

How to Take Truth as a Goal?

Marian David

Of the intellect which is contemplative, not practical nor productive, the good and the bad state are truth and falsity...

Aristotle, *Nicomachean Ethics*, vi.2

As to the preference which most people—as long as they are not annoyed by instances—feel in favor of true propositions, this must be based, apparently, upon an ultimate ethical proposition: ‘It is good to believe true propositions, and bad to believe false ones’. This proposition, it is to be hoped, is true; but if it is not, there is no reason to think that we do ill in believing it.

Bertrand Russell (1904, 76).

I

The idea that truth is a positive value, a good, is quite prominent in philosophy, also in science (or at least in popular science), and in religious thinking. In these fields, truth tends to be referred to as a great good, a worthy goal, something desirable to strive after, and talk of “the pursuit of truth” or “the search for truth” as eminently worthwhile is all but common-place, especially in philosophy. To my knowledge, Nietzsche was the only major philosopher who dared to deprecate, even ridicule, “the will to truth”, if only at times. With this exception, paying homage to truth is pretty much the norm in philosophy—though maybe not always easily distinguished from paying lip service.

When one talks about truth as a value, what is meant is not actually truth *per se*; what is meant, rather, is *possessing truth*. There are at least two important ways of possessing truth that are often not clearly separated: (a) having true beliefs, and (b) having knowledge. They are distinct, but not entirely: the second entails the first, but not the other way round. A perceived difference in value between the two prompted the question: “Why is knowledge better than mere true belief?” First raised in Plato’s *Meno*, the question is under discussion again in recent and very recent epistemology; it is part of the so-called “value problem”. However, here I will focus just on (a), which is typically referred to simply as the truth-goal, even though it would be more accurate to call it the true-belief goal.

If true belief is a positive value, then it is good, or a good, at least *prima facie*, to have true beliefs. This leads naturally to the view that it *ought* to be a goal, ought to be one of our goals, to have true beliefs, that we ought to want to have true beliefs. This normative thought, in turn, leads one naturally to look more closely at the truth-goal (the true-belief goal) itself, bracketing normative concerns, if only temporarily: What is it, anyway, to have having true beliefs as a goal? What is it to want to have true beliefs? Though I will take occasional glances at the idea that having true beliefs is good, in this

paper, I will be mostly concerned with the non-normative (preliminary) aspect of our topic, with having true beliefs as a goal, with desiring or wanting to have true beliefs.¹

II

I assume that belief (believing) is an attitude towards a *proposition* which encapsulates the content of the belief. To say that a belief is true is best understood as saying that *what is believed* is true, and the *what is believed* is a proposition. In this framework, talk of true (false) belief comes out as talk of believing true (false) propositions; and formulations pertaining to the truth-goal in sense (a) come out as: It is a goal, one of my goals, to believe true propositions; I want to believe true propositions. But the truth-goal is supposed to have two parts, a *positive* part and a *negative* part. For starters, the contents of the two parts might be put like this—using ‘*p*’ as an objectual variable ranging over propositions:

positive	and	negative
to believe <i>p</i> , if <i>p</i> is true		not to believe <i>p</i> , if <i>p</i> is not true
$Tp \rightarrow Bp$		$\sim Tp \rightarrow \sim Bp$

Below the (almost) ordinary language formulations, we find logical formulas, widely used as abbreviations. The formulas have limitations. The one intended to represent the content of the positive goal does not make clear that the wanting is—in some sense not easy to make precise—directed at the consequent, rather than the antecedent, of the conditional; the wanting points at the believing rather than the truth of the proposition:

I want to believe *p*, if *p* is true;

I want to be such that (I believe *p*, if *p* is true).

These are alternative acceptable formulations of what is intended. If, on the other hand, we read the formula with the wanting pointing at the formula’s antecedent instead, we get unintended results:

*I want *p* to be true, only if I believe *p*;

*I want to be such that (*p* is true, only if I believe *p*).

This is not what is intended. In fact, these formulations seem a bit crazy. They suggest that, in case the world goes beyond my beliefs, I expect *it* to change, to back down as it were, rather than being prepared to change my beliefs. This is the wrong direction of fit.

From the examples one might surmise that the ‘want’ always points at that part of the ordinary language conditional that is closer to it; however, that can’t be quite right, because:

I want to be such that (if *p* is true, then I believe *p*)

is an acceptable rendering of the content of the positive goal, with the ‘want’ properly pointed at the believing, even though the ‘believe’ is more distant than the ‘true’.

Similar considerations apply to the formula intended to represent the content of the negative goal. Here, too, the wanting should be taken to point at the consequent of the conditional formula: ‘I want to not believe *p*, if *p* isn’t true’; not: ‘I want *p* to be not true, only if I don’t believe it’.

The two parts of the goal are often combined into one with the formula: $(Bp \leftrightarrow Tp)$. This looks acceptable, *provided* it is read with the wanting pointing firmly at the believing: ‘I want to believe p , if p is true, and only if p is true’. Note that this requires reading the right-to-left direction of the formula, $(Tp \rightarrow Bp)$, as before, with the wanting pointed at the consequent (wanting to believe p , if p is true), while reading the left-to-right direction, $(Bp \rightarrow Tp)$, with the wanting pointed at the antecedent (wanting to believe p , only if p is true).

Concerning the latter, note also that ‘goal’ and ‘want’-talk give rise to opaque (intensional) contexts: even though $(\sim Tp \rightarrow \sim Bp)$ is logically equivalent to $(Bp \rightarrow Tp)$, it does not follow that the goal $(\sim Tp \rightarrow \sim Bp)$ is the same as the goal $(Bp \rightarrow Tp)$. But this issue is a bit confusing because the logical formulas can be read in different ways. As I have in effect suggested above, if the goal $(Bp \rightarrow Tp)$ is read as wanting to believe p , only if p is true, there doesn’t seem to be any important difference between this and the explicitly negative formulation: wanting to not believe p , if p is not true, i.e. $(\sim Tp \rightarrow \sim Bp)$; both will serve as abbreviations of the negative part of the goal. But if the goal $(Bp \rightarrow Tp)$ is read along the lines of wanting to be such that, if one believes p , then p is true, there is an important difference between this and the explicitly negative formulation, even though the formulas are logical equivalents. As Christian Piller (2009, 198) points out, wanting to be such that, if one believes p , then p is true, *wanting one’s beliefs to be true*, exhibits more of a concern for oneself—wanting to be right—than a genuine concern with the truth. A weather forecaster, having predicted bad weather, may want her prediction to be true, she may want to be right; but this is not a concern for the truth, it is more like wanting to win. Someone who wants her beliefs to be true, who wants to be right, would, in case of mismatch, prefer that the world conform to her beliefs, rather than make her beliefs conform to the world. Again, this is the wrong direction of fit (see also David 2005, 296). So, on an apparently natural reading of $(Bp \rightarrow Tp)$, it specifies a goal importantly different from the goal of not believing p , if p is not true, in spite of the logical equivalence with the formula $(\sim Tp \rightarrow \sim Bp)$.

III

Let me focus now on the positive part of the goal. The formulations used above are deficient; they don’t pay sufficient attention to quantifiers. Once we think about quantifiers, we notice that there are three different forms associated with the positive truth-goal, distinguished by the scope of the goal-operator relative to the quantifier ‘ $\forall p$ ’, i.e. ‘for all propositions p ’:

1. Goal: $(\forall p) (Tp \rightarrow Bp)$
2. $(\forall p)$ Goal: $(Tp \rightarrow Bp)$
3. $(\forall p) (Tp \rightarrow \text{Goal: } Bp)$

Note that only 1 can be taken as specifying a possible goal, the goal of believing all true propositions, or the goal of being such that one believes all true propositions. 2 and 3 can’t be taken as specifications of a goal at all, they make sense only when taken as statements—absurdly false statements, I should say. But let us look at 1 first.

I said that 1 can be taken as specifying a possible goal. This is not to say that it specifies a *plausible* goal. It is frequently pointed out that there are simply way too many,

too trivial and uninteresting truths for 1 to be a plausible goal: no one in their right mind would want to believe all the truths available from the local phone book, and all the truths available by counting the flowers on the wallpaper of one's hotel room, and similar trivialities. Consider especially the mindboggling redundancy of trivialities involved: every pair of boring truths of the 'At least n flowers'-variety gives rise to an equally boring conjunctive truth and an equally boring disjunctive truth; and there is much more "garbage" of this sort. Still, 1 specifies a (in principle) possible goal. Use the goal thus specified to form a statement ascribing it to some subject S , viz.: 'S has the goal of believing all true propositions'. The complaint is not that statements of this form cannot possibly be true. The complaint is, rather, that they will in fact be false for almost all S ; that most subjects won't have this goal and shouldn't be expected to have it either; that most of the few who may profess to have this goal will, on reflection, come to realize fairly quickly that they don't *really* have it; and that the remaining few, if any, are not quite of sound mind.

Maybe this is a bit too quick. Yes, we don't have desires for believing trivial truths and don't have desires for believing large sets of trivial truths (the phone-book truths, the wallpaper truths). It doesn't follow, though, that we don't have some desire for believing the whole truth. There is some allure to the idea of believing *all* the truths—the air of insanity flows, not so much from the goal itself, but from the all too obvious unrealizability of the goal: even if we had an infallible method for distinguishing truths from falsehoods, we couldn't possibly achieve this goal. Still, it does seem alluring somehow. In any case, remember that we are talking about the goal of having true beliefs in part because we are interested in the idea of true belief being a value, a good. And remember that many people ascribe to God attributes entailing that God believes all truths, apparently including the most stultifyingly trivial and redundant ones. This is evidence that many regard believing all truths as a good, as an in principle valuable state to be in, even though they may not be much inclined to adopt it as a goal for themselves.

This being said, it does seem that the goal of believing all true propositions is a bit extravagant. Looking for a more plausible goal, one will be propose restricting the quantifier in 1 to propositions that are *interesting*. Most likely, this means that a subjective element is being inserted into the goal. At least, it is difficult to see how being interesting could be an objective property of propositions: what is interesting seems to be relative to, well, our interests. It also means giving up on any idea of a "pure" truth goal. This is not an objection; just a reminder that there is now a second value in play: what is (we find) interesting. Alternatively, one might propose restricting the quantifier in 1 to propositions one has *considered*; more precisely, to propositions p with respect to which one has asked oneself whether p . By my lights, this is too restrictive. Insofar as I have the goal to have true beliefs, I also want to find truths beyond my horizon, truths concerning questions I have not considered.

Aristotle at one point mentions "indefinite" statements: statements that lack an explicit quantifier (cf. *De Int.*, 17^a38). I surmise that an indefinite formulation of the truth-goal would be the one people would be most likely to assent to when queried; something like: "I want to believe true propositions and not believe false ones". Putting in an explicit quantifier tends to end up sounding too bold (universal quantifier) or too meek (existential quantifier). Recently, there has been some interest in *generics*, i.e. claims like: Tigers have stripes (cf. Leslie 2012). Maybe a truth-goal plausibly

attributable to people, a psychologically realistic truth-goal, is one with a generic content: believing true propositions.

Let us consider more briefly item 2 from above. I pointed out that it does not specify a goal; instead, it makes a statement: it says that, with respect to each and every proposition, a subject *S* has the *conditional* goal to believe it, if it is true. This cannot be right; there are way too many propositions for this to be humanly possible, non-denumerably many; and even if we restrict the quantifier to interesting propositions, there are still way too many. We cannot have that many goals, that many wants. Of course, there may be many correct instances of 2. For many propositions *p*, a subject *S* may have the following conditional goal/want: to believe *p*, if it is true.

Consider 3. It too makes a statement, and again a false one. It says that, with respect to each and every true proposition, *S* has the non-conditional goal to believe it. This assumes, implausibly, that *S* has somehow managed to single out exactly the true propositions; in any case, there are still way too many true propositions (and way too many interesting true propositions) for this to be humanly possible.²

Item 3 has a special feature that makes even its instances problematic, or at least unusual. Each instance ascribes the non-conditional goal, with respect to a particular proposition, to believe *it*; each instance talks about wanting to believe a certain proposition. Now, wanting to believe a particular proposition is rather extraordinary. Not impossible, we can think of cases. Having witnessed, say, great suffering, a believer may have lost his faith in God's goodness; realizing this, he may very much *want to believe* this particular proposition, *that God is good*, but may find it hard going (at least initially). This sort of thing is possible, but it is a rare and rather special condition to be in. Believing is not under our voluntary control, at least not in any direct manner. We are to some extent aware of this, if only darkly, which goes some way towards explaining why we don't typically find ourselves wanting to believe particular propositions—except for extraordinary circumstances. Cynics may add the suspicion that our subconscious often enough manages to circumvent this problem; that below the threshold of conscious control we often want to believe particular propositions (ones that make us feel good about ourselves) and proceed to believe them because “we” want to.³

With respect to both, 2 and 3, one may observe that, while they are plainly false, attributing absurdly impossible numbers of goals, the results of replacing ‘goal’ with ‘good’ are not equally absurd. It is not impossible that every proposition is such that it is good to believe it, if it is true; nor is it impossible that every true proposition is such that it is good to believe it. Indeed, Michael Lynch (2004) holds that all these things are good, at least *prima facie* good.

IV

I now want to consider a puzzle that arises with respect to the positive part of the true-belief goal. To be honest, I am not entirely sure that there really is a puzzle—some try to tell me that there isn't one. Still, it seems to me that there is something puzzling, and I want to bring it up.

Consider first what appears to be a fairly plausible *transition principle* for conditional desires (wants), proposed by Piller (2009, 196). Here is how Piller puts it: “Wanting if A then B and noticing that A commits me to wanting B or, at least, gives me a reason to want B”. Note that the conditional desire is presented in the somewhat

artificial form ‘wanting if A then B’; the more natural form is “wanting B, if A’. Note also that ‘noticing that A’ involves believing that A; indeed, nothing more epistemically demanding is required here than belief.

Now consider the following scenario, for comparison. Imagine I have recently developed a nostalgic longing for the music of the 70ies glamor-rock band *T. Rex*. I want to buy their CDs (and only their CDs, but I’ll suppress this part for simplicity). With this longing, I enter a music store:

In the music store:

I want to buy all CDs by T. Rex

I want: $(\forall x)(Tx \rightarrow Bx)$,
quantifier restricted to CDs

I go through the CDs in the bargain bin,
for each one:

I want to buy it, if it is by T. Rex

I want: $(Tx \rightarrow Bx)$,
for each of the x ’s

I find one buy T. Rex:

I want to buy it.

I want: Bx .

This progression involves analogues of items 1, 2, and 3 discussed earlier. The transition from 2 to 3 goes by Piller’s principle. The earlier transition from 1 to 2 goes by a closely related principle. Considering a CD from the bin, the global goal turns into a conditional goal with respect to the CD under consideration. When I come to believe of such a CD that it is by T. Rex, the conditional goal turns into a non-conditional goal with respect to that particular CD: I want to buy it (unless it is too expensive, or broken, etc.). So far so good.

Now, imagine I have the goal of believing true propositions, and imagine that I am presented with a list of propositions (or if you want, a list of sentences expressing propositions):

A list of propositions:

I want to believe all the ones that are true

I want: $(\forall p)(Tp \rightarrow Bp)$,
quantifier restricted to the list

I go through the propositions on the list,
for each one:

I want to believe it, if it is true

I want: $(Tp \rightarrow Bp)$,
for each of the p ’s

I notice one that is true:

I want to believe it ?

I want: Bp ?

Unlike the first scenario, where there is nothing at all remarkable, the last row of this second scenario strikes me as odd. Considering an item on the list, my global goal (form 1 from above) is converted into the conditional goal, with respect to the proposition under consideration, to believe it, if it is true (an instance of 2). Now I notice of the proposition under consideration that it is true. What happens now? Remember, noticing that p is true amounts to believing that p is true: Given that I already believe that p is true, does it make sense to say that I want to believe p , or that I am committed to wanting to

believe p , or that my believing p gives me a reason to want to believe p ? I am not among those who subscribe to the strong deflationary view about truth, that *believing that p is true* = *believing p* . Still, the following does hold: if one believes that p is true, one believes p . How does believing p go together with *wanting* to believe p , or being committed to wanting to believe p , or having a reason to want to believe p ?

Admittedly, my setup is somewhat artificial. However, we certainly *can* do what I described. We can go through a list with the global goal in mind to believe the items that are true (the common activity of reading non-fiction literature is a bit like that). But when we do this, what happens is not at all like what happens when we go through some CDs with the global goal in mind to buy the ones by T. Rex. Why is that? Coming to believe that a CD has the “search property” *being by T. Rex* is quite different from buying it, so there is something additional left for me to do which I can want to do. But coming to believe that a proposition has the search property *being true* entails believing it, so there is nothing further for me to do and, it seems, nothing for me to want to do. The desire for music by T. Rex is *action guiding*; the desire for true belief is not.

Ernest Sosa (2001, 50) remarks briefly that the desire for *safety* cannot guide how we form beliefs. Sosa’s safety is of the form ($Bp \rightarrow Tp$), which is the positive-formula representation of the negative part of the truth-goal (discussed above in Section II). If this is read properly, as wanting to believe p , only if p is true, i.e. wanting not to believe p , if p is not true, his remark concerns a closely related point, namely the absence of guidance for non-belief: believing that p is not true, already entails not believing p , so once one believes that p is not true, there is nothing further left to do or to want, as far as one’s desire for safety is concerned.⁴

One feature that makes the second scenario odd is that it is a case of wanting to believe a particular proposition. I noted earlier (with respect to instances of 3) that this is strange or unusual at best, probably because believing is not under our direct voluntary control. However, I think this is not really the source of what is strange about the second scenario. One can remove this feature using a science-fiction example. Imagine I acquire a program that gives my computer access to my brain, specifically to the belief-forming module of my brain, say, via a thick cable plugged into my forehead. The program displays a list of propositions on my computer screen. By clicking the little “button” next to a proposition I can cause the program to cause me to believe that proposition. In this scenario, there is no in-principle problem with going through the list wanting, for example, to believe the propositions that will make me feel good about myself. Say I encounter the proposition that I am a better than average driver/lover. Will believing this make me feel good about myself? Yes, at least for a while. I want to believe it. Click. I believe it.

In this scenario, that particular source of oddness is removed. Here it makes sense for me to want to believe a given proposition under consideration, and my beliefs are (very close to being) under my direct voluntary control: I can turn on the light by flipping a switch; I can believe that I am a better than average driver/lover by clicking a button. In this science-fiction scenario, coming to believe, which normally isn’t much like an action, *is* much like an action, at least like an action of the ‘turning on the light by flipping a switch’-sort. In this scenario—and this is the feature I want to emphasize—my desires *can* guide (almost) directly how I form beliefs: I can believe because I want to believe (without subconscious subterfuge). Nevertheless, the case I am wondering about

remains odd. I go through the list of propositions displayed on the screen wanting to believe the propositions that are true. I encounter one that I believe is true. What would be the point clicking the button? And even if I do (as I probably would), nothing happens.

G. E. Moore once observed “that I cannot at any given moment distinguish what is true from what I believe” (1903, §80). Here is a final thought experiment by way of illustration. I present you with a list of propositions and ask you: Mark the ones that are true! You comply. Imagine now that, concerning the very same list of propositions, I had asked you: Mark the ones that you believe! You would have marked the very same propositions (which, of course, does not even begin to indicate that *being true* = *being believed*). I find Moore’s observation highly suggestive; it strikes me as being clearly related to the feature of the true-belief goal that I find puzzling; they are two sides of the same coin.

For the moment, though, I must leave it at that. I have not yet figured out something deeper to say on this topic.

Department of Philosophy
University of Notre Dame
Notre Dame, IN 46556
U.S.A.
e-mail: david.1@nd.edu

References

- David, M., 2005: On ‘Truth is Good’, *Philosophical Books* 46, 292–301.
- Leslie, S. J., 2012: Generics, *The Routledge Companion to the Philosophy of Language*, ed. D. G. Fara and G. Russell, New York: Routledge.
- Lynch, M., 2004: *True to Life: Why Truth Matters*, Cambridge, Mass: The MIT Press.
- Moore, G. E., 1903: *Principia Ethica*, Cambridge: Cambridge University Press.
- Moore, G. E., 1944: Moore’s Paradox, in G. E. Moore, *Selected Writings*, ed. T. Baldwin, London and New York: Routledge, 1993, 207–212.
- Nietzsche, F., 1886: *Beyond Good and Evil*; tr. H. Zimmern, Tribeca Books, 2012.
- Piller, C., 2009: Desiring the Truth and Nothing But the Truth, *Noûs* 43, 193–213.
- Russell, B., 1904: Meinong’s Theory of Complexes and Assumptions, reprinted in: B. Russell, *Essays in Analysis*, ed. D. Lackey, London: Allen and Unwin, 1973, 21–76.
- Sosa, E., 2001: For the Love of Truth?, in *Virtue Epistemology*, ed. A. Fairweather and L. Zagzebski, Oxford: Oxford University Press, 49–62.

Notes

1. Probably 'goal', 'desire', and 'want'-talk should be distinguished more than they usually are in this context; I will nevertheless follow widespread custom and be a bit cavalier about the distinctions.

2. Concerning the relations between 1, 2, and 3. Use 1 to form a statement ascribing the specified goal to *S*. We then have:

1. *S* has the goal to believe every true proposition.
2. For every proposition, *S* has the goal to believe it, if it is true.
3. For every proposition, if it is true, *S* has the goal to believe it.

There are no entailments here: 1 doesn't entail 2 or 3; nor does 2 entail 3; and 3 doesn't entail 2 or 1, nor does 2 entail 1.

3. ““I did that,” says my memory. “I could not have done that,” says my pride, and remains inexorable. Eventually—the memory yields” (Nietzsche 1886, §68).

4. If safety is read (improperly, by my lights) along the lines of wanting to be such that, if one believes *p*, then *p* is true, talk of belief or action guidance is of course entirely out of the question: *p*'s *being true* is not a belief or an action.