A Reliability Theory of Truth

Karl Schlechta

A Reliability Theory of Truth [*]

Karl Schlechta [†‡]

December 24, 2017

Abstract

Our approach is basically a coherence approach, but we avoid the well-known pitfalls of coherence theories of truth. Consistency is replaced by reliability, which expresses support and attack, and, in principle, every theory (or agent, message) counts. At the same time, we do not require a priviledged access to "reality".

A centerpiece of our approach is that we attribute reliability also to agents, messages, etc., so an unreliable source of information will be less important in future.

Our ideas can also be extended to value systems, and even actions, e.g., of animals.

Contents

1 Introduction

1.1 The Coherence and Correspondence Theories of Truth

See [Sta17a] and [Sta17b].

[*] File tru

[†] schcsg@gmail.com - https://sites.google.com/site/schlechtakarl/ - Koppeweg 24, D-97833 Frammersbach, Germany

[‡] Retired, formerly: Aix-Marseille Université, CNRS, LIF UMR 7279, F-13000 Marseille, France

We think that the criticisms of the coherence theory of truth are peripheral, but the criticism of the correspondence theory of truth is fundamental.

The criticism of the correspondence theory, that we have no direct access to reality, and have to do with our limitations in observing and thinking, seems fundamental to the author. The discussion whether there are some "correct" theories our brains are unable to formulate, is taken seriously by physicists, likewise the discussion, whether e.g. Quarks are real, or only helpful "images" to understand reality, was taken very seriously. E.g. Gell Mann was longtime undecided about it, and people perhaps just got used to them. We don't know what reality is, and it seems we will never know.

On the other side, two main criticisms of the coherence theory can be easily countered, in our opinion. Russell's objection, that ϕ and $\neg\phi$ may both be consistent with a given theory, shows just that "consistency" is the wrong interpretation of "coherence", and it also leaves open the question which logic we work in. The objection that the background theory against which we check coherence is undefined, can be countered with a simple argument: Everything. In "reality", of course, this is not the case. If we have a difficult physical problem, we will not ask our baker, and even if he has an opinion, we will not give it much consideration. Sources of information are assessed, and only "good" sources (for the problem at hand!) will be considered. (Thus, we also avoid the postmodernist trap: there are standards of "normal reasoning" whose values have been shown in unbiased everyday life, and against which standards of every society have to be compared. No hope for the political crackpots here!)

1.2 Our Approach: a Variant of Coherence Theory

Perhaps an example helps to illustrate ou view about truth.

Example 1.1

Suppose we are interested in the size of an object X. We cannot access X directly, and have to rely on witnesses.

Witness A had a meter, measured X, and says X is 120 cm long. Unfortunately, A is known to be crackpot.

Thus, we limit our sources of information to reliable ones.

Witness B had no meter, he measured using his thumb, and later calculated the length to be 90 cm.

Witness C had a meter, but the meter was old and twisted, so not very accurate. C says that X is 101 cm long.

By experience, C's method is superior to B's method.

This is all we know.

Based on this information, we say "our best estimate is that X is 101 cm long".

We do not doubt that there is some "real" length of X, but this is irrelevant, as we cannot know it. We have to do with what we know, but are aware that additional information might lead us to revise our estimate.

This story seems simple, but even much more complicated stories can be solved by essentially the same, simple ideas, we think.

The following extensions seem possible:

- Actions and animals

  We can apply similar reasoning to actions. The action of a monkey which sees a lion and climbs a tree to safety is "true", or, better, adequate.

- Values

  Values, obligations, "natural laws" are subjective. Still, influences are known, and we can try to peel them off. Religion, politics, personal history, influence our ideas about values. One can try to find the "common" and "reasonable" core of them.

We give another example. Reliabilities will be denoted $\rho(.)$.

Example 1.2

We have a meteorological station in Siberia. The thermometer is supposed to be reliable, it automatically records the current temperature (with time stamp etc.) reliably. ($\rho = 1$ below. This is "reality", which is introduced purely as a trick, to illustrate simple cases, we discard this later.) Sometimes, we want to know the current temperature immediately, we phone the human operator or operators and ask. Unfortunately, the line or lines is/are very noisy, and errors in transmission occurr. This is the common part.

Case 1:

The human operator is absolutely reliable. Later, we compare the temperature as transmitted by phone with the recorded temperature, and assign $\rho(t)$ to the transmission t.

The values of the absolutely reliable thermometer etc. are unchanged.

Case 2:

We have two reliable human operators, they use different unreliable phone lines. So we have transmissions t and t'. Assume that, e.g. based on previous experience, we gave t and t' already initial values $\rho(t)$, $\rho(t')$.

If T and T' agree (T, T' the temperatures received, without any knowledge of the "real" temperature!), we increase $\rho(t)$ and $\rho(t')$, they confirm each other.

If T and T' disagree (without any knowledge of the "real" temperature!), we decrease $\rho(t)$ and $\rho(t')$, as they contradict each other, based on their initial values.

Case 3:

As it is so cold, the human operators often drink too much, so they are not reliable. They make mistakes, and the transmission is not reliable, either. (To simplify, we assume that mistakes do not cancel each other, e.g. the operator reads 10 degrees too much, and the line transmits 10 degrees too little, so the correct value is transmitted.)

Case 3.1:

We consider only one human operator h, and one transmission line t, and compare the value with the recorded temperature. We first calculate the combined reliability of the chain ht, $\rho(ht)$.

If the transmitted temperature agrees with the recorded value, we increase the combined reliability $\rho(ht)$, and break this down to increases of $\rho(h)$ and $\rho(t)$, according to their previous values.

If they disagree, we first decrease the combined reliability $\rho(ht)$, and break this down again to decreases of $\rho(h)$ and $\rho(t)$.

Case 3.2:

We consider both operators and transmission lines, and compare the received values. Let h, h' be the human operators, t, t' the transmission lines. We proceed as in Case 3.1, but first adjust both $\rho(ht)$ and $\rho(h't')$, as in Case 2, and break the adjustments down to $\rho(h)$, $\rho(h')$, $\rho(t)$, $\rho(t')$ as in Case 3.1.

2 Details

2.1 Agents, Messages, and Reliabilities

Definition 2.1

(1) We have agents A, A' etc. and messages, M, M' etc.

Agents may be people, devices like thermometers, transmission lines, theories, etc.

Agents are named, otherwise we have to introduce each occurrence as a new agent.

(2) If agent A sends a message M, A will be called the source of M.

(3) Agents and messages have values of reliability, $\rho(A)$, $\rho(M)$, etc. Sometimes, it is more adequate to see reliability as degree of competence.

(4) Messages may be numbers, e.g. a temperature, but also opinions about something, e.g. "the earth is flat", also about the reliability of agents and messages. Messages may also be moral judgements. If a message is a moral judgement, then moral competence of the source of the message is important. If the source is the constitution of a country, the competence will probably be considered high, etc. In case of theories, the messages may be consequences of this theory, etc.

(5) A human agent may be a good chemist, but a poor mathematician, so his reliability varies with the subject. We neglect this here, and treat this agent as two diffent agents, A-Chemist, A-Mathematician, etc. Likewise, a thermometer may be reliable between 0 and 30 C, but less reliable below 0. Again, we may describe them as two different thermometers.

(6) Reliabilities (of agents or messages) will be multisets of the form $\{r_i a_i : i \in I\}$. r_i will be a real value between -1 and +1, a_i should be seen as a "dimension". $r_i = 1$ is, intuitively, maximal support (or competence) $r_i = -1$ maximal attack, and 0 neither support nor attack, etc.

This allows for easy adjustment, e.g. ageing over time, shifting importance, etc., as we will shortly detail now:

- the real values allow arbitrarily fine adjustments, it is not just $\{-1, 0, 1\}$,
- the dimensions allow to treat various aspects in different ways,
- for instance, we can introduce new agents with a totally "clean slate", 0 in every dimension, or preset some dimensions, but not others,
- the uniform treatment of all dimensions in Section 2.2 (page 5) is not necessary, we can treat different dimensions differently, e.g., conflicts between two agents in dimension a_i need not touch dimension a_j, etc.

We have arbitrarily many dimensions, with possibly different meaning and treatment, and within each dimension arbitrarily many values. This is not a total order, but within each dimension, it is.

(7) We define as usual $[-1, 1] := \{r \in \mathcal{R} : -1 \leq r \leq 1\}$, etc.

We will use some element-wise operations on reliabilities. If the operations work on more than one reliability, any dimension a which is present in some, but not all reliabilities, will be added where necessary with the factor $r = 0$.

Definition 2.2

(1) $\rho \leq \rho'$ iff $\forall a_i (r_i \leq r'_i)$ - where $0 r_i$ was added where necessary.

(2) for $r \in [-1, 1]$ $r\rho = \{(r * r_i) a_i : i \in I\}$

Note that, if $r < 0$, attack becomes support and vice versa.

(3) If $r, r' \in [0, 1]$, and $r + r' = 1$, we define the weighted mean of ρ and ρ' as $\{(r * r_i + r' * r'_i) a_i : i \in I\}$. (Similarly for more than two ρ's.)

(4) The average reliability of ρ is defined as $av(\rho) := \frac{\Sigma\{r_i : i \in I\}}{card(I)}$.

(5) We may want to give more reliable messages etc. more weight. E.g. for ρ, ρ', we may want to calculate the weighted mean of ρ, ρ', depending on their individual reliabilities. We can do this as follows:

Consider $av(\rho)$ and $av(\rho')$. If $av(\rho) \geq av(\rho')$, take $d := \frac{1}{2} + \frac{1}{4}(av(\rho) - av(\rho'))$. Then $d \in [0.5, 1]$, and this will be the weight for ρ. The weight for ρ' will be $1 - d$. If $av(\rho) < av(\rho')$, similarly.

(6) $min(\rho, \rho')$, $max(\rho, \rho')$, $\rho * \rho'$, and $\rho + \rho'$ are defined in the obvious way (after adding $0 a_i$ where necessary, as usual). In the case of $\rho + \rho'$ we have to cut off at the borders -1, 1.

2.2 Combinations

We discuss now a number of cases. The solutions are suggestions, often, one will find alternatives which might be as good or even better. Our discussion is more centered on basic ideas than on details - which might depend on context, too. A good overall algorithm will probably be quite robust against local changes.

In the following, we will treat only conflicts between two agents/messages. Of course, also situations like three values, 8, 9, and -1, need to be treated (example due to D. Makinson), where we will give more credibility to $\{8,9\}$ than to the exceptional value -1. We treat in the following only pairs, so $\{8,9\}$, $\{-1,8\}$, $\{-1,9\}$, which should go in the same sense as treating the triple $\{-1,8,9\}$.

2.2.1 $\rho(A)$ and $\rho(M)$

(1) From $\rho(A)$ to $\rho(M)$:

If agent A sends message M, the unadjusted reliability of M will by default be the reliability of A. It may be adjusted later, due to other messages.

(2) From $\rho(M)$ to $\rho(A)$:

When $\rho(M)$ was modified, this should have repercussions on $\rho(A)$. If $\rho(M)$ was increased, $\rho(A)$ should increase, too, if $\rho(M)$ was decreased, $\rho(A)$ should decrease, too. The effect should be "dampened" however. One wrong message should not totally destroy the reliability of the agent. We use a weighted mean, e.g. 0.9 weight for the old $\rho(A)$, and 0.1 for $\rho(M)$, to calculate the new $\rho(A)$.

2.2.2 Chains of Messages and Reliabilities

See Example 1.2 (page 3), Case 3.

Suppose agent A sends message M, and agent A' passes M on, perhaps with some modification, so this is message M'.

(1) The combined message MM' will have some $\rho(MM')$. It seems natural to set $\rho(MM') := min(\rho(M), \rho(M'))$, or similarly.

(2) Conversely: Suppose we have modified $\rho(MM')$, and given it a new value $\sigma(MM')$. We have to break down the modification to new $\sigma(M)$ and $\sigma(M')$, in a reasonable way, so that again $\sigma(MM') := min(\sigma(M), \sigma(M'))$.

We take a local approach, i.e. work for each dimension separately. Many other approaches are possible.

Recall that the old $\rho(MM')$ was calculated as the $min(\rho(M), \rho(M'))$.

Let $\sigma(MM')$ be $\{s_i a_i : i \in I\}$. Call the new r_i (to be computed) t_i, likewise t'_i for r'_i.

Case 1:

$s_i = min\{r_i, r'_i\}$.

We leave r_i and r'_i unchanged, i.e. $t_i := r_i$, $t'_i := r'_i$.

Case 2:

$s_i \neq min\{r_i, r'_i\}$.

Case 2.1:

$r_i = r'_i$. We set $t_i := t'_i := s_i$.

Case 2.2:

$r_i < r'_i$. (The case $r_i > r'_i$ is analogous.)

As r_i is less reliable, it should change more.

If $s_i < r_i$, set $t_i := s_i$, and, e.g., $t'_i := r'_i - 0.5 * (r_i - s_i)$, where the 0.5 expresses the smaller change.

If $s_i > r_i$, set $t_i := s_i$, and, e.g., $t'_i := min\{1, max\{s_i, r'_i + 0.5 * (s_i - r_i)\}\}$.

Again, many other solutions are reasonable.

2.2.3 Two Parallel Messages

Different agents might send messages about the same subject. Those messages might agree, or not.

Case 1:

The messages agree. The reliabilities support each other, and we set both to $max\{\rho, \rho'\}$.

Case 2:

The messages disagree. The stronger ρ' is, the more it contradicts ρ, and vice versa. We do again a local calculation.

Case 2.1:

$\rho_i \leq 0$, $\rho'_i \leq 0$. We decrease both, e.g. ρ_i to $\rho_i - 0.1 * (1 + \rho_i)$ and ρ'_i to $\rho'_i - 0.1 * (1 + \rho'_i)$.

Case 2.2:

$\rho_i > 0$, $\rho'_i > 0$. We decrease again both, e.g. ρ_i to $\rho_i - 0.1 * \rho_i$ and ρ'_i to $\rho'_i - 0.1 * \rho'_i$.

Case 2.3:

$\rho_i \leq 0$, $\rho'_i > 0$. We decrease e.g. ρ_i to $\rho_i - 0.1 * (1 + \rho_i)$ and leave ρ'_i unchanged.

2.2.4 Two Parallel Messages About Reliability

Suppose agent A sends message M with reliability ρ, A' M' with reliability ρ', the contents of M is a reliability μ, that of M' a reliability μ', where μ and μ' are about the same agent A'' or message M''. We want to calculate a new reliability ν (of A'' or M''). We proceed again by components (= dimensions). We put in 0's where necessary (as in Definition 2.2 (page 4)).

Case 1:

If $\mu_i = \mu'_i$, we set $\nu_i := \mu_i$.

Case 2:

$\mu_i \neq \mu'_i$. We take a mean value, depending on the strengths ρ_i and ρ'_i.

Case 2.1:

Suppose $\rho_i = \rho'_i$. We take $\nu_i := 0.5 * (\mu_i + \mu'_i)$.

Case 2.2:

Suppose $\rho_i < \rho'_i$. (The case $\rho_i > \rho'_i$ is analogous.) We take an approach slightly different from that of Section 2.2.2 (page 5).

If $\rho_i < 0$, and $\rho'_i \geq 0$, we forget μ_i (as it is only attacked), and set $\nu_i := \mu'_i$.

Otherwise, set $x := 0.25 * (\rho'_i - \rho_i)$, this is a value between 0 and 0.5, and set $\nu_i := (0.5 - x) * \mu_i + (0.5 + x) * \mu'_i$, giving μ'_i more weight, as $\rho_i < \rho'_i$.

3 Discussion

Our approach is very pragmatic, and takes its intuition from e.g. physics, where a theory is considered true - but revisably so! - when there is “sufficient” confirmation, by experiments, support from other theories, etc.

Many human efforts are about establishing reliability of humans or devices. An egineer or physician has to undergo exams to assure that he is competent, a bridge has to meet construction standards, etc. All this is not infallible, experts make mistakes, new, unknown possibilities of failure may appear - we just try to do our best.

Our approach has some similarities with the utility approach, see the discussion in [BB11], the chapter on utility. An assumption, though false, can be useful: if you think a lion is outside, and keep the door closed, this is useful, even if, in fact, it is a tiger which is outside. “A lion is outside” is false, but sufficiently true. We think that this shows again that truth should not be seen as something absolute, but as something we can at best approximate; and, conversely, that it is not necessary to know “absolute truth”. We go beyond utility, as improvement is implicit in our approach. Of course, approximation may only be an illusion generated by the fact that we develop theories which seem to fit better and better, but whether we approach reality and truth, or, on the contrary, move away from reality and truth, we cannot know.

Our ideas in Section 2 (page 3) are examples how it can be done, but no definite solutions. The exact choice is perhaps not so important, as long as there is a process of permanent adjustment.

There are many things we did not consider, e.g.:

(1) Loops, and more generally, complex structures with feedback: how to treat circular support, without excessive "auto-confirmation".

(2) How adjustments are propagated, how this is coordinated (locally or globally), and when does adjustment stop?

(3) Do more complicated structures have stronger inertia against adjustment?

On the other hand, these shortcomings are not very serious, as we have an example structure which handles these problems very well: our brain. (Attacks, negative values of reliability, correspond to inhibitory synapses, positive values, support, to excitatory synapses. Complex, connected structures with loops are created all the time without uncontrolled feedback, theories about the world survive some attack (inertia), until "enough is enough".)

3.1 Is this a Theory of Truth?

The author thinks that, yes, though we hardly mentioned truth in the text.

Modern physics seem to be perhaps the best attempt to find out what "reality" is, what "truly holds". We had the development of physics in mind, reliability of experiments, measurements, coherence of theories (forward and backward influence of reliabilities), reputation of certain physicists, predictions, etc. Of course, the present text is only a very rough sketch, we see it as a first attempt, providing some highly flexible ingredients for a more complete theory in this spirit.

4 Acknowledgements

The author would like to thank David Makinson for very helpful comments.

References

[BB11] A. G. Burgess, J. P. Burgess, “Truth”, Princeton University Press, Princeton, 2011

[Sta17a] Stanford Encyclopedia of Philosophy, “The coherence theory of truth”

[Sta17b] Stanford Encyclopedia of Philosophy, “The correspondence theory of truth”