Nick Treanor
University of Edinburgh

The Proper Work of the Intellect

> The virtue of a thing is relative to its proper work....[and] the proper work of the intellect is truth….
>
> [Of] the intellect which is contemplative, not practical nor productive, the good and the bad state are truth and falsity respectively (for this is the work of everything intellectual).
>
> - Aristotle, *Nichomachean Ethics*, Bk VI, sec. 2.

There is a familiar teleological picture of epistemic normativity on which it is grounded in the goal or good of belief, which is taken in turn to be the acquisition of truth and the avoidance of error. This traditional picture has faced numerous challenges, but one of the most interesting of these is an argument that rests on the nearly universally accepted view that this truth goal, as it is known, is at heart two distinct goals that are in tension with one another. This paper will look more closely at the standard way of understanding the truth goal, drawing out both its explicit and implicit features. My aim will be to show that the standard way of understanding the truth goal is deeply mistaken, to propose and defend an alternative model, and to show how this alternative model restores the unity of the goal and its potential to ground and explain the normative dimensions of belief.

1. Preliminaries

Talk of *the* standard model of the truth goal might seem retrograde. Perhaps as recently as two decades ago, the goal could have been uncontroversially stated as that of believing what is true and not believing what is false. But these days this formulation has been edged out by a variety of more subtle constructions. The goal might instead be that of believing all of the true propositions and none of the false propositions *that we are capable of grasping*. Or perhaps the goal really concerns propositions *that we entertain*, or *that we are curious about*, or *that involve matters of practical importance*. There is, one might think, no standard model of the truth goal, but only a plethora of distinct and competing conceptions.

This line of thought is mistaken, however. Each of these is a conception of some quite different goal: the goal of believing the truth and nothing but the truth with regard to some

restricted domain of propositions. However interesting such goals may be, none is the truth goal, any more than the goal of being kind to every compatriot is the goal of being kind to everyone. In other words, although I will focus on an unrestricted goal and discuss "the standard model" of this goal, what I will allege is standard is not that the goal is unrestricted but rather a certain way of thinking about that goal. To say that the goal is unrestricted is merely to say which goal it is.

There are three reasons to focus on the unrestricted truth goal and on how it is understood. First, because in many cases philosophers have been concerned to articulate other, distinct goals because the prevailing understanding of the truth goal is deeply mistaken. I will defend this claim in more detail as the paper develops, but in brief, the standard way of thinking of the truth goal misconstrues what it is for a person to fulfill the truth goal to a greater degree, and consequently many epistemologists have found the truth goal implausible as the epistemic, or even as an epistemic, goal. Second, the features of the standard model of the truth goal that I will identify and criticize can also be found in models of other, restricted goals. I will not be able to elaborate on this claim in detail, but it should be clear from the discussion to come that the problems I identify are not unique to the standard understanding of the truth goal. If we can achieve a better understanding of that goal, we will be better placed to understand other, restricted goals we may be interested in for this or that good reason. Finally, as I said above, the standard model of the truth goal is one that leaves it unable to serve as the ground of an epistemic normativity worth having. If we understand the truth goal as we should, this problem dissolves.

2. The Standard Model of the Truth Goal

There are two main aspects to the standard model of the truth goal. The first is that the goal is taken to have two distinct elements, while the second is a certain picture of what it is to satisfy the goal to a greater and lesser degree. I will discuss each in turn, and then show how taken together they have serious consequences for the prospect of the truth goal serving as the foundation of epistemic normativity.

2.1 The Two Elements of the Truth Goal

On the standard understanding of the truth goal, it is taken to have two distinct elements. As

Marian David puts it:

> Let us characterize the truth-goal, somewhat loosely, as the goal of *believing truths and not believing falsehoods….* Note that the goal has two parts, a positive part (believing truths) and a negative part (not believing falsehoods). The label 'truth-goal' is less than ideal because it de-emphasizes the negative part. (2001: 153)

This claim has become a platitude that introduces discussions of the truth goal. It was not always

so. James had to argue for, or at least insist upon, the distinction between what David calls the

positive and negative parts of the truth goal:

> There are two ways of looking at our duty in the matter of opinion—ways entirely different, and yet ways about whose difference the theory of knowledge seems hitherto to have shown very little concern. *We must know the truth*; and *we must avoid error,* – these are our first and great commandments as would-be knowers; but they are not two ways of stating an identical commandment, they are two separable laws. (1969: 203)

As everyone recognizes, these dual elements can be written as a single goal:

> believe p if and only if p

or:

> believe exactly the true propositions

but this doesn't change the fact that, at least typically, they are thought of as two distinct goals:

> if p, then believe p
> if not p, then do not believe p

2.2 Satisfying the Truth Goal to Greater and Lesser Degree

The second aspect of the standard model of the truth goal is less explicit. That is that the degree to

which one satisfies it is given by, roughly, the ratio of true beliefs to truths balanced against the

ratio of false beliefs to falsehoods. We can put it more precisely as the idea, implicit rather than

explicit in the standard picture of the truth goal, that one satisfies the first element:

> if p, then believe p

to whatever extent the truths one believes comprise all the truths there are, and one satisfies the

second element:

> if not p, then do not believe p

to the inverse of whatever extent the falsehoods one believes comprise all the falsehoods there are. The basic idea is that there are some number of true propositions and one satisfies the first element incrementally as one comes to believe them; with each new true belief, a true proposition moves from the 'true but unbelieved' ledger to the 'true and believed' ledger. Similarly for the second element: there are some number of false propositions and as one comes to believe them, one satisfies the second element less; with each new false belief, a false proposition moves from the 'false and unbelieved' ledger to the 'false and believed' ledger.

These two aspects of the standard model of the truth goal raise or are bound up with two further issues. The first is that of weighing how much satisfaction of the overall, two-part goal is determined by how much one satisfies the first element and by how much one satisfies the second element. To say that one satisfies the first element of the truth goal more with each new truth believed, and that one satisfies the second element less with each new falsehood believed, leaves unsettled the issue of how these ledgers balance against each other. One might think the degree to which one fulfills the positive element makes a greater contribution to how much the goal is satisfied than does the degree to which one fulfills the negative element. Or one might think the degree to which one fulfills the negative element makes the greater contribution. Or one might think the elements are perfectly balanced, so that the degree to which one satisfies the overall goal is a simple average or sum.

It would be a mistake to think that only the equal contribution view is coherent. One might be attracted to the simple sum view in the case of the two elements of the truth goal because there is a very obvious symmetry or complementarity to the idea of believing what's true and not believing what's false. But we can recognise this while realizing it does not settle the issue. How much each element contributes is an additional aspect, something that is not internal to the goal so long as it is interpreted as that of believing what is true and not believing what is false. James recognized this, and recognized that how they are balanced makes a very big difference to the assessment of one's epistemic position:

Believe truth! Shun error!—these, we see, are two materially different laws; and by choosing

> between them we may end by coloring differently our whole intellectual life. We may regard the chase for truth as paramount, and the avoidance of error as secondary; or we may, on the other hand, treat the avoidance of error as more imperative, and let truth take its chance. (1969: 204)

This issue has been recognized, but there is another that is subtler and perhaps harder to see. Merely saying that the truth goal consists in the sub-goals of:

> if p, then believe p
> if not p, then do not believe p

leaves open whether the satisfaction or fulfillment curve of each of these elements is regular, and if it is, *why* it is that way. To illustrate what is at issue, suppose for simplicity there are 50 truths and at time t1 Priya believes none of them. She then comes to believe one of those truths, so she believes $1/50^{th}$ of the true propositions. A little later, she believes a second of the truths, such that she believes $2/50^{ths}$ of the true propositions. Then she comes to believe a third truth, so she believes $3/50^{ths}$ of the true propositions. Eventually she believes 25 of these truths, or half of the true propositions. (For simplicity we will ignore the negative element and assume that she has refrained from forming any false beliefs.) It is clear that in the scenario as described, Priya satisfies the truth goal more at the end than at the beginning. But there is still an intelligible question concerning whether her progress toward satisfying the goal was regular. That is, does believing an additional truth make the same degree of contribution toward how much one satisfies the truth goal independently of the degree to which one already satisfies the truth goal at the point of acquiring that belief? Or does the degree of contribution made by an additional true belief grow or diminish as one comes closer to satisfying the truth goal (and if so, is the rate of growth or diminution fixed or exponential)? Or is there perhaps an irregular pattern, or no pattern at all? In the same vein, one can also ask whether each true or false belief makes the same degree of positive or negative contribution as any other. Do some make more of a contribution and some less? Or does a person satisfy the truth goal more to such-and-such degree with each new true belief, regardless of which belief it is, and satisfy it less to such-and-such degree with each false belief, regardless of which belief it is? And finally, distinct from this, does a particular belief's contribution to a person's satisfying (or failing to satisfy) one of the elements of the truth goal vary depending on what else

the subject believes? These are all intelligible questions in want of an answer. Moreover, we need

not only an answer to these questions, but an explanation of the answer, a story that makes

intelligible why the answer is what it is.[1]

These questions have not been asked. We have already said enough to see, however, that the

standard model of the truth goal does take a stand on these questions, albeit an implicit and

undefended one. It takes the progress toward satisfying the truth goal to be regular, both in the

sense that any given belief makes the contribution it does regardless of what else or how much one

truly or falsely believes, and in the sense that each belief makes the same positive or negative

contribution as any other. This follows from the fact it construes progress toward satisfying the

truth goal by appeal to the ratios of true beliefs to truths and false beliefs to falsehoods. It isn't clear

what is supposed to justify this, as it is an additional element beyond the mere statement of the goal

as that of believing what's true and not believing what's false. To be sure, it might be natural to

think the degree of contribution is indifferent to the proposition believed and to what else is

believed. First Priya knew nothing, then she acquired one truth's worth of truth, then two truths'

worth, then three truths' worth, and so on. But natural is not the same as correct, and we should

want an explanation.

2.3 The Truth Goal and Epistemic Normativity

Both aspects of the standard model of the truth goal will be broadly familiar. Nonetheless it is

worth considering an example of the model at work in philosophy. The example illustrates not only

both aspects of the standard model, but the deep consequences of construing the truth goal in this

way.

Consider Wayne Riggs's paper "Balancing our epistemic goals" (2003). Riggs begins by

noting that "There seems to be little disagreement within epistemology these days about the view

that something like 'truth,' 'having true beliefs,' or 'gaining truth' is our major, perhaps even our

---

[1] One might find this demand unreasonable. If one of the sub-goals is {if p, then believe p}, then surely (one wants to say) a person satisfies that sub-goal more with each new true belief, and one satisfies it more to the same degree. But this is to insist on an answer to a question without being willing to ask it.

only, purely epistemic goal." (2003: 343) He then points to the first aspect of the standard model,

the duality stressed by James, saying:

> But, as James points out, we also value avoiding false beliefs (or perhaps better: we disvalue having false beliefs). Unfortunately, pursuit of these two goals can pull us in opposite directions. Concern to avoid having false beliefs naturally prompts skepticism and caution, while the desire to accumulate true beliefs urges us toward acceptance, though not to the point of recklessness. (2003: 343)

What follows is an extended analogy between cognition and a widget factory, which is worth

quoting at length:

> MegaCorp, Inc. is a multinational company specializing in the manufacture, sales, and service of widgets. MegaCorp has production facilities all over the world, each with its own quality control inspector to ensure that the standards handed down from company HQ are met.
>
> At this year's management retreat, the high officers of the company decided on two goals for the upcoming year. First, to put a great many *operative* widgets on the market, and second, to put no *defective* widgets on the market. In order to facilitate these goals, the CEO had a memo distributed to the quality control inspectors of all production facilities advising them of the goals and of their importance.
>
> However, the CEO was worried about how motivated the inspectors would be to follow the guidelines rigorously. Therefore, she sent out another memo, announcing that all quality control inspectors would be evaluated in the following year regarding how successfully they met the company's goals. After the year was over, the CEO looked over the performance records of the inspectors in order to evaluate them. Two inspectors, in particular, caught her eye. One of these, Mr. Nervous, was notable for having put no defective widgets on the market. Ms. Careless, on the other hand, an inspector at a different plant, allowed *thousands* of widgets to leave the factory floor that were defective. Here seem to be paradigm cases of inspectors that deserved promotion and dismissal, respectively.
>
> Just as the CEO was about to write memos to this effect, the Vice President in charge of Production barged irritatedly into her office. "Look at these production figures," he said angrily. The plant whose records he showed her were from the plant managed by Mr. Nervous. It turned out that he sent out only a single widget in the whole year! The sum output of his factory was 1 widget. "Here, this is more like it," the VP said with satisfaction. This time the VP showed her the records from Ms. Careless's plant. Her factory had sent out 100,000 widgets this year. The VP informed the CEO that this was more than any other plant had produced that year. "That inspector deserves a promotion for sure," said the VP on his way out of the CEO's office.
>
> Putting all this information together, the CEO discovers the following. Mr. Nervous allowed 0 defective widgets on the market, but also allowed only 1 operative widget. Ms. Careless, on the other hand, allowed 10,000 (let us say) defective widgets on the market, but also allowed 90,000 operative widgets. So which of the two has done the better job this year? While Ms. Careless is not, perhaps, a model inspector, neither is Mr. Nervous. It is an open question which of the two has done better for the company. It all depends upon the relative weights of the two goals that MegaCorp is pursuing.
>
> The parallels between this scenario and the epistemic evaluations of cognizers are obvious, yet instructive. (2003: 343-4)

As Riggs points out, a factory that produces widgets can be assessed by how many operative

widgets it produces and by how many defective widgets it produces. Because these dimensions of

assessment come apart, to assess how a factory is doing overall with regard to widget-production one needs to assign a specific weight to each dimension, and a principled reason for assigning the weights that one does.

As Riggs emphasizes, the parallel with epistemic evaluation is obvious. We can assess cognizers by how much they get right, and by how much they get wrong. Once we've done this, however, there is a further question: how to assess them overall:

> Finally, we come to the most fundamental question posed above. "What *is* the correct weighting of the relative values of our two cognitive goals?" Do we, in general, value having truths more than we disvalue believing falsehoods, or vice versa? Or are they completely on a par with each other? Where are we even to look for answers to such questions? (2003: 349, italics in original)

The wording seems to run together the question of what the correct weighting is with the question of what we actually value or care about. I take it though that the former question is his real focus and it is certainly ours.

Riggs argues that there are two strategies we could adopt to determine the correct weighting of the relative values of what he calls our two cognitive goals. We could accept that believing more truths and fewer falsehoods are the ultimate epistemic goals, and then look for an overarching, non-epistemic goal to ground a relative weighting of these epistemic goals (e.g., the goal of being able to act in ways that best serve our practical interests). Or we could look for a deeper, genuinely epistemic goal that is not essentially veridic or alethic, not ultimately the goal of believing the truth and nothing but (perhaps, he suggests, the goal of acquiring understanding is such a goal). On either strategy, we should notice, the truth goal retreats in importance, unable to serve as the foundation of an epistemic normativity worth having. On the first strategy the truth goal remains the ground of epistemic normativity, but epistemic normativity contracts severely, to the point where there is no epistemically normative correct answer to the question of whether one weighting of true versus false beliefs is better than another. On the second strategy, there may be an epistemically normative correct answer to the question of what relative weighting of each goal is best, but that answer won't rest on a connection to the good of believing more truth and less falsehood. If anything, as Riggs notes, on this second strategy the epistemic value or disvalue of

believing truths and falsehoods might rest on this other, non-veridic good.

Looking at this discussion by Riggs is valuable for several reasons. First, it illustrates the prevalence of both aspects of the standard model of the truth goal – that it consists in two distinct, independent goals, and that one satisfies each goal in regular increments as one acquires truths and avoids falsehoods. But the discussion also shows some of how the standard model of the truth goal leads us astray. I will defend this claim in more detail later, but note for now that Riggs's pessimistic reflections on whether the truth goal is suited to serve as the ultimate ground of an epistemic normativity worth having rest on his construing the truth goal as decomposing into two distinct independent goals, each of which is served incrementally by the acquisition or avoidance of beliefs. Finally, the extended analogy wherein human cognition is compared to a widget factory illustrates the firm grip in epistemology of a conception of belief on which how much a person believes truly, and how much falsely, is a matter of how many true and false beliefs that person has, just as the output of a factory consists in some number of good and bad widgets. This is part and parcel of the standard model of the truth goal and at the very heart of where it goes wrong.

3. How is the Standard Model Wrong?

In this section, I will argue that we have gone astray, and that the standard way of understanding the truth goal is deeply mistaken.[2]

3.1 Counting the uncountable

For simplicity, we will consider just the positive goal and focus on someone with no false beliefs. According to the standard model of the truth goal, that person fulfils the positive element of the truth goal to a greater degree to the extent that the true propositions she believes comprise all the true propositions there are. There are two different ways we could understand this. First, by appeal to the notion of a ratio: Take the set of true propositions and sort them into two subsets, the set of true propositions believed by her and the set of true propositions not believed by her, and then compare the cardinality of these sets. How she is doing, vis-a-vis the truth goal or at least its

_____

[2] The following two subsections draw on Treanor 2013, Treanor 2014 and Treanor 2018.

positive aspect, is determined by this ratio. Alternatively, we could simply look at how many of the true propositions she fails to believe: the fewer she fails to believe, the better, and the absolute measure of how she is doing vis-a-vis the truth goal (or its positive aspect) is given by the cardinality of the set of true propositions she fails to believe.

Whichever method we choose, however, we run into a problem: there are too many true propositions for this method to work. Consider, first, that if there is a set of true propositions, then its cardinality is at least as great as the cardinality of the real numbers. Hence there will be uncountably many true propositions the person fails to believe. But then it does not make any sense to talk about there being a ratio of true propositions believed to true propositions; the idea that the person's true beliefs comprise some extent of the true propositions is unintelligible. Moreover, since the cardinality of the set of true propositions she fails to believe does not change (it is always the cardinality of the reals), it is simply false to say that as she acquires true beliefs, the cardinality of the true propositions she fails to believe becomes smaller. In either case, although it may make sense to say that she believes more, or fewer, truths than she did before, it does not make sense to say that she believes more, or fewer, *of the* truths than she did before.

One may think this objection merely helps refine the account. What we should want, this response says, is an account that takes a cognizer to satisfy the positive element of the truth goal to a greater degree if and only if she believes a greater number of true propositions. What matters, in other words, is sheer numbers rather than ratio. This is more than a minor amendment however. It abandons the idea that a typical person, at a time, satisfies the truth goal to some degree. It allows that the person satisfies the truth goal more or less than some other person, but not that there is some degree to which she satisfies it herself. This would permit comparative ranking against other cognizers, but not an absolute standard vis-à-vis the goal itself. In other words, there would be no fact of the matter concerning how much she satisfied the goal. If the degree to which a person satisfies the goal determines the degree to which a person is an epistemic success, then there would be no fact of the matter concerning the degree to which she is an epistemic success.

But there is another problem with this refinement. Consider a cognizer who believes of each

point on some one inch line, that it is a point. Now consider a cognizer who believes, of each point on some one inch line, that it is a point, and of everything in the *Encyclopedia Britannica*, that it is true. The second cognizer comes closer to satisfying the truth goal. Yet the cardinality of the set of true propositions she believes is identical (and there isn't even a subset relation, so long as the lines are different). So the 'sheer numbers' approach fails to even capture the thought that she is doing better than the other cognizer vis-a-vis the goal.

The same points hold if we consider the negative sub-goal: if a proposition is false, then do not believe it. If there is a set of false propositions, then its cardinality is at least as large as the cardinality of the reals. Hence if you have 8 false beliefs and I have 408 false beliefs, the cardinality of the set of false propositions you have managed to avoid believing is identical to the cardinality of the set of false propositions I have managed to avoid believing. To be sure, you believe fewer false propositions than I do. But you do not believer fewer of the false propositions. Moreover there are pairs of possible cognizers each of whom believes continuum-many false propositions, and yet one is worse off vis-a-vis the negative goal than the other. One might believe of every point on some one inch line that it could have had a colour. Another might believe of every point on some one inch line that it could have had a colour, and that everything in the *Encyclopedia Britannica* is false. The cardinality of the set of false propositions believed by the first cognizer is identical to the cardinality of the set of false propositions believed by the second cognizer, but the first has done better vis-à-vis the truth goal.

3.2 What to Count?

The above arguments may strike some readers as unconvincing, in that they turn on considerations about the cardinality of infinite sets that have an air of mystery about them. I do not think this reaction is warranted, but let's consider, nonetheless, a distinct set of problems for the standard model of the truth goal. There are several ways to develop the line of argument. One way is to consider the notion of a truth or a falsehood as implicit in the standard model of the truth goal. What is the notion? There has been little explicit discussion of this question with regard to the truth

goal, but it is clear that the epistemologists who have thought about the truth goal and endorsed the standard model of it have in mind something like what is expressed by an ordinary declarative sentence of natural language. It will be useful here to draw examples from classic discussions of why, allegedly, truth is not our principal or only epistemic goal. Consider these well-known passages:[3]

> The most obvious pure epistemic goal is truth….But, in my judgment, truth is not the important part of the story….The trouble is that most of the truths that can be acquired in these ways are boring. Nobody is interested in the minutiae of the shapes and colors of the objects in your vicinity, the temperature fluctuations in your microenvironment, the infinite number of disjunctions you can generate with your favourite true statement as one disjunct, or the probabilities of the events in the many chance setups you can contrive with objects in your vicinity. What we want is *significant* truth. (Kitcher 1993: 93-94)

> What is the 323rd entry in the Wichita, Kansas, telephone directory? Who placed sixth in the women's breast stroke at the 1976 Summer Olympics? What was the full name of Domenico Scarlatti's maternal grandmother? (Goldman 1999: 88)

> If random telephone numbers do not elicit a wide enough yawn, consider a randomly selected cubic foot of the Sahara. Here is a trove of facts, of the form grain x is so many millimeters in direction D from grain y, than which few can be of less interest. (Sosa 2000: 49)

From these passages we can extract the following examples of what a truth or falsehood, for the purposes of reflecting on the truth goal, is thought to look like:

> There is a blue cube to my left
> It is slightly warmer in this room than it was a few minutes ago
> Snow is white v p v q v r v s (ad infinitum)
> The cell is the basic structural, functional and biological unit of all known living organisms
> The 323rd entry in the Wichita, Kansas phone book is 'Gale, Dorothy. 1900 Emerald Drive. 316-688-1945'
> Gold is an element
> Maryna Yurchenia of Russia placed sixth in the 100 metre women's breaststroke at the 1976 Summer Olympics
> Space and time emerged together 13.799±0.021 billion years ago with a fixed amount of energy and matter that has become less dense as the Universe has expanded
> Domenico Scarlatti's maternal grandmother was named Sicuramente Non Questo
> Grain of sand x is 4 inches in direction 289.452343 from grain of sand y

The problem, however, is that if this is what is meant by *a* proposition, or *a* truth, or *a* falsehood, then it is just implausible that one makes regular incremental progress toward or away from satisfying the truth goal as one comes to believe truths or falsehoods. Some clearly contain more information than others. For example, compare these propositions:

---

[3] Haack 1993 makes the same point, but without giving examples.

> Jane is female.
> Mark is a bachelor.

For simplicity, let's assume that 'Jane' and 'Mark' function purely referentially, having no descriptive content. It is clear that the second proposition says more about Mark than the first one says about Jennifer. Or, more generally, the second proposition says more (full stop) than the first. For just this reason, if both propositions are true, then coming to believe that Mark is a bachelor, from a position of total ignorance regarding Mark, makes more of a difference to how much one satisfies the truth goal than does coming to believe that Jane is female, from a position of total ignorance regarding Jane. The standard model of the truth goal would certainly mandate believing each of them. It's not easy to see, however, how it can capture the evident fact that believing the second one is better, vis-a-vis the most intuitive sense of what the truth goal is, than believing the first.

With this particular example, it's tempting to say that the standard model of the truth goal *can* capture the fact it's better – and not only capture, explain and predict. After all, one wants to say, when a person believes that Mark is a bachelor they ipso facto believe that Mark is male, that Mark is unmarried, and that Mark is an adult. In other words, they believe three truths[4] rather than merely one and that – just as the standard model would explain – is why they fulfill the truth goal to a greater degree. The problem with this response, however, is that its promise is illusory. For consider this list of ordinary truths:

> Jane is female.
> Mark is a bachelor.
> It is raining.

Intuitively, believing that Mark is a bachelor satisfies the truth goal more than does believing that Jane is female. The response we're considering alleges that it can explain this by pointing to the fact that the second truth is really (or at least involves) three distinct truths, whereas the first is merely one. But what of the third? How many truths is that? Should we say it is exactly one, on the grounds that it doesn't wear any complexity on its face? We should not, because we can also compare that truth with this one:

> There is precipitation now.

---

[4] Or 4, or 5, or 6, depending on the mereology endorsed.

If 'it is raining' were exactly one truth, then 'there is precipitation now', which seems to say less about the world, would seem to have to be less than one.[5] That can't be right.

The problem with the response we're imagining is that it is tempting only in a very small range of cases, those relevantly similar to the 'Mark is a bachelor' example. When we consider the full range of truths and falsehoods expressed by ordinary sentences of natural language, the promise dissipates. It isn't merely that we have no idea how to even begin deciding how many truths each of (e.g.) 'there is a blue cube to my left' or 'the cell is the basic structural, functional and biological unit of all known living organisms' contains. It is that it is hard to see that, even in principle, a person who believes them believes ipso facto some number of simple truths, in the sense that each of these truths is or decomposes into some number of simple truths.

This leaves us with a significant puzzle concerning the standard model of the truth goal. What is believing more truth or more truths supposed to be, according to that model? We know the answer is something that has to do with the cardinality of truths believed, each belief being a sort of widget produced by cognition. But we also know, after reflection, that insofar as we think of individual beliefs or individual truths by appeal to the individual sentences that express them – as we do and perhaps must – individual beliefs and individual truths are not equimagnitudinous. Just as individual sentences differ in how much they say about the world, the individual truths expressed by such sentences, and the individual beliefs we would use such sentences to attribute to ourselves and others, differ in how much of reality they represent. Hence the puzzle: According to the standard model of the truth goal, and setting aside the issue of weighting the positive and negative aspects, one satisfies the truth goal more when one has a greater number of true beliefs and a lesser number of false beliefs, or to put the claim slightly differently, when one believes a greater number of truths and a lesser number of falsehoods. But if we ought not to count truths and falsehoods as expressed in ordinary sentences of natural language – as only a little reflection tells us we shouldn't or even cannot – then what should we count? What is there to count? This is a serious problem.

---

[5] If one worries that 'there is precipitation now' actually says more, in that it includes the truth that there is at least one non-rain way of being precipitation, notice that this would prove the point I am trying to make, just in a different way.

4. What, then, is the correct model?

The standard model construes the truth goal as two distinct goals. It takes it that some other, external factor has to be imposed to generate a balance between them. And it fails, as a model, because when you drill down to the details, it isn't possible to sustain the idea that progress toward each sub-goal is a matter of the number of truths and falsehoods believed. Can we do better?

I think we can, and that there is a way of preserving the core idea behind the standard model of the truth goal while avoiding its mistakes. The correct model must, it should go without saying, be one on which believing truths and disbelieving falsehoods is central. But it has to do this in a way that makes intelligible how one can satisfy the goal to a greater and lesser extent.

Here is the model in outline: The truth goal should be construed not by appeal to the idea of maximizing the cardinality of the set of truths you believe while minimizing the cardinality of the set of falsehoods you believe. Rather, it should be construed by appeal to the idea of increasing the similarity between the world-as-represented and the world, or to put the idea in more familiar language, between your picture of the world and the world. To be sure, true and false beliefs matter, for generally speaking true beliefs increase the similarity and false beliefs decrease the similarity. But just as sharing property A can make more of a difference to how similar two objects are than does sharing property B, representing the world as being such that A is instantiated can make more of a difference to how much one's picture of the world resembles the world than does representing the world as being such that B is instantiated. The mistake we have been making is to assume that because satisfying the truth goal involves having true and false beliefs, there is such a thing as the number of truths and falsehoods we believe and it's the number of these that matters. But this is as wrong as the view that since being similar involves having properties in common, there is such a thing as the number of properties any two objects share and it's the number of shared properties that matters.

That's the sketch. It is difficult to develop a more detailed model because many of the central issues such a model would need to resolve are among the most deeply contested in metaphysics and philosophy of language. I have in mind here not only questions about the nature of similarity, about

what it is and what makes for it, but also questions about the nature of properties, about the relation between predicates and properties, and about naturalness. We can illustrate the guiding idea, however, with an analogy, which I offer as a better way to think about cognition than the widget factory.

Suppose MIT sponsors an engineering competition where the challenge is for teams to build an AI robot that will strive to duplicate some part of reality at 1:1 scale, say some 1 cubic mile as it was in the first minute of 2018 EST. For example, if the centre of the cube is located at the summit of Chimborazo, the point on the earth's surface furthest from its centre, at some fixed angle relative to the earth, what is relevant is everything within that cubic mile – the glaciers and igneous rock, the microorganisms, the lichen and cruciferous plant species, and all the cold and quiet air that surrounds the summit and its upper slopes. The competition task is to build a robot that will duplicate this bit of the world as it was in the first minute of 2018 EST, where each robot has some set amount of time, say 5 years, to do this task.

Fast forward five years. Three robots were entered and each has produced a physical object, or more precisely a dynamic system on a one-minute loop. Robot A has produced something that closely resembles a deflating football. Robot B has produced something that closely resembles the cubic mile centred on the summit of Chimborazo, with just the following differences: what is air is water, the rock is sedimentary rather than igneous, and in place of glaciers there is only mayonnaise. Robot C has produced something that closely resembles the cubic mile centred on the summit of Chimborazo, with just these differences: the mountain is a tiny bit smaller, say one millimetre less prominent, and the volume of air a tiny bit bigger to compensate. The judges at MIT have three prizes to award. A first place trophy, a second place plaque, a third place participation ribbon. I take it as very strongly intuitive that Robot C should get the trophy, Robot B the plaque, and Robot A the participation ribbon. The strange deflating football that Robot A produced bears little similarity to the cubic mile around the summit of Chimborazo; nice effort though. The mountain of just the right size, made by Robot B, but of the wrong kind of rock, covered in mayonnaise rather than ice, and surrounded by water rather than air, is much more similar to the cubic mile around the summit

of Chimborazo than is the deflating football, but still goes substantially wrong. The object that Robot C produced, in contrast, stands out as very similar, as very close to being a duplicate. It's not a perfect duplicate, to be sure, but boy is it close.

What makes it the case that this is how the prizes should be distributed? Is it the *number of properties* that Robots A, B and C got right, and the number that each got wrong? At first glance we may find this tempting, but it is a familiar and widely accepted point about similarity between objects or worlds that similarity doesn't reduce to the number of shared properties. Each of the objects the robots produced shares and fails to share infinitely many properties with the cubic mile around the summit of Chimborazo. Moreover, properties do not all make the same contribution to similarity – some make a much greater contribution than others, and some make perhaps no contribution at all. As Lewis memorably put both points:

> Because properties are so abundant, they are undiscriminating. Any two things share infinitely many properties, and fail to share infinitely many others. That is so whether the two things are perfect duplicates or utterly dissimilar. Thus properties do nothing to capture facts of resemblance…. Properties carve reality at the joints—and everywhere else as well. (1983: 346)

Consider, for example, an apple and an electron: they are quite dissimilar, but there are nonetheless infinitely many properties they share (being more than 1 second old, being more than 1.1 seconds old, being more than 1.11 seconds old, not to mention the infinitely many shared properties more gruesome than these). Consider now two oranges. They are quite similar, but they are much more similar by virtue of sharing the property of being a citrus fruit than they are by virtue of sharing the property of not being Jupiter. In the case of the robots, what we should say is that the robot that won the trophy did so because the object it produced was most similar, overall, to the cubic mile it was trying to duplicate. The robot that got the participation ribbon deserved to come in last because the object it produced was least similar. And the second-place robot's object was somewhere in the middle. To be sure, we have no good theories of what such similarity consists in; that problem is as old as philosophy. But it is still the right answer.

Let's move closer now to the issue that concerns us, that of how the truth goal should be understood. Suppose in future years the competition committee decides that to reduce the carbon

footprint of the competition, the goal will be changed from that of duplicating an object *in reality*, by producing a perfectly similar object, to that of duplicating an object *in thought*, by producing a mental representation that perfectly resembles it, or that gets it exactly right. In other words, the AI robot doesn't have to actually make a physical object, it only has to form beliefs about what the target cubic mile of the world is like. The cubic mile is again centred on the summit of Chimborazo and three robots enter the competition. One produces a mental representation according to which that cubic mile is represented as having exactly the properties that a deflating football has. A second produces a representation according to which that cubic mile has a mountain in it composed of sedimentary rock, topped with mayonnaise and covered in water. The third produces a mental representation that gets the cubic mile exactly right save a bit of inaccuracy about the height of the mountain and the volume of air. In other words, the three robots produce mental representations that correspond exactly, save being mental rather than real, to the physical objects produced the first time the competition ran. For every property each of those objects had, these robots respectively have beliefs that ascribe that property to the target cubic mile.

How should the competition be judged this time? Is it by appeal to how many truths and falsehoods each robot represents? The problem with this story is that if the original objects produced the first time the competition ran each share, and fail to share, infinitely many properties with the cubic mile that surrounds the summit of Chimborazo, then the current robots, with their mental versions of these objects, each has infinitely many true and infinitely many false beliefs about that cubic mile. The number of truths and falsehoods believed, therefore, doesn't capture the difference in the faithfulness with which each robot represents the target cubic mile. By that measure, they are doing equally well. Moreover, just as we recognise that the sharing of some properties makes more of a difference to how similar two objects are than does the sharing of other properties, we should recognise that the representing of an object as having certain properties makes more of a difference to how much one's representation resembles that object than does the representing of that object as having other properties. This follows from the fact the robots in this competition have just as many true and false beliefs as one another, for the difference between

them therefore has to be not *how many* true and false beliefs they have, but *which* true and false beliefs they have. Recall also the two oranges we considered earlier: They both bear the properties of being a citrus fruit and of not being Jupiter, but they are more similar by virtue of sharing the former than they are by virtue of sharing the latter. If you form a mental representation of one of the oranges that includes the fact that it is not Jupiter and I form a mental representation of it that includes the fact that it is a citrus fruit, then by believing it is a citrus fruit – that is, in correctly ascribing to it the property of being a citrus fruit – I increase the similarity between the orange and my picture of it more than you do by believing that it is not Jupiter – that is, by correctly ascribing to it the property of not being Jupiter. By learning it is a citrus fruit, I close in on what it is. By learning it is not Jupiter, however, you are still almost wholly in the dark about what it is; for all you know, it could be anything that is not Jupiter, as wildly different in kind and nature as those are.

Reflection on this competition does not yet tell us how the truth goal should be understood. We have one idea, that of increasing the number of truths believed and decreasing the number of falsehoods believed, and a different idea, that of increasing the similarity between object or world and one's representation of the object or world. Why should we think this alternative conception is a better way to understand the truth goal, rather than just something else, some different thing entirely? There are three reasons.

First, when we think about the truth goal in its most basic form, before it is interpreted by theory, it is the goal of getting the world right, of representing the world as it is and not as it is not. That notion more closely corresponds to the account on which the truth goal is a matter of increasing the similarity between your representation of the world and the world. This can be hard to see because we are very used to thinking of the truth goal in a way that is laden with theory, a way according to which serving the truth goal is a matter of having more true and fewer false beliefs. But this is why the example of the artificial intelligence competition is illuminating. We can make sense of the robots being unleashed on the world with the command to duplicate, as closely as possible, some part of the world. If that is their aim, then increasing similarity rather than maximizing the number of properties in common while minimizing the number of properties not

in common is the right way to understand their task. My claim now is that when we think about our most basic grasp on the truth goal, and about the environmentally friendly version of the competition, where the robots are asked to duplicate the cubic mile in thought rather than in reality, we can see that the robot who deserves to win that competition deserves to win because it has better served the truth goal. It was unleashed on the world, or more precisely a circumscribed part of it, and given the job of knowing it, of bringing the world or that part of it within its ken. One robot did a poor job of that, one a brilliant job of that, and one was somewhere in the middle. We *could* say that all three robots serve the truth goal equally well, and that, so far as the truth goal is concerned, the robot that represents the cubic mile as a deflating football is no better or worse than the robot who represents the cubic mile as it is save the mountain being a millimetre shorter and the volume of air correspondingly greater. We could say that, and insist that the goal of cognitive duplication, which distinguishes the robots, is something else, something that doesn't have to do with the truth goal. But this doesn't seem right. The cognitive duplication ideal just is the ideal of getting the world right, of representing the world as it is and not as it is not.

A second reason, closely related to this, is that the similarity or resemblance of a representation to what it represents is properly understood as a veridical or alethic notion. Similarity between objects, and similarity between representations and objects, are both a matter of likeness; in both cases it is a notion of matching or closeness of fit. The similarity that holds between objects is not a matter of veridicality, however, since one object doesn't represent another, at least in ordinary circumstances. But as soon as representation is brought into the picture, then the likeness or closeness of fit becomes veridical. To be sure, the only things that can be true or false, in the dominant sense of those words, are individual sentences, beliefs, or propositions. But there is a broader, more inclusive notion of veridicality, which includes but is not exhausted by this narrower notion. This is the notion of faithfulness or fidelity to an original, and it is just this notion that is captured by the similarity approach to the truth goal. This is why the AI competition is an apt analogy. The objects the robots produce in the first version of the competition each stand in a certain similarity relation to the cubic mile around the summit of Chimborazo. If we construe those

objects as representations of that cubic mile, then the similarity relation qua object becomes a resemblance relation that holds qua representation. Each object qua representation faithfully represents the cubic mile to the same degree that the object qua object is similar to the cubic mile.

The third reason we should think of this proposal as a way to understand the truth goal, rather than just some different thing entirely, is because it preserves the best part of the standard model of the truth goal while avoiding its mistakes. That model starts in the right place, by recognizing the central importance of believing truths and avoiding falsehoods. But it errs in elevating this starting point to the content of the truth goal itself, in construing the truth goal as nothing more than the goal of maximizing the number of true beliefs and minimizing the number of false beliefs. We know that cannot be right. The proposal that recalibrates our understanding of the truth goal toward similarity preserves the importance of true and false beliefs, just as an understanding of what is involved in one object being similar to another cannot but give central place to the sharing and failing to share of properties.

5. A Unified Goal

In this paper I explored the standard model of the truth goal, drawing out its explicit and implicit features, and showed how conceiving of the truth goal this way undermines its potential to ground epistemic normativity. I then argued that this model of the truth goal is deeply mistaken and defended an alternative way of understanding that goal, one that doesn't require an appeal to the number of truths and falsehoods believed but that nonetheless preserves the central place of true and false beliefs. To close the paper, I want to consider whether the truth goal so construed is a potential ground of epistemic normativity. To put the question more precisely, I will not here try to develop an account of epistemic normativity that is based on this conception of the truth goal. Instead, the focus will be on whether the proposed model of the truth goal is one that fractures that goal into two, as the standard model does, leaving it unfit to serve as the foundation of epistemic normativity.

At first glance one might think it fares no better in this regard, since it too affirms (as how

could it not) the importance of believing truths and avoiding falsehoods. True beliefs increase the similarity between the world and one's representation of it, while false beliefs decrease the similarity. The goal again has both a positive and negative element, therefore, and moreover these again pull in different directions. The positive element is indifferent to risk and urges believing, while the negative element is wary of error and cautions restraint. We need, as before, some other factor or principle to weight these elements or sub-goals against one another. This alternative construal of the truth goal, on this line of thought, may offer a better understanding of that goal, but it doesn't salvage the possibility that the truth goal could serve as the foundation of epistemic normativity.

This response is mistaken, however. Construing the truth goal as a matter of the similarity between a representation and what it represents does make intelligible how believing truths and not believing falsehoods are part of one unified goal, and moreover a veridic or alethic one at that. The parallel with similarity between objects is again instructive. The robots in the first version of the AI competition do not have two separate goals, those of creating an object that (i) is as much like the cubic mile as possible, while (ii) being as little unlike the cubic mile as possible, where they fulfil these distinct goals by giving the objects they create properties the cubic mile has while refraining from giving them properties the cubic mile doesn't have. They have, rather, one goal, that of creating an object that is as similar to the cubic mile as possible, and *that* involves creating an object that has properties the cubic mile has and doesn't have properties that the cubic mile doesn't have (where we remember that some properties make more of a difference than others). In the same way, on the model of the truth goal that I have defended, the goal is increasing similarity between a representation and what it represents, where believing truly contributes to that and believing falsely detracts from that. True belief and false belief on this picture are akin to opposing vector forces, with true belief a vector force pushing in the direction of greater similarity and false belief a vector force in the opposite direction, pulling away from similarity. Acquiring true belief and avoiding false belief, therefore, are oriented in the same direction, toward the same single end.

It was once orthodoxy that the proper work of the intellect was truth, at least as an unrefined

slogan. Over time, the details were filled in, and we moved too quickly from the evident fact that both truth and falsehood matter to a picture that rests on implausible assumptions about truth and falsehood being countable and about that countability yielding the correct measure on what is true. Moreover, this picture fractures the truth goal in a way that leaves it unable to do the job it had long been thought to do, namely ground and explain the normative dimensions of belief. The alternative I have offered here is intended, principally, to be a better model of the truth goal itself. But it also restores the promise of that goal, the potential for that unalloyed goal to be the proper work of the intellect.

Bibliography

David, Marian. (2001) Truth as the epistemic goal. In M. Steup (ed.) *Knowledge, Truth and Duty*, New York: Oxford University Press, 151-169.

Goldman, Alvin. (1999) Knowledge in a Social World. Oxford: Clarendon Press.

James, William. (1969) "The Will to Believe," in *The Moral Philosophy of William James,* New York: The Thomas Y. Crowell Co.

Haack, Susan. (1993) *Evidence and Inquiry*. Oxford: Basil Blackwell Press.

Kitcher, Philip. (1993) *The Advancement of Science*. New York: Oxford University Press.

Lewis, David. (1983) New work for a theory of universals. *Australasian Journal of Philosophy* 61:4, 343-377.

Riggs, Wayne. (2003) Balancing our epistemic goals. *Noûs*, Vol. 37, No. 2 (Jun., 2003), 342-352.

Sosa, Ernest. (2000) For the love of truth. In Linda Zagzebski & Abrol Fairweather (eds.), *Virtue Epistemology: Essays on Epistemic Virtue and Responsibility*. Oxford: Oxford University Press. 49-62.

Treanor, Nick. (2013) The measure of knowledge. *Noûs*, Vol. 47, No. 3 (Sept. 2013), 577-601.

----- (2014) Trivial truths and the aim of inquiry. Philosophy and Phenomenological Research, Vol. 84, No. 3 (Nov. 2014), 552-559.

----- (2018) Truth and epistemic value. European Journal of Philosophy, Vo. 26, No. 3 (Sept. 2018), 1057-1068.