Title
Overconfidence as Truth Approximation

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 47(0)

Authors
Cevolani, Gustavo
Coraci, Davide

Publication Date
2025

Overconfidence as Truth Approximation

Gustavo Cevolani (gustavo.cevolani@imtlucca.it)
IMT School for Advanced Studies Lucca
Lucca, Italy

Davide Coraci (davide.coraci@imtlucca.it)
IMT School for Advanced Studies Lucca
Lucca, Italy

Abstract

Human reasoning and decision-making under uncertainty often deviate from normative standards of rationality. Over the past decades, cognitive scientists have extensively investigated heuristics and cognitive biases, such as overconfidence—the tendency to overestimate the probability that one's judgments are correct. Meanwhile, philosophers have explored different "cognitive utilities" that guide both scientific and everyday reasoning, including the concept of truthlikeness, i.e., how well an hypothesis, be it a statement or a numerical interval, approximates the whole truth about a target domain. In this paper, we integrate empirical findings with philosophical perspectives, showing how formal models of truthlikeness offer valuable insights for empirical research on overconfidence. In particular, by conceptualizing overconfidence through the lens of expected truthlikeness maximization, we argue that many instances of this phenomenon may be construed not as cognitive biases, but rather as rational strategies for approaching the truth under conditions of uncertainty.

Keywords: Overconfidence; Overprecision; Overestimation; Cognitive Bias; Miscalibration; Interval Estimation; Truthlikeness; Verisimilitude; Closeness to the Truth.

Introduction

Human reasoning and decision-making under uncertainty often deviate from normative standards of rationality. Over the past decades, cognitive scientists have investigated various heuristics and cognitive biases of human reasoning, documenting how they depart from theoretical prescriptions (Kahneman, 2011; Tversky & Kahneman, 1983; Gigerenzer, 2015). A widely studied bias is overconfidence in judgments (Lichtenstein, Fischhoff, & Phillips, 1977; Russo, Schoemaker, et al., 1992; Moore & Healy, 2008), occurring when people systematically overestimate the probability that their own judgments are approximately correct.

Independently, philosophers have examined the "cognitive" or "epistemic" utilities—such as probability, accuracy, confirmation, explanatory power, and others—that guide reasoning and cognitive decision-making in both scientific and everyday contexts (Niiniluoto, 1987; Sprenger & Hartmann, 2019; Pettigrew, 2016). Among them, verisimilitude or truthlikeness (TL), as originally introduced by Popper (1963), refers to how well an hypothesis, be it a statement or a numerical interval, approaches the truth about the target domain.

In this paper, we integrate empirical findings with philosophical perspectives, arguing that formal models of truthlikeness can provide important insights into the study of overconfidence and miscalibration in judgment. In particular, we show that, when overconfidence is conceptualized through the lens of expected TL maximization, many instances of this phenomenon may be construed not as cognitive biases, but rather as rational strategies for approaching the truth under conditions of uncertainty.

The paper proceeds as follows. First, we review some results concerning overconfidence and miscalibration in the psychology of reasoning and discuss some interesting theoretical models proposed by psychologists to account for such phenomenon (Yaniv & Foster, 1995, 1997). Second, we introduce the notion of TL, as developed in the philosophy of science literature (Oddie & Cevolani, 2022), focusing in particular on Oddie's (1986) measure. Third, we assess available experimental evidence through formal models of expected TL. We show how these models combines probable truth and informative content and precisely explain the accuracy/information trade-off underlying the phenomenon of overconfidence (Yaniv & Foster, 1995, 1997; Teigen, 1990). To further corroborate this theoretical interpretation of overconfidence, we use computer-based simulations to predict empirical assessments that can be compared with empirical data about overconfidence in interval estimation.

On the basis of our results, we argue that overconfidence, when conceptualized through the notion of TL, is not necessarily a cognitive bias, but can be seen, in a more positive light, as a rational strategy to approximate the truth under uncertainty. Finally, we conclude with some suggestions for future work, in particular on how to empirically test our TL-based model of overconfidence.

Overconfidence and the Preciseness Paradox

Overconfidence, broadly defined as people's tendency to overestimate the confidence in the correctness of their judgments, emerges from miscalibration in judgments. An "epistemic agent" is perfectly calibrated when its degree of confidence in a certain judgment, i.e., how probable the agent thinks the judgment is true, matches the degree of accuracy of the judgment. For instance, if an agent is 90% sure of its performance on a quiz and answers correctly 9 out of 10 questions, the agent is perfectly calibrated. On the contrary, a mismatch between confidence and accuracy leads to miscalibration, implying either overconfidence (when confidence exceeds accuracy) or underconfidence (when confidence undershoots accuracy). For instance, if the agent is 90% sure of

In D. Barner, N.R. Bramley, A. Ruggeri and C.M. Walker (Eds.), *Proceedings of the 47th Annual Conference of the Cognitive Science Society*

its performance but answers correctly only 6 out of 10 questions in the quiz (60% of accuracy), he is overconfident.

Miscalibration leading to overconfidence is a well-known and studied bias in cognitive sciences and psychology (Lichtenstein et al., 1977; Alpert & Raiffa, 1982; Russo et al., 1992; Liberman & Tversky, 1993; Baranski & Petrusic, 1994; Harvey, 1997; Keren & Teigen, 2001; Teigen & Jørgensen, 2005). People show a systematic tendency in overestimating their ability to provide accurate judgments. This phenomenon, cited as the most significant among cognitive biases (Kahneman, 2011), is often related to real-world issues, such as financial crises and corporate failures (Malmendier & Tate, 2005; Ho, Huang, Lin, & Yen, 2016; Leng, Ozkan, Ozkan, & Trzeciakiewicz, 2021).

While behavioral experiments clearly show the pervasiveness of overconfidence, the theoretical interpretation of this phenomenon is less clear, and no agreement exists on its main determinants. In part, this is due to a lack of conceptual clarity, since overconfidence appears to be an umbrella term covering quite different constructs and experimental paradigms. Following Moore and Healy (2008), it is worth distinguishing at least two constructs of overconfidence relevant for the current paper, i.e., *overestimation* and *overprecision*.[1]

Overestimation consists in thinking that you are better than you are, being then related to one's actual ability and the chance of success in a task, like the knowledge quiz in the example above. Overprecision instead concerns the excessive faith that you know the truth, and typically occurs in cases of miscalibrated interval estimations. Suppose, for instance, that you are presented with a question such as: "How long is the Nile River?". You are asked to provide an interval estimate (range) with a "confidence interval" $CI = 90\%$, that is, to provide a low and a high estimate such that you are 90% certain the correct answer will fall within these limits. If the interval estimate you provide is too narrow (and hence corresponds to a CI lower than 90%), you are overprecise.

Experimental research is consistent in demonstrating that people prefer estimates corresponding to CIs around 60%, or even lower (e.g., 20% and 40%), even when asked to provide estimates with higher CIs, like 90% or more (Lichtenstein et al., 1977; Russo et al., 1992; Teigen & Jørgensen, 2005; Keren & Teigen, 2001). This robust finding motivated a systematic investigation of overprecision in interval estimation. The systematic tendency of preferring precise estimates, at the cost of reducing their probability of being correct has been dubbed the "preciseness paradox" by Teigen (1990). Indeed, even though the most precise of two estimates has necessarily, for purely logical reasons, a smaller likelihood of being true, it has typically greater chances of being preferred by human judges. In the most advanced investigation of overprecision (Yaniv & Foster, 1995, 1997), participants were directly presented with the true value of some quantity, thereby eliminating any possible uncertainty. When asked to choose between two alternative intervals representing possible answers to the question about the relevant true value, participants tended to still prefer narrower intervals, even if they didn't contain the target value (see below for an example).

Different models have been proposed to account for the data on overprecision (Tversky & Kahneman, 1974; Gigerenzer, Hoffrage, & Kleinbölting, 1991; Yates, 1994; Tversky, 1969; Juslin, Winman, & Hansson, 2007).[2] Some point to a possible trade-off between probability of truth and informativeness as the main determinant of the effect: participants may value the greater information provided by an interval estimate at the expense of its lower probability of being true, i.e., containing the relevant true value. In particular, the so-called "additive model" proposed by Yaniv and Foster (1995) assumes that, when asked to provide or assess an interval estimate, people tend to balance two, potentially conflicting, dimensions: the informativeness of the estimate, i.e., the coarseness of the interval, roughly measured as its width, and its accuracy, i.e., the distance of the interval from the true value as measured by the distance of its midpoint.

Interestingly, their idea is quite close to the ones that independently motivated the development of models of verisimilitude or truthlikeness by philosophers of science, to which we now turn.

Truthlikeness and expected truthlikeness

In scientific and ordinary contexts, truth is often assumed to be the aim of rational inquiry. However, some truths are better than some other truths; and sometimes even falsehoods may realize the aim of inquiry better than some truths do. For instance, suppose to compare two different estimates, A and B, of the height of Mont Blanc, knowing that its actual height is 4 808 m. Suppose further that A states that the Mont Blanc is between 2 000 m and 10 000 m high, while B says that the Mont Blanc is between 4 000 m and 4 500 m high. Note that A is a true estimate, since it contains the true value, whereas B is false, since it excludes it. Still, B may appear as clearly a better estimate than A in approximating the height of the Mont Blanc. The reason is that B, although false, is much more informative than the rather vague estimate provided in A, and still quite close to the truth (even if not as close as A).

This example shows that truth is a too coarse-grained property to evaluate the epistemic status of propositions and scientific hypotheses. A more fine-grained notion is needed to assess the relative closeness to the truth of such hypotheses and propositions. Starting from Popper (1963), philosophers of science have then tackled this issue by defining formal notions of truthlikeness or verisimilitude, construed as closeness to the whole truth about a target domain (Popper, 1963; Oddie, 1986; Niiniluoto, 1987; Oddie & Cevolani, 2022). In the philosophical jargon, we say that estimate B in the exam-

[1] Moore and Healy (2008) also analyze a third construct, *overplacement*, defined as an exaggeration of the degree to which one's judgments are better—more accurate—than others'. The discussion of overplacement is beyond the scope of the present work.

[2] See Moore and Healy (2008); Moore, Tenney, and Haran (2015) for a review of the different approaches proposed to explain overconfidence and related phenomena.

ple above is false but closer to the truth, more verisimilar, or more truthlike, than estimate A. The "logical" problem of TL amounts to rigorously define what truthlikeness is: given a certain truth t, tell when a theory (proposition, hypothesis, interval) is closer to the truth than another one. The "epistemic" problem of TL consists in assessing when a theory (proposition, hypothesis, interval) is estimated as closer to the truth than another without knowing the actual truth, but possibly some (true) evidence about the relevant domain.

The philosophical literature has developed different formal models to approach these issues (Oddie, 1986; Niiniluoto, 1987; Schurz & Weingartner, 2010; Cevolani & Festa, 2021; Kuipers, 2024).[3] All such models aim at providing adequate measures (or comparative orderings) of the closeness of an hypothesis to the truth. Such measures can be construed as striking a balance between the truth and the informative content of the hypothesis; in this way, they provide a way of assessing, in a principled and independently motivated way, the trade-off between accuracy and informativeness studied by psychologists.

For the sake of comparison, in the following we shall focus on cases where hypotheses are expressed as finite, closed intervals I of points of a discrete "universe" U, and "the truth" is one of this point. As an example, U may be the line of natural numbers from 1 to 10000 expressing the heights of mountains in meters; $t = 4808$ may be the true value of some relevant quantity like the height of Mont Blanc; and $A = [2000, 10000]$ and $B = [4000, 4500]$ may be two hypotheses about t. A definition of the truthlikeness of such hypotheses amounts to a measure of the distance of the corresponding intervals from t. For normalization purposes, let's assume the absolute difference between two points a and b of U, divided by the cardinality $card(U)$ of U, as the measure of their distance:[4]

$$\delta(a,b) = \frac{|a-b|}{card(U)} \quad (1)$$

One natural way to extend measure δ to a measure Δ of the distance of an interval I from a point a is taking the average of the distances from a of the points in I:

$$\Delta(I,a) = \frac{1}{card(I)} \times \sum_{x \in I} \delta(x,a) \quad (2)$$

where $card(I)$ is the number of points in I. Note that measure Δ expresses a trade-off between the accuracy of I (depending on the distance δ of its points from a) and the informativeness of I (as measured by its cardinality or width). In fact, I could be very accurate (i.e., overall close to the truth) but, at the same time, poorly informative (i.e., "too wide"); or, vice versa, it could be very informative (i.e., "narrow"), but inaccurate (i.e., far from the truth).

[3]For surveys, see Niiniluoto (1998, 2020); Oddie and Cevolani (2022).

[4]This particular measure is chosen for the sake of simplicity. Other possible measures, not referring to the cardinality of U, are discussed for instance by Kuipers (2024) and Niiniluoto (1987).

The truthlikeness of I is then defined as the complement of its distance from the truth, as follows:

$$\begin{aligned} TL(I) &= 1 - \Delta(I,t) \\ &= 1 - \tfrac{1}{card(I)} \textstyle\sum_{x \in I} \delta(x,t) \end{aligned} \quad (3)$$

This is known as the (Tichý-Oddie) average measure of truthlikeness. Again, note that $TL(I)$ depends both on how close to the truth I is (in terms of δ) and how informative I is (in terms of $card(I)$). In this precise sense, truthlikeness is a trade-off between truth (accuracy) and information content.

Now suppose that the truth t is unknown. Suppose further that an epistemic probability distribution p is defined over U (expressing the degrees of belief of a rational agent about the truth in U), such that, for each $x \in U$, $p(x)$ expresses the probability that point x is the true value (possibly given some evidence, in which case we use conditional probability). The expected truthlikeness of I can then be defined as (Oddie, 1986; Niiniluoto, 1987):

$$ETL(I) = \sum_{x \in U} TL(I|x) \times p(x) \quad (4)$$

i.e., as the expected value of $TL(I)$, i.e., the (weighted) sum of the truthlikeness degree of I relative to each possible point, weighted by the probability that the point represents the truth. In the overconfidence literature, the "confidence interval" CI associated with I is the probability that I contains the truth t, i.e., it is simply defined as the proportion $card(I)/card(U)$. Interestingly, CI and ETL are independent notions: a narrow interval is bound to have a low CI but may have high ETL if all its points are expected to be close to the truth. Vice versa, a wide interval with high CI could have a low degree of ETL (see below for examples).

From a philosophical point of view, measures TL and ETL, as defined above, solve the logical and epistemic problem of truthlikeness, respectively. $TL(I)$ expresses how close to the truth t an interval is, assuming t is known; $ETL(I)$ represents the best estimate, given the available evidence E, a rational agent can give of the unknown, actual truthlikeness of I when t is unknown. As we shall see in a moment, such measures may also account for the empirical data on overconfidence produced by psychologists.

Modeling overconfidence as (expected) truthlikeness maximization

Let's come back to the problem of overprecision as defined above. The two main empirical findings are as follows. First, when participants are asked which, between two intervals A and B, is a better estimate of the known, true value t of some quantity, they tend to prefer the narrower of the two intervals, even if it doesn't contain t (recall the Mont Blanc example above). This is the preciseness paradox. Second, when participants are asked to give, in the form of an interval, their best estimate of the unknown value t at some confidence level CI, they tend to answer with an interval which is "too narrow", i.e., corresponding to a probability of being correct much

Table 1: Comparisons of truthlikeness for some interval estimates from Yaniv & Foster (1995). The star (∗) identifies the interval preferred by participants.

(1.a)		(1.b)		(1.c)	
$U=[1,100]$		$U=[1800,1950]$		$U=[500,900]$	
$t=22.5$		$t=1894$		$t=713$	
Intervals	TL	Intervals	TL	Intervals	TL
$[20,40]$	0.92	$[1870,1890]$	0.90	$[800,850]$	0.72
$[18,20]*$	0.96	$[1875,1925]*$	0.91	$[600,800]*$	0.87

lower than CI. This is the overprecision bias in its standard form.

From the present perspective, the tasks participants face in these two cases are, respectively, assessing the relative truthlikeness of the given intervals and providing their best estimate in terms of expected truthlikeness. In the following, we discuss these two cases in turn. The underlying assumption is that participants are behaving as if they were ideal, rational agents who choose the most truthlike hypothesis, when the truth is known, or maximize expected truthlikeness when it is unknown.

Truthlikeness and the preciseness paradox

Let consider the following trial of an experimental session (Yaniv & Foster, 1995). Participants are asked to choose between two estimates of the amount of money spent on education by the US federal government in 1987. The first estimate corresponds to the interval A "20 to 40 billion," the second to the narrower interval B "18 to 20 billion." Participants know the true answer, which is $t = 22.5$ billions. Experimenters found that around 80% of participants preferred B over A as an estimate of t, despite B being false and A true. How can we make sense of this preference?

A natural answer is provided by truthlikeness. While false, B is much more informative and precise than A, and still it is reasonably close to the truth, in the sense that both of its extremes are quite close to t. On the contrary, even if A is true, it is quite vague and uninformative; moreover, it contains point estimates which are very far from the truth. In short, B can be well estimated as more truthlike than A. Indeed, this is what our formal model delivers. Consider the two intervals $A = [20,40]$ and $B = [18,20]$ relative to some sensible range of values providing the relevant universe such as $U = [1,100]$. Then, if the truth is $t = 22.5$, by applying eq. 3 we obtain that $TL(B) = 0.96 > 0.92 = TL(A)$. Further examples are provided in Table 1, showing how TL can rationalize participants' choices by capturing the accuracy-information tradeoff in judgmental estimation. Note that narrower, i.e., more informative intervals are preferred to wider and less informative ones if the former are sufficiently close to the truth (as in case 1.a), but not otherwise (as in cases 1.b and 1.c)

Table 2: Summary results of the simulations: expected truthlikeness and confidence levels of some intervals in the universe $U = [1,100]$.

	Uniform distr.		Normal distr.	
	Interval	ETL	Interval	ETL
ETL_{max}	[51]	0.75	[30]	0.99
ETL_{median}	[9,80]	0.7	[4,85]	0.77
ETL_{min}	[1]	0.5	[100]	0.3
CI=20%	[41,60]	0.75	[21,40]	0.95
CI=40%	[31,70]	0.74	[10,49]	0.9
CI=60%	[21,80]	0.72	[1,60]	0.85
CI=90%	[6,95]	0.68	[1,90]	0.75
CI=100%	[1,100]	0.67	[1,100]	0.71

Overconfidence as expected truthlikeness maximization

Typical experiments on overprecision require participants to provide interval estimates with a given degree of confidence CI specified by the experimenter (Russo et al., 1992; Keren & Teigen, 2001; Teigen & Jørgensen, 2005). Differently from the previous case borrowed from Yaniv and Foster (1995), in such studies the truth is unknown. Participants have to estimate it by providing the upper and the lower bound of an interval such that they are, say, 90% certain the truth falls within the interval.

As said above, participants tend to provide too narrow intervals, having a CI much lower than the one required by the experimenter. For instance, Russo et al. (1992) find that participants provide interval ranges corresponding to confidence levels of 60%, 40%, and even 20%, despite being instructed to respond based on confidence intervals of 90% or 95%. While such answers are usually interpreted as pointing toward an overprecision bias, they are to be expected if one assumes that participants are trying to maximize expected truthlikeness.

To better explore this view of overprecision based on expected truthlikeness maximization, we conducted a computer-based simulation.[5] In the simulation, we assume a fixed, abstract universe $U = [1,100]$. Relative to this U, we generated all possible intervals, representing possible participants' answers, including the 100 "degenerate" intervals amounting to point estimates, like for instance [42]. This produced a total of 5 052 intervals. For each of these intervals I, we computed its degree of expected truthlikeness $ETL(I)$ by applying Eq. 4, as well as its CI in percentage. For the former, we need to define a probability distribution on U. For the sake of illustration, we considered two different probability distributions: a uniform distribution $p(x) = 1/card(U)$ for all $x \in U$, and a normal distribution with mean $\mu = 30$ and fixed standard deviation $SD = 1$.

Summary results of the simulation are reported in Table 2. The general message is clear. In order to maximize expected

[5] Available at: osf.io/Over_Truth_CogSci2025.

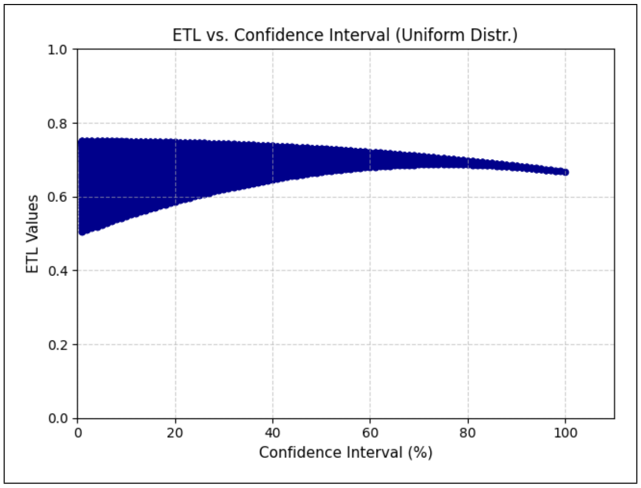

Figure 1: Scatter plot of the ETL values of intervals (for uniform probability distribution only), grouped by CI.

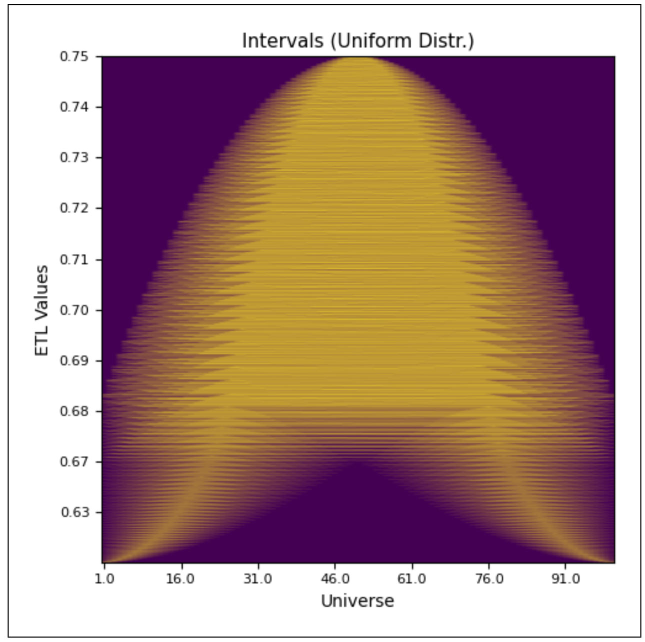

Figure 2: ETL values of all the intervals in U, represented as yellow horizontal segments (the width corresponds to the CI). A uniform probability distribution is assumed.

truthlikeness, one needs to choose comparatively narrower intervals, or even degenerate ones. Of course, such intervals will have lower confidence levels. For example, in the case of an uniform probability distribution, the degenerate interval $[51]$ ($CI = 1\%$) has the maximum expected truthlikeness of 0.75, a value equal or very close to that of wider intervals, such as $[41,60]$ ($CI = 20\%$) and $[31,70]$ ($CI = 40\%$), which respectively have ETL of 0.75 and 0.74. On the contrary, to meet the expectations of experimenters, one would need to sacrifice expected truthlikeness: indeed, wider intervals like $[6,95]$ ($CI = 90\%$), have lower ETL, i.e., 0.68. Moreover, it is worth noticing that intervals with higher CIs show a small variance in ETL, meaning that the ETL values associated with less informative intervals are approximately constant (see Figure 1).

With an uniform probability distribution—corresponding to a state of complete ignorance, on the side of the agent, about the likely location of the truth in U—values of ETL over the various intervals of the universe don't differ so much. The trade-off between expected truthlikeness and confidence levels is better appreciated when considering a normal probability distribution with $\mu = 30$ and SD=1. Again, narrower intervals are favored in terms of ETL, at the price of lower levels of CI. The "best" interval, having an ETL value of 0.99, close to the maximum, is the degenerated one corresponding to the mean of the distribution ($[30]$, $CI = 1\%$). Similarly high ETL values are observed for intervals around μ with CIs of 5%, such as $[29,33]$ (ETL=0.98). Wider intervals have higher CIs but lower values of ETL: for instance $[21,40]$ ($CI = 20\%$) has ETL 0.95, $[10,49]$ ($CI = 40\%$) has ETL 0.9, and intervals like $[1,90]$ ($CI = 90\%$) or similar have ETL around 0.75.

Figures 2 and 3 display the ETL values of all intervals in U. The trade-off between expected truthlikeness and confidence is clear, since only narrow intervals have high ETL values. Wider intervals, corresponding to higher CI levels, tend to have lower ETL. However, narrow intervals in the tails of the distributions have both low CI and ETL values.

This is in line with the theoretical interpretation proposed here. Since truthlikeness is a mixture of accuracy and informativeness, only relatively narrow, i.e., highly informative, intervals can have a high degree of ETL. However, this is not enough. To maximize expected truthlikeness, an interval needs not only to be informative, but also close to the probable truth. This explains why, in Figure 3, narrow intervals centered on the mean of the probability distribution maximize expected truthlikeness. On the contrary, narrow intervals centered on values (like 100) very far from the most probable truth minimize expected truthlikeness.[6] As for confidence levels, these are maximized by very wide and uninformative intervals, which however are bound to have average values of ETL. Again, this is as expected: as Popper (1963) pointed out long ago, maximizing the probability of being correct, i.e., of hitting on the truth, cannot be the only guide in our cognitive choices. We aim not just at truths, but at informative truths, and these are bound to have a low probability. By preferring narrow intervals probably close to the truth, experimental participants behave in a way consistent with Popper's intuition.

[6]The same happens in Figure 2. Here, however, since no cues is available about the probable location of the truth, ETL is maximized by simply choosing narrow intervals centered in the middle point (50), thus minimizing the chances of hitting on values very far from one of the extreme values in U.

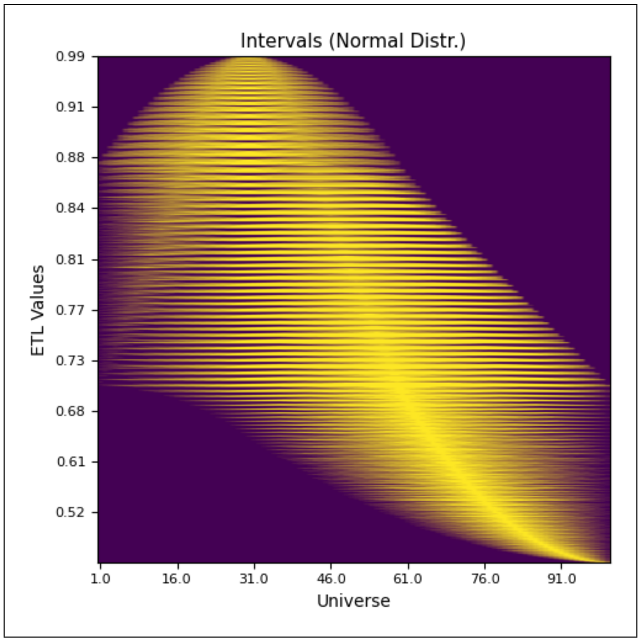

Figure 3: ETL values of all the intervals in U, represented as yellow horizontal segments (the width corresponds to the CI). Normal probability distribution with $\mu = 30$ and SD=1 is assumed.

Discussion and conclusion

In this paper, we proposed a novel theoretical interpretation of overconfidence (and overprecision in particular) based on the notion of (expected) truthlikeness as studied in the philosophy of science. Overprecision refers to the human tendency to estimate unknown quantities with interval judgments which are too precise (narrow) to be confidently accepted with high probability of being correct. This robust phenomenon is usually interpreted as a bias, showing people's overconfidence in assessing their own ability to provide accurate judgments. We suggested a different interpretation. According to truthlikeness theory, for purely cognitive reasons, a rational agent should prefer hypotheses and judgments that are both informative and close to the truth. In the present case, this translates in intervals that are narrow (and hence have a low probability of being true) and centered on the most probable truth. The proposed interpretation is in line with previous models like the one proposed by (Yaniv & Foster, 1995, 1997), but it is normatively motivated and non-*ad hoc*. Moreover, it can in principle account for the available evidence about people's preference for precision in judgment, as shown by the computer simulations above.

To be sure, much remains to be done to further test our interpretation and to explore its limits and prospects. First, one may explore how our interpretation deals with other cases of overconfidence besides overprecision of interval judgments, i.e., overestimation and overplacement. Second, further work needs to be done at the theoretical level, to study the relations between truthlikeness models (besides Tichý-Oddie's average measure) and the various models proposed by psychologists.

Third, our truthlikeness-based model should be empirically tested. In this regard, our model may offer new, fruitful ideas to enrich the standard experimental design of studies on overconfidence, along three main directions. One concerns the possibility of studying overprecision not just in the case of numerical intervals, as currently done, but also in the propositional case, since truthlikeness theory rigorously defines the information/accuracy trade-off for both intervals and statements, offering also a way to translate between the two cases. Another advantage of our model is making possible precise normative predictions about more pairs of intervals or statements than those currently studied (cf. Table 1). In particular, the theory allows to unambiguously rank all possible combinations of any kind of hypotheses, like true and informative, true and uninformative, false and informative, false and uninformative ones, in terms of their informativeness and of their overall closeness to the truth. Lastly, the distinction between the logical and the epistemic problem of truthlikeness offers a more rigorous way to explore the relative role of factors relative to the logical features of competing hypotheses (i.e., information content and closeness to the truth) and of those concerning their cognitive assessment (in terms of probability and confidence). Overall, our TL-based model of overconfidence promises to significantly enrich our way of studying the information/accuracy trade-off in human reasoning under uncertainty.

As a last comment on possible future work, if our model is found to be promising both at the empirical and at the theoretical level, we aim to explore further purported biases in human reasoning under uncertainty, like for instance the conjunction fallacy (Tversky & Kahneman, 1983; Festa, Cevolani, & Crupi, 2011; Cevolani & Crupi, 2022), the inclusion fallacy (Osherson, Smith, Wilkie, Lopez, & Shafir, 1990; Sloman & Lagnado, 2005), and related empirical phenomena which appear to crucially depend on the information/accuracy trade-off. These are all venues for future research.

Acknowledgments

The authors acknowledge financial support from the Italian Ministry of University and Research (MUR) through the PRIN 2022 project "Reasoning with hypotheses: Integrating logical, probabilistic, and experimental perspectives" (grant n. 2022ARRY9N; CUP: D53D23009570006), funded by the European Union (Next Generation EU PNRR M4C2 I.1.). We thank five anonymous reviewers for the CogSci 2025 Conference for very detailed and helpful comments on the original version of the paper.

References

Alpert, M., & Raiffa, H. (1982). A progress report on the training of probability assessors. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 294–305). Cambridge University Press.

Baranski, J. V., & Petrusic, W. M. (1994). The calibration and resolution of confidence in perceptual judgments. *Perception & psychophysics*, *55*(4), 412–428.

Cevolani, G., & Crupi, V. (2022). Truth, probability, and evidence in judicial reasoning: The case of the conjunction fallacy. In *Judicial decision-making: Integrating empirical and theoretical perspectives* (pp. 105–121). Springer.

Cevolani, G., & Festa, R. (2021). Approaching deterministic and probabilistic truth: a unified account. *Synthese*, *199*(3), 11465–11489.

Festa, R., Cevolani, G., & Crupi, V. (2011). The whole truth about linda: Probability, verisimilitude and a paradox of conjunction. In *New essays in logic and philosophy of science* (pp. 603–615). College Publications.

Gigerenzer, G. (2015). *Simply rational: Decision making in the real world*. Oxford University Press.

Gigerenzer, G., Hoffrage, U., & Kleinbölting, H. (1991). Probabilistic mental models: a brunswikian theory of confidence. *Psychological review*, *98*(4), 506.

Harvey, N. (1997). Confidence in judgment. *Trends in cognitive sciences*, *1*(2), 78–82.

Ho, P.-H., Huang, C.-W., Lin, C.-Y., & Yen, J.-F. (2016). Ceo overconfidence and financial crisis: Evidence from bank lending and leverage. *Journal of Financial Economics*, *120*(1), 194–209.

Juslin, P., Winman, A., & Hansson, P. (2007). The naïve intuitive statistician: A naïve sampling model of intuitive confidence intervals. *Psychological review*, *114*(3), 678.

Kahneman, D. (2011). Thinking, fast and slow. *Farrar, Straus and Giroux*.

Keren, G., & Teigen, K. H. (2001). Why is p=. 90 better than p=. 70? preference for definitive predictions by lay consumers of probability judgments. *Psychonomic Bulletin & Review*, *8*(2), 191–202.

Kuipers, T. A. (2024). Truthlikeness and the number of planets. *Journal of Philosophical Logic*, *53*(2), 493–520.

Leng, J., Ozkan, A., Ozkan, N., & Trzeciakiewicz, A. (2021). Ceo overconfidence and the probability of corporate failure: evidence from the united kingdom. *The European Journal of Finance*, *27*(12), 1210–1234.

Liberman, V., & Tversky, A. (1993). On the evaluation of probability judgments: Calibration, resolution, and monotonicity. *Psychological Bulletin*, *114*(1), 162–173.

Lichtenstein, S., Fischhoff, B., & Phillips, L. D. (1977). Calibration of probabilities: The state of the art. In *Decision making and change in human affairs: Proceedings of the fifth research conference on subjective probability, utility, and decision making* (pp. 275–324).

Malmendier, U., & Tate, G. (2005). Ceo overconfidence and corporate investment. *The journal of finance*, *60*(6), 2661–2700.

Moore, D. A., & Healy, P. J. (2008). The trouble with overconfidence. *Psychological review*, *115*(2), 502–517.

Moore, D. A., Tenney, E. R., & Haran, U. (2015). Overprecision in judgment. *The Wiley Blackwell handbook of judgment and decision making*, *2*, 182–209.

Niiniluoto, I. (1987). *Truthlikeness* (Vol. 185). Springer Science & Business Media.

Niiniluoto, I. (1998). Verisimilitude: The third period. *The British Journal for the Philosophy of Science*, *49*(1), 1–29.

Niiniluoto, I. (2020). Truthlikeness: old and new debates. *Synthese*, *197*(4), 1581–1599.

Oddie, G. (1986). *Likeness to truth*. Springer Dordrecht.

Oddie, G., & Cevolani, G. (2022). Truthlikeness. In E. N. Zalta & U. Nodelman (Eds.), *The Stanford encyclopedia of philosophy* (Winter 2022 ed.). Metaphysics Research Lab, Stanford University.

Osherson, D. N., Smith, E. E., Wilkie, O., Lopez, A., & Shafir, E. (1990). Category-based induction. *Psychological review*, *97*(2), 185–200.

Pettigrew, R. (2016). *Accuracy and the laws of credence*. Oxford University Press.

Popper, K. R. (1963). *Conjectures and refutations*. Basic Books Publishers.

Russo, J. E., Schoemaker, P. J., et al. (1992). Managing overconfidence. *Sloan management review*, *33*(2), 7–17.

Schurz, G., & Weingartner, P. (2010). Zwart and Franssen's impossibility theorem holds for possible-world-accounts but not for consequence-accounts to verisimilitude. *Synthese*, *172*(3), 415–436.

Sloman, S. A., & Lagnado, D. (2005). The problem of induction. In *The cambridge handbook of thinking and reasoning* (pp. 95–116).

Sprenger, J., & Hartmann, S. (2019). *Bayesian philosophy of science*. Oxford University Press.

Teigen, K. H. (1990). To be convincing or to be right: A question of preciseness. *Lines of thinking: Reflections on the psychology of insight*, *1*, 299–313.

Teigen, K. H., & Jørgensen, M. (2005). When 90% confidence intervals are 50% certain: On the credibility of credible intervals. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, *19*(4), 455–475.

Tversky, A. (1969). Intransitivity of preferences. *Psychological review*, *76*(1), 31–48.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases: Biases in judgments reveal some heuristics of thinking under uncertainty. *Science*, *185*(4157), 1124–1131.

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

Yaniv, I., & Foster, D. P. (1995). Graininess of judgment under uncertainty: An accuracy-informativeness trade-off. *Journal of Experimental Psychology: General*, *124*(4), 424–432.

Yaniv, I., & Foster, D. P. (1997). Precision and accuracy of judgmental estimation. *Journal of behavioral decision making*, *10*(1), 21–32.

Yates, J. F. (1994). Subjective probability accuracy analysis. *Subjective probability*, 381–410.