On the Probabilistic Convention T

Hannes Leitgeb

Abstract

We introduce an epistemic theory of truth according to which the same rational degree of belief is assigned to $Tr(\neg \alpha)$ and $\alpha$. It is shown that if epistemic probability measures are only demanded to be finitely additive (but not necessarily $\sigma$-additive), then such a theory is consistent even for object languages that contain their own truth predicate. As the proof of this result indicates, the theory can also be interpreted as deriving from a quantitative version of the Revision Theory of Truth.

Keywords: Probability, Truth, Liar Paradox, Revision Theory of Truth.

Rational believers are rational concerning what they believe, and thus also rational concerning what they believe to be true. But what should one rationally believe in that respect? Tarski’s[11] classical results on truth entail that we cannot take for certain, i.e., assign a rational degree of belief of 1 to, all sentences of the form

$Tr(\neg \alpha) \leftrightarrow \alpha$

where $Tr$ is the truth predicate and $\alpha$ is an arbitrary sentence of a sufficiently expressive language that includes $Tr$ as a predicate. For otherwise one would have to be certain of

$Tr(\neg \lambda) \leftrightarrow \lambda$

where $\lambda$ is an appropriately encoded Liar sentence which intuitively says about itself ‘I am not true’. But since a truly rational agent must be certain of every axiom of (the weak and uncontroversial system of) Robinson arithmetic, one would also have to be certain of the diagonal equivalence

$\lambda \leftrightarrow \neg Tr(\neg \lambda)$

which is a theorem of Robinson arithmetic. Since the assignment of rational degrees of beliefs to sentences obeys the axioms of the probability calculus,
the set of sentences with degree 1 must be closed under logical consequence, so we would have to be certain of everything that is logically implied by these two equivalences for \( \lambda \), i.e.: We would have to be certain of a *contradiction*, which would not be rational at all (begging pardon to dialetheists about truth).\(^1\)

But taking every T-biconditional as certain is perhaps not what ought to be demanded of a rational believer anyway: rather, a rational believer should adjust her degrees of belief in a way such that \( \text{Tr}(\triangledown \alpha) \leftrightarrow \alpha \) always receive the *same* such degree. Note that if \( \text{Tr}(\triangledown \alpha) \leftrightarrow \alpha \) is regarded of degree 1, the axioms of probability entail that \( \text{Tr}(\triangledown \alpha) \leftrightarrow \alpha \) indeed have the same degree associated with them, while the probability axioms do not necessitate the converse direction.

Given that we accept probability measures to be the obvious means of assigning rational degrees of belief to sentences (or the propositions expressed by them), we may thus turn to what might be called

- *Probabilistic Convention T* (short: PCT):

  The epistemic probability measure\(^2\) \( P \) of a rational believer is such that for all sentences \( \alpha \):

  \[
  P(\text{Tr}(\triangledown \alpha)) = P(\alpha)
  \]

Can a rational believer adopt PCT in the context of a language that allows for self-referentiality and ungroundedness? Yes, indeed she can, by the help of idealized frequentistically defined probability measures (as e.g. investigated in a non-truth-theoretic context by Schurz&Leitgeb[10]), but only if \( \sigma \)-additivity is not considered to be necessary for epistemic probability measures. We will show this in the case of the language of arithmetic extended by \( \text{Tr} \).

Some terminology and basics: In the following, let \( \mathcal{L} \) be the usual first-order language of arithmetic. Our object language \( \mathcal{L}_{Tr} \) is the result of extending \( \mathcal{L} \) by a unary truth predicate \( \text{Tr} \) of sentences in \( \mathcal{L}_{Tr} \). We may presuppose substitutional quantification both inside and outside of the context of the truth predicate, due to the fact that every natural number \( n \) is denoted by its standard numeral \( \overline{n} \) in \( \mathcal{L} \). By Gödelization, the arithmetic language \( \mathcal{L} \) contains a theory of syntax for \( \mathcal{L}_{Tr} \). In particular, if \( \alpha \) is a sentence of \( \mathcal{L}_{Tr} \), then there is a singular term \( \triangledown \alpha \) in \( \mathcal{L} \) (and thus also in \( \mathcal{L}_{Tr} \)) which denotes the Gödel code of \( \alpha \). Accordingly, let \( \cdot \) be the arithmetically definable function sign that denotes the function that maps numbers to their numerals, and let \( \triangledown \alpha[\overline{x}] \) denote the arithmetically definable function which,
for $n$ as an argument, takes the code of $\alpha[n]$ as value. Finally, let $Val$ be the arithmetically definable function $\mathrm{sign}$ the interpretation of which is the semantic value mapping for constant arithmetical terms.

This is what we are going to prove:

**Theorem 1** There exists a function $P$, such that

1. $P$ is a finitely additive probability measure on $\mathcal{L}_{Tr}$, i.e.,
   - $P$ maps the sentences of $\mathcal{L}_{Tr}$ into $[0,1]$
   - For all logically true $\alpha \in \mathcal{L}_{Tr}$: $P(\alpha) = 1$
   - For all $\alpha,\beta \in \mathcal{L}_{Tr}$ such that $|=\neg(\alpha \land \beta)$:
     $$P(\alpha \lor \beta) = P(\alpha) + P(\beta)$$

2. $P$ satisfies PCT, i.e., for every $\alpha \in \mathcal{L}_{Tr}$:
   $$P(Tr(\neg\alpha)) = P(\alpha)$$

3. $P$ assigns 1 to all arithmetically true sentences of $\mathcal{L}_{Tr}$, i.e., to all the sentences that are true independently of how the standard model of arithmetic is expanded by an interpretation of $Tr$

4. $P$ assigns 1 to all of the standard recursive Tarskian truth clauses, i.e.: for all $\alpha, \beta$ and all constant terms $t, t'$ in $\mathcal{L}_{Tr}$, $P$ assigns 1 to
   - $Tr(\neg\forall x \alpha[x]) \leftrightarrow \forall x Tr(\neg\alpha[x])$
   - $Tr(\neg\alpha \land \beta) \leftrightarrow Tr(\neg\alpha) \land Tr(\neg\beta)$
   - $Tr(\neg\neg\alpha) \leftrightarrow \neg Tr(\neg\alpha)$
   - $Tr(\neg t = t') \leftrightarrow Val(t) = Val(t')$

Furthermore, no function which satisfies 1., 2., 3., 4. can also satisfy the following principle of continuity (see e.g. Earman[1], p.37), which is a weak version of $\sigma$-additivity for probability measures on sentences:

$$P(\exists x \alpha[x]) = \lim_{n \to \infty} P(\alpha[1] \lor \ldots \lor \alpha[n])$$

**Proof.** For the first part of the proof, in which we show the existence of a function $P$ as claimed above, consider a revision sequence in the sense of Belnap&Gupta[2] up to the ordinal $\omega$, i.e.: take the standard model of arithmetic as a ‘ground model’ for the language without truth predicate and
choose some initial extension of $Tr$; this gives us the initial model $M_1$. Define the next extension of $Tr$ to be the set of (codes of) sentences satisfied by $M_1$, which leaves us with the second model $M_2$; let the subsequent extension of $Tr$ be the set of (codes of) sentences satisfied in $M_2$, which defines $M_3$, and so forth. Thus we get an infinite sequence $M_1, M_2, M_3, \ldots$ of models in which, except for the initial one, $Tr$ is interpreted by the set of sentences true in the predecessor model. Now keep one such sequence $M_1, M_2, M_3, \ldots$ of models fixed: The guiding thought of the subsequent proof is to conceive of the probability of a formula $\alpha$ as the limiting frequency of the truth of $\alpha$ in this sequence. However, there is a complication: it might well be that for some $\alpha$ the frequency of its being true in our given sequence of models does not converge at all, which is why we have to turn to generalized and more abstract limit functions – so-called Banach limits – the existence of which follows from a well-known application of the classic Hahn-Banach theorem.

We will presuppose the standard terminology of functional analysis as well as the standard ways of deriving results on Banach limits from the Hahn-Banach theorem; for more details, see typical introductions to functional analysis (such as Werner[12], pp.93ff and pp.126f).

Let $F$ be the linear space of bounded real sequences $(x_n)$ for which $\lim_{n \to \infty} \frac{1}{n}(x_1 + \ldots + x_n)$ exists; note that for every convergent sequence $(x_n)$ it holds that $\lim_{n \to \infty} x_n = \lim_{n \to \infty} \frac{1}{n}(x_1 + \ldots + x_n)$. If we define $\varphi : F \to \mathbb{R}$, s.t. $\varphi((x_n)) = \lim_{n \to \infty} \frac{1}{n}(x_1 + \ldots + x_n)$, then $\varphi$ is a linear functional on $F$. Let $E = l^\infty(\mathbb{R})$ be the linear space of all bounded real sequences $(x_n)$, and let $p : E \to \mathbb{R}$, s.t. $p((x_n)) = \limsup_{n \to \infty} \frac{1}{n}(x_1 + \ldots + x_n)$. It follows that $F$ is a linear subspace of $E$, $p$ is a sublinear functional on $E$, and $\varphi$ is bounded from above by $p$. By the theorem of Hahn-Banach, $\varphi$ can be extended to a linear functional $\tilde{\varphi}$ on $E$, such that $\tilde{\varphi}$ is still bounded from above by $p$ (this extension is known not to be uniquely determined). By standard arguments one can prove that $\tilde{\varphi}$ is a Banach limit, i.e., (i) $\tilde{\varphi}(T((x_n))) = \tilde{\varphi}((x_n))$ for all $(x_n) \in E$, where $T$ is the shifting operator, i.e., $T((x_1,x_2,x_3,\ldots)) = (x_2,x_3,x_4,\ldots)$, (ii) if $x_n \geq 0$ for all $n \in \mathbb{N}$, then also $\tilde{\varphi}((x_n)) \geq 0$, (iii) $\tilde{\varphi}((1,1,1,\ldots)) = 1$. In particular, property (i) will turn out to be crucial to guarantee that $Tr(\overline{\alpha})$ and $\alpha$ will have the same probability; call (i) the Shifting Property.

Now we can exploit the one-to-one correspondence between subsets of \{\$M_1,M_2,M_3,\ldots\}$ and infinite binary sequences: for $X \subseteq \{M_1,M_2,M_3,\ldots\}$, let $f(X)$ be a real bounded sequence, so that $f(X)(n) = 1$ iff $M_n \in X$ and $f(X)(n) = 0$ else ($n \geq 1$). We define $P^*$ by $P^*(X) = \tilde{\varphi}(f(X))$. Finally, we can define a probability measure $P$ on $\mathcal{L}_{Tr}$ by means of:
for every sentence $\alpha \in \mathcal{L}_{Tr}$: $P(\alpha) := P^*(\{\mathcal{M}_n : \mathcal{M}_n \models \alpha\})$.

From (i)-(iii) above and $\tilde{\varphi}$ being linear, it follows that $P$ is indeed a finitely additive probability measure on sentences. $P$ also satisfies PCT from above, because:

$P(Tr(\alpha)) =, \text{ by the definition of } P,$

$= P^*(\{\mathcal{M}_n : \mathcal{M}_n \models Tr(\alpha)\}) =, \text{ by the definition of } P^*,$

$= \tilde{\varphi}(f(\{\mathcal{M}_n : \mathcal{M}_n \models Tr(\alpha)\})) =, \text{ by notating the value of } f \text{ by } (x_n),$

$= \tilde{\varphi}((x_1, x_2, x_3, \ldots)) =, \text{ by the Shifting Property},$

$= \tilde{\varphi}(f(\{\mathcal{M}_n : \mathcal{M}_n \models \alpha\})) =, \text{ by the definition of } P^*,$

$= P^*(\{\mathcal{M}_n : \mathcal{M}_n \models \alpha\}) =, \text{ by the definition of } P \text{ again},$

$= P(\alpha), \text{ which was to be proven.}$

Finally, 3. and 4. follow directly from the definition of $P$, the properties of Banach limits, and the determination of the underlying revision sequence of models.

For the second part of the proof, in which we will show the non-existence of a function $P$ which satisfies 1., 2., 3., 4. from above and continuity, we consider (a slight variant of) McGee’s ‘paradoxical’ sentence $\mu$ as introduced by McGee[9]: he shows that an equivalence $\mu \leftrightarrow \exists x \neg Tr(g(x, \lfloor \mu \rfloor))$ is derivable in Robinson arithmetic, where the arithmetically definable function sign $g$ denotes an arithmetical function $G$ with the property that $G(n, \text{code}(\alpha))$ is the Gödel code of the sentence $Tr(\ldots Tr(\lfloor \alpha \rfloor)\ldots)$ in which $Tr$ is iterated $n$ times (for $n \geq 0$). Intuitively, $\mu$ says of itself: there is an iterated application of the truth predicate to my name such that the result of that application is not true. Now assume there is a $P$ which satisfies 1., 4. and continuity: Since $P(\mu \leftrightarrow \exists x \neg Tr(g(x, \lfloor \mu \rfloor))) = 1$ by 3., we conclude from 1. that $P(\mu) = P(\exists x \neg Tr(g(x, \lfloor \mu \rfloor)))$. By 3. all axioms of Robinson’s arithmetic have probability 1; 1. implies that the set of formulas with probability 1 is closed under logical consequence; from 2. it follows that the set of formulas with probability 1 is closed under $\alpha Tr(\alpha)$ (necessitation); by 4. all the closure conditions on $Tr$ considered by McGee receive probability 1 as well. Hence, by the proof of McGee’s theorem, $P(\mu)$ and thus $P(\exists x \neg Tr(g(x, \lfloor \mu \rfloor)))$ are identical to 1, while at the same time $P(\neg Tr(g(\overline{1}, \lfloor \mu \rfloor)) \lor \ldots \lor \neg Tr(g(\overline{n}, \lfloor \mu \rfloor))) = 0$ for all $n$ (since from McGee’s proof, $P(\neg Tr(g(\overline{i}, \lfloor \mu \rfloor))) = 0$ for all $i \in \mathbb{N}$). But this contradicts continuity. Therefore, 1.–4. and continuity cannot be satisfied simultaneously. Accordingly, $P$ as defined in the first part of the proof invalidates continuity, and the corresponding measure $P^*$ is not $\sigma$-additive in the standard sense of probability theory. ■
Discussion: Since we now know that there indeed exist probability measures $P$ which satisfy PCT as well as all of the other conditions from the list above, let us have a look at what this tells rational believers to believe about truth (as far as our representative example of $\mathcal{L}_{Tr}$ is concerned).

First of all, since $P(\alpha) = 1$ for all logical truths $\alpha$, we also have that $P(Tr(\lnot \alpha)) = 1$ for such $\alpha$, and the same holds for all arithmetical truths $\alpha$. Furthermore, $P(Tr(\lnot \alpha)) = 1 - P(\alpha) = 1 - P(\lnot \alpha)$, by PCT again, so it follows from PCT and the axioms of probability that $Tr$ commutes with negations in a probabilistic context. Therefore, going back to our theorem above, it is not necessary to employ 4. in order to derive a result such as $P(Tr(\lnot \alpha)) = P(\lnot Tr(\alpha))$.

However, aiming at a probability measure which satisfies PCT and the clauses in 4. is still advisable: rational believers might simply want to be certain about the compositionality of truth where this certainty is expressed internally by assigning probability 1 to the compositionality clauses.

What about the Liar sentence? Since $\lambda \leftrightarrow \lnot Tr(\lambda)$ is derivable arithmetically, we have that $P(\lambda \leftrightarrow \lnot Tr(\lambda)) = 1$, which entails $P(\lambda) = P(\lnot Tr(\lambda)) = 1 - P(Tr(\lambda))$. Together with $P(Tr(\lnot \lambda)) = P(\lambda)$, which follows from PCT, this implies that $P(\lambda) = 1 - P(\lambda)$ and thus

$$P(\lambda) = P(\lnot \lambda) = \frac{1}{2}$$

This shows the strength of PCT: while it is very natural to assume that we should believe the Liar to the same degree as we ought to believe its negation – are they not indistinguishable on epistemic grounds? – we are not inclined at all to regard the material equivalence of the Liar and its negation as certain. This result might look somewhat similar to Kripke’s Strong and Weak Kleene approaches to the Liar in which both $\lambda$ and $\lnot \lambda$ are considered to be ‘undefined’, and even more so to the fuzzy logic approach by Hajek et al. [4], where $\lambda$ and $\lnot \lambda$ have the truth value $\frac{1}{2}$. But the similarity is only superficial: probability measures differ a lot from three-valued truth-functional evaluations and fuzzy possibility measures. In particular, while neither of the latter assign 1 to the laws of classical logic, probability measures indeed do so. In this respect, if anything, our theory is probably closest to Kripke’s supervaluationist account of truth.

There is also a connection to another formal theory of truth: Reconsider the proof of our theorem above; in the cases where $\varphi(f(\{M_n : M_n \models \alpha\}))$ exists, $P(\alpha)$ is nothing but the frequency of $\alpha$’s being true in the given revision sequence in the long run, which can be seen as the natural probabilistic expression of the concept of convergence as employed by the Revision The-
ory of Truth (Gupta&Belnap[2] and Herzberger[6]). Indeed, if a sentence \( \alpha \) is stably true within our \( \omega \)-sequence of models, i.e., if it remains true after some natural number, then \( P(\alpha) = 1 \). But even if the truth value of a sentences does not stabilize on 1, the sentence may still end up having probability 1 as long as the frequency of its being true tends towards 1. This is close to the spirit of so-called nearly stable sentences in the Revision Theory ([2], p.169). On the other hand, \( \lambda \) and \( \neg \lambda \) are not simply classified as ‘paradoxical’ as in the Revision Theory, but they are rather assigned a probability of \( \frac{1}{2} \), reflecting the fact that they are both true in ‘half’ of all the models. For these reasons, PCT, and more particularly the conditions on probability which are stated in our theorem, may be regarded justified by a variant of the Revision Theory of Truth for \( \omega \)-sequences, once the Revision Theory has been lifted from the ordinal to the quantitative scale.

The attractiveness of probability measures \( P \) as determined by the theorem above comes with a price, though. The set of sentences which are certain according to any such measure \( P \) is \( \omega \)-inconsistent: From 3. and 4., and since \( P(\alpha) = 1 \) iff \( P(\text{Tr}(\neg \alpha)) = 1 \) for all \( \alpha \) by PCT, it follows that every sentence of the theory FS of truth studied Halbach[5] must have probability 1. FS may be seen as an axiomatization of the revision semantics for finite levels, which supports the point made in the last paragraph. However, FS is also a supersystem of McGee’s[9] system, which McGee proved to be \( \omega \)-inconsistent. We leave open to what extent this speaks against probability measures as determined in the proof of our theorem – after all, \( \omega \)-inconsistency is not so uncommon in the realm of theories of truth for semantically closed languages (see Leitgeb[8]). Those who deem such a consequence unacceptable might want to study epistemic theories of truth which result from (completely or partially) dropping the conditions in 4. above while keeping our Probabilistic Convention T.

By similar formal means as the ones employed in this paper, also theories of type-free probability can be proven consistent: in such cases PCT gets replaced by probabilistic reflection principles the study of which might turn out to be useful for a better understanding of higher-order probabilities and probabilistic bridge postulates such as van Fraassen’s Reflection Principle and Lewis’ Principal Principle. But going into further details would mean going far beyond the topic of this article.

Acknowledgements: We are particularly grateful to Gerhard Schurz, Branden Fitelson, Reinhard Wolf, John Shepherdson, Theo A.F. Kuipers, the organizers and participants of the 2007 Banff workshop on Mathematical
Methods in Philosophy, and – as always – the members of the Luxemburger Zirkel.

Notes

1All of these inference steps are meant to be internal to a rational agent who is able to reason about truth, i.e., who can assign rational degrees of belief to certain statements involving the truth predicate. However, the agent is not assumed to reason about the probabilities that she assigns to statements involving the truth predicate. So her probabilistic reasoning is taken to be based on an ‘internal’ consequence relation that is defined on a language with a truth predicate rather than on an ‘external’ consequence relation which holds (among others) between sentences that include a symbol for the agent’s subjective probability function; the agent is only presupposed to be omniscient about the former consequence relation, not about the latter.

2Instead of treating probability measures as epistemic – as we do in this paper – one might perhaps treat them as ‘purely semantic’. In this way, PCT would turn into a semantic thesis rather than an epistemic one. The reason why we stick to the epistemic interpretation is that the notion of a semantic probability measure is not clear enough and thus demands further justification, while the epistemic approach is well-developed and constitutes the dominating view on probabilities. For the rare case of a more semantic view on probability measures, see e.g. Hailperin[3].

3So we consider probability measures to be defined directly on sentences. Alternatively, the results and proofs in this paper could be restated in terms of probabilities of sets of possible worlds or models, where the worlds or models in question would be expansions of the standard model of first-order arithmetic.

References


HANNES LEITGEB  
Departments of Philosophy and Mathematics  
University of Bristol  
9 Woodland Road  
Bristol BS8 1TB, UK  
e-mail: Hannes.Leitgeb@bristol.ac.uk