissible guess than the latter. Furthermore, the black Pareto front reflects a guessing scenario that is *strictly better* than that reflected by the gray Pareto front.

Q, and a credence distribution over those answers, P: $G = (Q, P)$. We begin by defining a *better permissible guess* as follows.

> Consider two different guessing scenarios: $G_1 = (Q, P_1)$ and $G_2 = (Q, P_2)$—where the only difference between these scenarios is that the probability distribution over complete answers is different. Let p_1 be a permissible guess in the context of G_1 (so that p_1 lies on the Pareto front of G_1), and let p_2 be a permissible guess in the context G_2 (p_2 thus lies on the Pareto front of G_2). Then p_1 is a *better permissible guess* than p_2 if and only if p_1 weakly dominates p_2.

Note that whether or not p_i is permissible in the above definition depends on whether it is permissible in the context of G_i (only). Indeed, p_2 will not be permissible relative to G_1. For the horse-race example immediately above, $p_1 \equiv$ "*Ajax* or *Benji*" is a permissible guess in the first context [with an accuracy of 0.94 and a specificity of $f(|Q|/c_{p_1}) = f(5/2)$] and $p_2 \equiv$ "*Benji* or *Cody*" is a permissible guess in the second context [with an accuracy of 0.42 and a specificity of $f(|Q|/c_{p_2}) = f(5/2)$]. However, because $p_1 \equiv$ "*Ajax* or *Benji*" (as a guess in G_1) performs better along at least one of the two qualities of interest (either accuracy and/or specificity) than $p_2 \equiv$ "*Benji* or *Cody*" (as a guess in G_2), p_1 is a *better permissible guess* than p_2. Fig. 5 illustrates the key concepts.

We can broaden this definition to one that relates to guessing scenarios as a whole.

> A guessing scenario $G_1 = (Q, P_1)$ is *strictly better* than a guessing scenario $G_2 = (Q, P_2)$, if and only if all permissible guesses for G_2 (save for the guess with an accuracy of 1) are weakly dominated by permissible guesses for G_1.

So for each permissible guess for G_2 (save for the guess with an accuracy of 1) there is a *better permissible guess* that lies on the Pareto front of G_1. In terms of trade-off lines (those lines that connect the dots of the Pareto front): the trade-off line for G_1 lies entirely above that of G_2 [save for when those two lines must meet at $(f(1), 1)$]. In the horse-race example immediately above, the first guessing scenario is *strictly better* than the second guessing scenario.

For completeness, we'll present two further definitions that relax assumptions in the above definition.

> A guessing scenario $G_1 = (Q, P_1)$ is *better* than a guessing scenario $G_2 = (Q, P_2)$ if there is at least one point on the Pareto front of G_2 that is weakly dominated by a point on the Pareto front for G_1 and no points on the Pareto front for G_1 that are weakly dominated by a point on the Pareto front for G_2.

So the trade-off line for $G_1 = (Q, P_1)$ lies above the line for $G_2 = (Q, P_2)$ for some segment and otherwise the two trade-off lines coincide. Finally, note that it is possible for there to be no such strict ordering. Then, if neither G_1 is better than G_2 nor G_2 is better than G_1, we call G_1 and G_2 *incomparable*.[17]

3.2 Generating better permissible guesses: Fermi problems

We turn to the issue of how better permissible guesses may be generated. In particular, we focus on a novel type of guessing situation that has received almost no attention in the recent literature on guessing, in which the express goal is to generate better permissible guesses and, more generally, better guessing scenarios.

The novel type of guessing situation involves *Fermi problems* [see, for an introduction, Von Baeyer (1993, Ch. 1)]. We will define these as numerical estimation problems for which: (i) the agent has little background information; and where (ii) a particular broad approach is taken to solve the problem.[18] Here are some examples: *How many penguins can you fit into your kitchen?*; *How many quarters would you have to stack to reach the Sun from the Earth?*; *What is the volume of a cat in inches cubed?* Regarding (i): as opposed to the types of scenarios described in the recent literature on guessing, a set of complete answers is not provided nor are any credences assumed. And yet there is a clear sense in which a response to such questions could be described as *guessing*. Regarding (ii): the broad approach that is used to address such problems uses only 'ad hoc knowledge'—that is, knowledge that is immediately available to the guesser (indeed one is not allowed to consult books or other external resources). And, more specifically, the approach involves a decomposition of a given problem into subproblems for which one may more credibly be able to provide guesses. The guesses to these subproblems are then combined into an overall guess for the original question.

[17] We have again presented a minimal set of definitions to capture what we believe to be salient about guessing scenarios for our purposes in this paper. But one could imagine distinguishing G_1 from G_2 even if they were incomparable by, for example, comparing the 'areas' under their Pareto fronts. But such a construction requires formalism that isn't needed for our purposes, here.

[18] They are more familiar in scientific and educational contexts. In science, they function as back-of-the-envelope consistency checks on more detailed calculations and can thus also serve to reveal a more basic (quantitative) understanding of complex scenarios. In educational contexts they are used to develop and improve intuition for quantities, scales, and the modular nature of modeling—a feature not lost on certain corners of the business world in which such problems have been used to assess job applicants' quantitative reasoning skills.

To illustrate this methodology, as well as our claim that such problems are involved in generating better permissible guesses, consider perhaps the most famous such problem:

Tuners : *How many piano tuners are there in Chicago?*

Assume your ad hoc knowledge includes the fact that there are roughly 3 million inhabitants of Chicago. How could one proceed?

An initial guess might just be based on rough orders of magnitude. You might reason that there must be at least 10—for, just one tuner would be much too low in such a big city. Tuners numbering 1,000,000 sounds like a lot, as does 100,000 tuners. To be safe (that is, to guarantee *accuracy*), 10,000 tuners might sound like a good upper bound. And thus you might conservatively guess somewhere between 10 and 10,000 tuners. Perhaps you're indifferent over this range, in which case a uniform distribution over this range could be used to represent your credences. Or perhaps your credence distribution might be a little peaked. In any case, such reasoning provides us with an initial set of *complete answers* $Q_{tuners} \equiv \{10, 11, 12, \ldots, 9999, 10000\}$, as well as a first-pass at a credence distribution over these answers. And, following the algorithm described in Sec. 2.2, one may generate permissible guesses in such a context.[19]

But, one can do better, by adopting the methodology outlined above. You might, for instance, first estimate the number of pianos in Chicago. This would provide a means to estimate the number of tunings that are required per year (say), yielding an estimate of the number of tuners that might be needed to meet the demand for tunings. Thus, one might proceed as follows.

> **Fermi solution**: Divide the assumed number of people in Chicago (call this a_1) by an estimate of the typical number of people in a household (a_2). Thus the total *number of households in Chicago* is $a_1 \times (1/a_2)$. Assume a value for the fraction of households with a piano (a_3). Thus the *number of pianos in Chicago* is $a_1 \times (1/a_2) \times a_3$. Assume a value for the average number of times a piano needs to be tuned in a year (a_4). Thus the *number of piano tunings required in Chicago in a year* is $a_1 \times (1/a_2) \times a_3 \times a_4$. Assuming further that the tunings that can be provided meet the demand for tunings, let one tuner perform a_5 tunings in a day, where that tuner works a_6 days per year. Thus the *number of tunings a tuner will complete in a year* is $a_5 \times a_6$. And so an estimate for the number of piano tuners, meeting the demand for tunings, is given the ratio of the last two estimates:
>
> $$\textit{Number of piano tuners in Chicago} = a_1 \times (1/a_2) \times a_3 \times a_4 \times (1/(a_5 \times a_6)).$$
>
> If one chooses $(a_1, a_2, a_3, a_4, a_5, a_6) = (3 \times 10^6, 4, 1/50, 3, 3, 150)$, then the number of tuners one obtains is 100.

This is a significantly more crisp estimate than the initial range represented by the complete answers in Q_{tuners}. But an agent committing to precisely 100 tuners as a result of the above reasoning is perhaps acting irrationally. For there is significant uncertainty about this value, which arises from uncertainty related to guesses for each subproblem. We will describe in more detail, shortly, how such uncertainty carries over to the final estimate but note that even if it were

[19] Admittedly, guessing in a state of perfect indifference (where one's credence distribution is exactly uniform) yields rather strange permissible guesses. For instance, one guess on the Pareto front will be "11 *or* 37 *or* 235" tuners. But such an outcome is a result of the idealization involved in the adoption of an exactly uniform distribution. More plausibly, the distribution won't be uniform in which case the permissible guesses will be less strange.

the case that the uncertainty in the final estimate was of an order of magnitude on either side, so that the agent's credence distribution could be represented as one that is uniform over $(10, 1000)$ (and zero outside of this range), this would yield a significantly better guessing scenario than the initial response outlined to **Tuners**. Indeed, such a guessing scenario contains within it, better permissible guesses as defined in Sec. 3.1.

Let us analyze, in more detail, the types of reasoning patterns that allow one (more generally) to arrive at such a better guessing scenario. Plausibly, the agent is decomposing a problem into subproblems about which the agent is less uncertain than the original problem. Now, despite the fact that agent may still not have very much directly applicable knowledge that can guide guesses about the subproblems they may still be able to provide relatively accurate guesses to the subproblems. Two patterns of reasoning that they may employ are as follows.

(1) In the **Fermi solution** to **Tuners**, there is very little about the city of Chicago itself that needs to be included to render reasonable the guess that the typical number of people in a household there is four. Perhaps knowing that Chicago is a city of a certain size with various economic advantages and a certain distribution of wealth is enough to secure that the (probability) distribution over the number of people in a household in Chicago is peaked at or near four people.

(2) For other Fermi-type guessing scenarios (such as the penguins question introduced the outset of this paper), an agent may estimate size by iterating based on a familiar reference quantity—as for when the agent may guess the volume of a room by imagining how many large moving boxes would fit into the room, as opposed to something less familiar (such as how many penguins might fit into the room).

How tight the analogy may be (in the first case) or how familiar the reference quantity may be (in the second case) dictates the degree of uncertainty that the agent may have about the estimate provided.[20] This uncertainty is naturally cashed out as a spread of possible values surrounding the estimate provided and it carries over into uncertainties about the final estimate provided for the original Fermi problem.

In more formal terms, the spread of values about any estimate provided (either for a subproblem or for the original problem) is naturally characterized by the *variance* (the square of the standard deviation) of the agent's credence distribution. And, it turns out, variances can add in a very simple way to provide a variance for the overall estimate provided for the original problem. In appendix A we describe some elementary considerations that exhibit how the variance in the overall estimate is simply a weighted sum of the variances of the estimates provided for the different subproblems (assuming the subproblems are probabilistically independent of each other—an assumption we will return to, below). Such a relationship provides a clue as to what a rational agent who successfully solves a Fermi problem may be sensitive to. In particular, if an agent can (perhaps even subconsciously) assess (or intuit) when the variance of the overall estimate—formed by summing variances for estimates for each subproblem—is significantly less than the variance associated with some initial (non-highly-peaked) credence distribution over

[20] Note that we've described two general heuristics in addressing Fermi problems—there are others that we do not have space to go into. [See, for a selection: Ärlebäck and Albarracín (2019); Albarracín and Gorgorió (2014); Dowker (1992).] In addition, the focus here on multiplicative decompositions is also a simplification that could be dropped in certain guessing scenarios. After all, natural joints or kinds (or appropriate categories for a decomposition) relate to how we view the world and this does not limit decomposition to multiplicative decompositions.

possible answers (indeed before the agent decomposes the problem), then the agent can potentially solve the problem. And such an assessment by the agent goes hand-in-hand with the assessment that *further* decomposition—so, the introduction of more subproblems about which guesses need to be made, together with their attendant uncertainties—will only significantly increase the variance of the overall estimate. In this way, an agent who successfully solves a Fermi problem is able to choose the rough size of the decomposition of the original problem into subproblems. The agent's guesses about the subproblems have a relatively low uncertainty such that the overall variance of the final estimate is as low as is necessary to provide a successful guess.[21]

As mentioned above (and described formally in the appendix) estimating the variance of the overall estimate is computationally more straightforward when the guesses to subproblems are (probabilistically) independent of each other. But, prima facie, Fermi-style decompositions can include correlations. Consider the **Fermi solution** to **Tuners**. There are plausible correlations between the various pieces of the calculation. The estimate of the size of a family (a_2) is plausibly correlated with the size of town considered (a_1). Larger cities such as Chicago attract young singles and couples who may not have children whilst people further along in their careers, with perhaps larger families, may move out of the city in search of more space. This might reduce the average number of people in a household in larger cities compared to smaller ones. What is important, however, is whether such correlations can be neglected: in the specific case above it seems reasonable that one can ignore such a correlation, for family size in the city might more strongly correlate with other factors (that may already have been taken into account by the agent) such as prevailing cultural trends. Moreover such differences can be negligible relative to the nature of the estimate required (as for when only order of magnitude estimates may be required).

To sum up, Fermi problems involve various inter-twined estimation problems. Crucially, the aim is to generate a final credence distribution that is more peaked than one with which one begins (upon hearing the question). When successfully done, the typical procedure followed in solving a Fermi problem thereby creates a *better guessing scenario* together with the possibility of *better permissible guesses*. Of course, in response to **Tuners**, after having effected the **Fermi solution** (say), an agent may still answer "anywhere between 10 and 10,000 tuners" (for they may still place a premium on being accurate); or they may more closely follow the final credence distribution (and place a higher premium on specificity): in which case they may answer "anywhere between 50 and 200". But this just reflects the general setting in which such guessing scenarios are, we believe, to be understood: as involving trade-offs between accuracy and specificity. Our analysis of Fermi problems strengthens the case for thinking about guessing as such a trade-off. And our analysis broadens the scope of applicability of the minimal formal account described above. Whether and how one might look to address such scenarios in other cases (as for if one were to employ the techniques of Dorst and Mandelkern), we'll leave for future work.

[21] A less agent-centric way of explaining the success of decompositions in solving Fermi problems proceeds via the observation that errors in guesses to different subproblems in a decomposition can cancel each other out. [See Von Baeyer (1993, Ch. 1): lecture notes by Sussman (2020) provide a model for this idea based on random walks.] We leave a more complete epistemological analysis of Fermi problems for future work.

4 Conclusion

In this paper we have provided a new way to understand the nature of guesses, namely the answers that rational agents provide to questions about which they are uncertain. The recent philosophical literature on guessing has focused on *permissible guesses*—guesses that are rational and that capture observable aspects of the practice of guessing (Dorst and Mandelkern, 2022; Skipper, 2023). Such guesses have been argued to optimize, using epistemic utility theory, a trade-off between accuracy (the quality of not being incorrect) and specificity (the quality of having ruled out possible options). We have presented an alternate, minimally formal account of permissible guesses by treating the trade-off as it is often captured in economics and in the sciences. This account is independent of those that rely on epistemic utility theory. It is parsimonious (compared to such accounts) and is plausibly consistent with more specific models (including those involving epistemic utility theory) that attempt to capture what an agent is doing when they rationally guess. Our work provides a basis for understanding rational constraints on guessing and can also account for predictions about guessing that have recently been developed.

Furthermore, we have extended the literature in describing what makes for *better permissible guesses*, as for when an agent's credence distribution over possible outcomes evolves over time, in response to further deliberation. We identified and analyzed Fermi problems—quantitative estimation problems in which precisely such an evolution in an agent's credence distribution can occur. Our analysis of Fermi problems: (i) describes how better permissible guesses can arise; (ii) strengthens the case for thinking about guessing as a trade-off between accuracy and specificity; and (iii) exhibits the scope of applicability of our minimal formal account. Our minimal formal account thus strengthens the case for understanding guessing (now, very broadly considered), in terms of a trade-off between accuracy and specificity.

A Uncertainties in Fermi estimates

To understand where uncertainties in the final estimate of a Fermi problem come from, we'll collect some basic results in this appendix.

Let the quantity to be estimated in a Fermi problem be denoted by f. For ease of exposition, we'll assume that this is a real number. (The case where the number to be estimated is an integer, for example, can be treated in a similar way.) Assume the agent follows the usual method of decomposing the guessing problem n subproblems (labeled in some order by $i = 1, 2, \ldots, n$)—and then combines guesses to those subproblems in some way (for example by multiplying them). Let the *possible* guesses for the i'th subproblem be denoted by x_i and assume that the agent has some (implicit) credence distribution over x_i. We'll denote the actual value the agent assumes for the i'th subproblem by a_i. In this way one can think of x_i as denoting (some value of) a random variable. The final estimate is then a function of these different x_i's: one can write, in general, $f \equiv f(x_1, x_2, \ldots, x_n)$, where the agent's final estimate (not taking uncertainties into account) is $f(a_1, a_2, \ldots, a_n)$.

Described in this way, the final estimate can also be thought of as a random variable which can be characterized by various statistical properties. Indeed, we're primarily interested in exploring how (relatively small) uncertainties in guesses to the subproblems carry over to uncertainties in the final estimate. We can do this by first expressing the dependence of the final estimate on small deviations in guesses to the subproblems. Formally, we can write (using elementary calculus, to 'first order'):

$$f(x_1, x_2, \dots, x_n) \approx f(a_1, a_2, \dots, a_n) + \sum_{i=1}^{n} c_i (x_i - a_i),$$

where the $c_i \equiv \partial f / \partial x_i|_{\vec{a}}$ are constants expressing the dependence of f on the guess to the i'th subproblem [here, $\vec{a} \equiv (a_1, a_2, \dots, a_n)$]. This formula tells us how the final estimate $f(a_1, a_2, \dots, a_n)$ changes if we adopt, for each subproblem, x_i instead of a_i (where x_i is 'close' to a_i).

One can show [by calculating the variance of a linear combination of random variables, see: Devore and Berk (2012, Ch. 6)] that the typical spread of values for $f(x_1, x_2, \dots, x_n)$, denoted by σ_f^2 [that is, the variance of $f(x_1, x_2, \dots, x_n)$], is related to the typical spread of values for x_i, denoted by σ_i^2, and covariances $\sigma_{i,j}$ (a measure of how strongly x_i and x_j are related), via:

$$\sigma_f^2 = \sum_{i=1}^{n} c_i^2 \sigma_i^2 + \sum_{i=1}^{n} \sum_{j \neq i} c_i c_j \sigma_{i,j}.$$

Assuming (say) probabilistic independence of the subproblems (uncorrelated subproblems will suffice), the second term on the right-hand side of the above equation vanishes and the formula simplifies dramatically:

$$\sigma_f^2 = \sum_{i=1}^{n} c_i^2 \sigma_i^2.$$

We see, then, that the variance (or the typical spread of values) for the final Fermi estimate can be very simply related (that is, linearly related) to variances for guesses to each subproblem.

It should hopefully be evident that Fermi problems that can be decomposed into (probabilistically) independent subproblems can more easily be solved than those Fermi problems for which correlations between the subproblems exist. The above analysis highlights that working with correlated subproblems requires the agent to not only intuitively keep track of the size of uncertainties for individual subproblems but also to keep track of uncertainties that arise as a result of 'cross-talk' (encoded in the covariances) between the different subproblems. Working with (say) independent subproblems requires keeping track of uncertainties over the individual subproblems while allowing one to ignore such complicating cross-talk.

Acknowledgements

We thank Paddy Blanchette and Nic Teh for helpful discussions. N. L. acknowledges the support of the Swiss National Science Foundation as part of the project 'Philosophy Beyond Standard Physics' (105212_207951). F. A. acknowledges support from the Faculty Research Support Program (FY20FRSPE) at the University of Notre Dame.

References

Albarracín, L. and Gorgorió, N. (2014). Devising a plan to solve Fermi problems involving large numbers. *Educational studies in mathematics* **86**, 79–96.

Ärlebäck, J. B. and Albarracín, L. (2019). The use and potential of Fermi problems in the STEM disciplines to support the development of twenty-first century competencies. *ZDM Mathematics Education* **51**, 979–990.

Carter, J. A. (2020). Sosa on knowledge, judgment and guessing. *Synthese* **197**, 5117–5136.

Devore, J. L. and Berk, K. N. (2012). *Modern Mathematical Statistics with Applications*. Second Edition. New York: Springer.

Dorst, K. and Mandelkern, M. (2022). Good guesses. *Philosophy and Phenomenological Research* **105**, 581–618.

Dowker, A. (1992). Computational estimation strategies of professional mathematicians. *Journal for Research in Mathematics Education*. **23**, 45–55.

Giagkiozis, I. and Fleming, P. J. (2014). Pareto front estimation for decision making. *Evolutionary Computation* **22**, 651–678.

Holguín, B. (2022). Thinking, guessing, and believing. *Philosophers' Imprint* **22**(6).

Kahneman, D. and Tversky, A. (1982). Variants of uncertainty. *Cognition* **11**, 143–157.

Levi, I. (1973). *Gambling with Truth: An Essay on Induction and the Aims of Science*. Cambridge: MIT Press.

Skipper, M. (2023). Good guesses as accuracy-specificity tradeoffs. *Philosophical Studies* **180**, 2025–2050.

Sussman, D. M. (2020). How things work. Lecture notes for *PHYS121* (Emory University). Accessed: February 6, 2024.

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

Von Baeyer, H. C. (1993). *The Fermi Solution: Essays on Science*. New York: Dover.