

# Truths are Valuable, Truth isn't

Alexander auf der StraÙe

Department of Philosophy, Heinrich-Heine-Universität Düsseldorf, Germany  
aufderstrasse@hhu.de

## Abstract

This paper deals with the relationship that, according to some, holds between true beliefs and success. It argues for truth-theoretic minimalism. In particular, minimalism will be defended against a particular objection against deflationism raised by Michael Lynch. The paper denies that truth has any non-instrumental value in the sense that truth is pursued for its own sake. Moreover, the instrumental value of true beliefs will be explained in terms of psychological regularities of agents' 'correct' beliefs about the world, rather than in terms of truth as such. The argument concludes with the result that – in the strict sense – truth is valueless because truth is no genuine property. However, the value of individual true beliefs is acknowledged, insofar as they foster one's behavioural success.

## 1 Introduction

At least since Dummett's seminal paper, 'Truth', dating back to 1959, people have repeatedly remarked that deflationary theories of truth miss the *point* of truth – that truth is an aim. The question whether modern variants of deflationism lack the resources to account for this is still controversial. Among the more recent critics of deflationism, one of the most prominent ones is Michael Lynch, who took up Dummett's original line of argument and developed it further in several respects (Lynch 2004, 2009). He defends a position according to which truth plays a significant role in our everyday life. Truth, from his point of view, has value. We pursue truth for its own sake, he says. This paper aims at undermining this claim by arguing exemplarily against Lynch's particular position. I examine Lynch's argument as a typical example of an argument for the desirability of truth. I shall argue that truth-theoretic minimalism is compatible with the legitimate claims about the value of truth. In contrast to Lynch's view, then, a position will be defended according to which one may acknowledge the value of *particular* truths but can still deny that truth *as such* has any value. In a nutshell, this is because truth is no genuine property.

The following very rough overview will serve as a backdrop against which the relevant theories of truth can be positioned. Briefly speaking, the world of truth theories divides into deflationary views and inflationary ones. Deflationism about truth is a family of varied theories, all of which have in common that some form or other of Tarski's convention T

(T) "p" is true iff p<sup>1</sup>

plays a prominent role. "P" is a variable ranging over declarative sentences of a specified natural language (English, say) and "p" (that is, "p" in quotation marks) is a meta-level name for that very sentence. (For sake of simplicity, English serves double purpose here both as meta-

---

<sup>1</sup>This way of stating the schema ignores some technical niceties, such as the problem that it actually requires quasi-quotation marks. (For "p" here is only a variable.) Still, I adopt this notation since it is one of the most common ones, and the notational technicalities do not affect the present argument.

language and object language.) Depending on what one takes primary truth bearers to be, the schema will be relativised accordingly either to propositions, sentence types, beliefs or utterances. The exact status of the schema varies from theory to theory. Importantly, in their ‘metaphysical’ part deflationary theories deny that truth has an underlying nature that can be revealed or anyway determined.

Inflationism about truth, on the other hand, is often characterised as the denial of deflationism, i. e. in inflationary theories, convention T plays no significant explanatory role (which, of course, is not to say that it plays no role whatsoever). Moreover, truth is analysable on these accounts: an inflationary theory would state the basic properties that constitute truth. Lynch’s theory of truth is of this sort.<sup>2</sup> One of the properties he ascribes to truth is that it is good. It is this claim that I will be focussing on in what follows. In particular, I will defend truth-theoretic minimalism, which is one of Lynch’s specific targets (cf. Lynch 2004: ch. 7, therein esp. pp. 107–116, 2004a).

Minimalism<sup>3</sup> says (i) that truth has no underlying nature, (ii) that the truth predicate denotes only a logical property in virtue of being a predicate (but that it denotes no ‘real’ property), (iii) that all its uses derive – in a yet-to-be-specified way – from (our underived acceptance of) instances of the T schema, and (iv) that everyone who understands how the schema works knows the meaning of “true” entirely. The reason that there is a truth predicate at all in natural languages lies in its usefulness, the minimalist claims. It allows formulating otherwise ungraspable generalisations like “Theory T is true” (read: every single sentence of which theory T consists is true) and so-called blind ascriptions like “What Johnson thought this morning is true”, where truth is ascribed to an entity whose identity is potentially unknown to the ascriber (hence the name). Roughly, this is all there is to know about truth, according to minimalism.

Obviously, if minimalism about truth holds, truth is valueless. According to minimalism, the only reason to ‘keep’ the word “true” in English is because otherwise certain things would either become inexpressible – like “Everything the pope says is true” – or too uneconomical to express – like “The Gravity Theory is true”.<sup>4</sup> The truth predicate ‘behaves’, linguistically speaking, like any other old predicate. In this sense, it denotes a property – truth – because *every* predicate, in a minimal sense, denotes a property qua being a predicate. But since “true” is *not analysable* analogously to, say, “red” or “water”, we may say that it denotes only a logical property. To be sure, this claim needs independent justification, which, however, can’t be given here.<sup>5</sup> If we only have “true” at the semantic level but no corresponding entity denoted in the

<sup>2</sup>Note that this applies to Lynch’s earlier view as exemplified by his 2004 book as well as his more recent ‘alethic functionalism’ (Lynch 2009, 2012), according to which truth is a functionalistic property, multiply realisable by different properties in different discourse domains. Irrespective of the additional functionalistic analysis Lynch proposes, his latter theory qualifies to be handled on a par with other, more paradigmatic inflationisms (coherentism, correspondence, etc.), since, on his view, truth is similarly analysable in terms of a constituting property. The claim that x is true iff ‘x has a property that plays the truth-role’ (Lynch 2009: 72) is theoretically on a par with such claims as that x is true iff x is part of a perfectly coherent belief system; that x is true iff x corresponds to reality, and so on. Thanks to an anonymous referee for this journal for pointing out to me that this needed clarification.

<sup>3</sup>By ‘minimalism’ I mean the deflationary theory of truth defended most prominently by Paul Horwich (e. g., Horwich 2005, 2010). This is not to be confused with Crispin Wright’s completely different truth theory of the same name. For the latter, see, e. g., Wright (1994), and the elaborations in Wright (2003).

<sup>4</sup>Examples of the first kind would, as a matter of fact, become inexpressible because their “true”-rectified counterpart involves an infinite conjunction (If he says that God exists, then God exists; if he says that Europe lies south of Canada, then Europe . . . and so on), but infinite conjunctions can’t be expressed by us finite beings. In the latter case, “true” might *in principle* be eliminated, but only at a very high cost (in terms of time and cognitive effort): every time one justifies a certain claim by referring to the truth of the Gravity Theory one would have to recite each of its axioms.

<sup>5</sup>See Horwich (1998: 141–144).

world, then a fortiori this 'thing' – truth – can't bear any properties. In particular, it can't have value.

Now, this flies in the face of common sense. 'Of course', you might think, 'truth has value'. We all pursue the truth. We, especially philosophers and scientists, value truth for its own sake, no matter for what purposes it may turn out to be useful in the future. All this is basically right, and, at least at first sight, it may be doubted whether minimalism is able to explain these alleged 'facts'. But: as always, it highly depends on the details whether the objection actually succeeds. In the following paragraphs, I shall (i) demonstrate that truth does not possess all the properties you might previously have thought it does and (ii) show that all properties that may be reasonably ascribed to truth can be accounted for within a minimalist framework.

In detail, I will proceed as follows. In section 1.1, the 'ceteris paribus' condition in Lynch's truth norm will be discussed; followed by a brief review of the scope of the norm and the exact meaning of "good" in section 1.2. 2.1 analyses the main line of argument as a struggle about the correct 'directionality' of explanations (from generalisations to individual rules, and vice versa). Accordingly, in 2.2 the deflationist position on this issue will be defended. Section 2.3 is a brief discussion of which phenomena should be accounted for by truth theories, and which should rather be left to related theories. Eventually, section 2.4 argues that, in sharp contrast to instrumental beliefs, 'non-instrumental' beliefs are probably not desirable. A short conclusion (3) ends the paper.

### 1.1 Unqualified CP clauses

Lynch (2004) claims that it is good to believe what is true. He is aware of the problem that the truth predicate serves here only as a device of generalisation, as described above. So, he reformulates the same 'norm' without the predicate. In reconstructing (and undermining) Lynch's argument we need both formulations anyway, so I restate them here in their original form:

(TN) Other things being equal, it is good to believe that p if and only if it is true that p. (Lynch 2004: 108)

(B) Other things being equal, it is good to believe that p if and only if p. (109)

It is pretty obvious that both formulations amount to the same, since their respective right-hand side is derived by substituting "p" with "it is true that p" (and vice versa), which is just an instance of the T schema.<sup>6</sup> Before discussing Lynch's actual argument, let me very briefly point out some problems of the 'norms' as such.

First of all, there is the ceteris-paribus (CP) clause. Clearly, very often it is good to believe true things. But that means in reverse: sometimes it is *not* good to do so.<sup>7</sup> The CP clause tries

<sup>6</sup>More exactly, substituting "p" for "it is true that p" is an instance of the equivalence schema for propositions, i. e. Tarski's Convention T applied to propositions. (There are equivalence schemata for utterances and sentences as well.) Both Horwich and Lynch agree that propositions are the primary bearers of truth, and both work with the equivalence schema that is relativised to propositions. Accordingly, I shall assume that, on either account, TN is derivable from B, and vice versa, by substituting "p" for "it is true that p", and vice versa.

<sup>7</sup>Often, arguments to the effect that false beliefs might have positive effects refer to folk psychology. Here is a real-life example:

[... S]elf-deception tends to lead to positive beliefs about oneself, which in turn trigger the subsequent display of the winner effect [i. e., 'an increased ability to win fights and social conflicts following prior victories']. [... P]ositive beliefs about oneself, and the adaptive benefits that go with it, can be reached through self-deception, past wins, justified true belief, or any number of other sources. The fact that self-deception often leads to these types of beliefs shows that it, like the past wins, can offer some adaptive benefits. (Lopez, J.K. & Fuxjager, M.J. (2012), Self-

to capture this. But it is really just a fig leaf: it plays such an important role for the overall argument that it needs to be spelled out in more detail.<sup>8</sup> An example will help to illustrate what is at stake here: on my (the author's) desk, the keys lie to the right of the candle. So – by the T schema – it is true that the keys lie to the right of the candle. Hence – by TN – it is good to believe that the keys lie to the right of the candle (on the author's desk). Or is it? For almost everyone besides a handful of people this information is irrelevant. Considering the limited cognitive capacities of humans, it is in no reasonable sense 'good' for any of the other seven billion people on planet Earth to have this belief (not to mention beliefs about certain molecule constellations in far-away galaxies).

But maybe Lynch had in mind something different. Maybe the rule is intended to mean it is good to believe that the keys lie to the right of the candle rather than believing that they lie to the left of it (given that, in fact, the keys lie to the right of the candle), so that *if* you believe anything at all about the keys on the author's desk, then you better believe that they are wherever they really are. At least this is not what B (or TN, for that matter) says.<sup>9</sup> But even such a restricted reading of the norm is implausible. Of course, it is very often highly useful to believe falsehoods. And Lynch acknowledges this. This is why he embeds the alleged norm in a CP clause. But to spell out what the CP condition exactly consists in means to commit oneself. And this would reveal the deficits of the norm, for there are counterexamples to even any restricted variant of it. When it comes to norms the burden of proof is not on the side of those who deny their validity.

The difficulties surrounding the CP clause are widely acknowledged in the literature. (One of the most illuminating pieces of work in this regard surely is Heal (1987). For more recent criticism of the *ceteris-paribus* condition, see Coates (2009); cf. also Piller (2009).) The fundamental problem consists of two aspects. Firstly, there are unbelievably many *trivial* truths. On the one hand, the 'trivially' trivial ones include propositions like 'that the keys lie to the right of the candle on the author's desk', or the contents of perceptual states. On the other hand, there are trivial truths build by running logical operations on atomic truths: for example, the truth of 'that today the weather is nice', combined with any mathematical truth. The CP condition is primarily intended to handle these trivial truths.

The other aspect of the problem is that sometimes it is better not to believe the truth, even in cases where the truth in question is substantive or non-trivial. Lynch's own example

---

deception's adaptive value: Effects of positive thinking and the winner effect, *Consciousness and Cognition* 21, 322)

<sup>8</sup>David raises a similar worry against Lynch's strategy:

It is difficult to criticise claims with 'prima facie/other things being equal'-qualifiers: objections tend to receive the response that that's a case where other things aren't equal. (2005: 297)

For general concerns regarding CP qualifications in (scientific) theories, see Earman, Roberts & Smith (2002). CP laws — TN is such a 'law' in the relevant sense — seem to be ultimately flawed, for they can't be tested against *any* available evidence:

Consider the putative law that CP, all Fs are Gs. The information that x is an F, together with any auxiliary hypotheses you like, fails to entail that x is a G, or even to entail that with probability *p*, x is a G. For, even given this information, other things could fail to be equal, and we are not even given a way of estimating the probability that they so fail. (Earman et al. 2002: 293)

<sup>9</sup>Note that Lynch says at one point in the discussion of why believing truths is good: 'it is better to believe something when and only when it is true. Or more loosely: *it is better to believe what is true than what is false*' (2004: 13, emphasis added). This *suggests* a reading like: regarding a particular proposition, if someone is about to form a belief about this very proposition, it is generally better when the belief is true than when it is false. But such a reading is incompatible with TN and B, which are generalisations concerning all propositions, including trivial ones.

of how believing falsities may foster one's success is this: a talent-free climber who thinks that a particular summit is reachable is more likely to get further than the one whose relevant beliefs are accurate in this respect. This is a general phenomenon. Many people reach (certain of) their aims better or equally well when believing falsehoods. Accordingly, the CP condition must be conceived of as also excluding these cases.

Generally, the problem is that the 'ceteris paribus' condition is *unspecified*. Considering all trivial and not-worth-the-effort, non-trivial truths, it is arguably the case that: *other things being equal*, it is good not to believe the truth. This is certainly true (or could be true) if we are allowed to leave the CP clause unspecified, for in that case every justified instance of good, true beliefs would qualify as a case where not everything is equal.

Note that Lynch thinks that it is actually just the other way around:

[...] Sometimes, believing what is true isn't the best thing—some falsehoods might be better to believe in certain circumstances and some trivial or dangerous truths may not be worth pursuing all things considered. But these cases are the exceptions that prove the rule: *other things being equal*, true beliefs are worth pursuing. (2009: 12)

The idea here is that truths that are not worth pursuing are clearly exceptional and can, therefore, be excluded with the CP condition. Note that Lynch does not have a proof for this. And nor have I for the reverse when playing devil's advocate here. The difference, though, is that the burden of proof is on his side, for the burden of proof is arguably *always* on those who posit norms, not on those who deny them or remain neutral.

## 1.2 Cognitive Goods

The second minor issue with Lynch's norms is that he is rather silent on the semantics of "good". Good for whom, good in which respect? On this we only read:

In believing, we operate under the norm of truth: *other things being equal*, it is good to believe a proposition when and only when it is true. [...] I don't mean that it is necessarily morally better. Things can be better or worse, good or bad in different ways. Clear writing is an aesthetic good; tasty food is a culinary good; and believing true propositions, we might say, is a cognitive or intellectual good. (Lynch 2004: 13, emphasis omitted)

What is a cognitive good? As we have seen, it is not something that lets people reach their respective aims. That believing true propositions is a cognitive good might mean that we praise others for holding true beliefs (like when they have written clearly or cooked well). That is implausible, if anything is. Many of us do not (always) care whether others believe correctly. (Just to state the obvious: this might be true even if most of us care most of the time what others believe.) What, then, is a cognitive good? Or, in other words, in which respects is it good to believe true propositions? Unless this is specified in some more detail, Lynch's theory of truth is trivially false, because it is uncontroversial that believing true propositions *tout court* is not good.

In sum, these are the minor objections against Lynch's norm of truth. Firstly, the ceteris-paribus clause is undefined. Secondly, the scope of the norm is unclear. Does it apply to everyone? In which respect is it good for people to conform to the rule? What does "good" mean in this context? These problems undermine the plausibility of the account but do not affect the actual core of the argument against minimalism. It is to this issue that I now turn.

## 2.1 Directions of Explanation

The minimalist, remember, says that our concept of truth is constituted by our underived disposition to accept instances of the T schema. From this – it is granted – TN, the truth norm, can't be derived. In the above paragraphs we have seen some reservations one may have about the rule. However, let us now suppose that something like the truth norm in fact holds, i. e. that it is actually good for everyone to believe the truth. A first simple distinction suggests itself: the distinction between, in Lynch's terms, 'being instrumentally good' and the rest. Accordingly, Lynch offers a third norm, which he also claims to hold:

(BI) It is *more than instrumentally* good to believe that p if and only if p.  
(Lynch 2004: 111, emphasis added)

This helps understanding how Lynch conceives of TN. Applying the instrumental/non-instrumental distinction, TN is to be understood thus: it is instrumentally good to believe the truth (all else equal). The following can be considered common sense. Humans have certain aims they want to reach. They act according to the beliefs they hold. Simply put, having true beliefs increases one's chances to reach one's aims. For example, you believe, say, that swallowing green pills makes you immortal. It is true (let's suppose) that green pills make immortal. You want to be immortal. Therefore, you swallow green pills, which, by assumption, do make you immortal. Now, you have reached your aim because you truly believed that the pills would have the desired effect. The general phenomenon is: having true beliefs helps, on the whole, reaching one's respective aims. This is meant by 'being instrumentally good': true beliefs are good instruments to succeed in life.

The explanation above is fine as far as it goes. But: what *really* explains success in this case – in terms of causal efficacy – is not the *truth* of the proposition that green pills make people immortal. Rather, the explanation should recur to the *fact* (if it is a fact) that green pills make immortal. No truth is *required* to explain instrumental success.<sup>10</sup> Lynch challenges the claim that this direction of explaining things is correct. It is not, he argues, that a single belief that p helps reaching a single aim if and only if p (under relevant circumstances), from which it is then inferred – after some such occurrences – that, generally, beliefs that p help reaching one's aims if and only if p (viz. iff they are true). According to him, the exact opposite holds:

[W]hy do we accept the infinite list of little belief norms? Answering this question is crucial, because the nonminimalist [i. e. the inflationist] has a ready and obvious answer. The reason we should accept that it is good to believe that snow is white just when snow is white, and good to believe that Socrates was a philosopher just when he was, is that it is good to have true beliefs. What makes it good to believe a proposition is that proposition's *being true*. (Lynch 2004: 110)

Accordingly, he concludes that 'minimalists must either come up with some other explanation, or admit they can't explain every fact about truth' (2004: 110). Here we see that really, so to say, the 'direction' of explanation is at issue. We could reasonably ask back again: why

<sup>10</sup>As I will argue in section 2.3, a theory of truth is the wrong place to look for an answer to the question why some actions are more successful than others. Please note two important things in this context. Firstly, "fact" is usually defined in terms of 'true proposition'. That itself, however, does not undermine our claim – that truth itself is causally inefficacious – for the relevant point is that, in regard to any particular successful action, its success can be explained *without* invoking the notion of truth. Secondly, facts and truth figure prominently in the area of causation (cf. Schaffer 2008: section 1). In particular, facts (true propositions) are promising candidates for being the relata of causation. Yet, this way of talking can plausibly be traced back to the utility of the truth predicate as the only common device of formulating particular generalisations – a feature that is perfectly explicable in deflationary terms.

is it instrumentally good to believe true propositions? We may visualise the alternatives like this:<sup>11,12</sup>

$$TN \succ G(B\phi \leftrightarrow \phi) \succ G(Bp_1 \leftrightarrow p_1), G(Bp_2 \leftrightarrow p_2), \dots, G(Bp_n \leftrightarrow p_n);$$

$$G(Bp_1 \leftrightarrow p_1), G(Bp_2 \leftrightarrow p_2), \dots, G(Bp_n \leftrightarrow p_n) \succ G(B\phi \leftrightarrow \phi) \succ TN,$$

where “TN” is the truth norm, “B” stands for ‘belief’, “G( )” means ‘... is good’, and the respective second step is a schema that abstracts from particular beliefs. The first option is Lynch’s. First a cognitive agent believes that the truth norm (TN) holds, from which she then infers single belief norms (e.g., ‘It is good to believe that snow is white iff snow is white’). The alternative is to start from the observation that humans initially form *individual beliefs* that concern particular situations; they then *abstract* from these beliefs (“Every substitution instance of ‘ $\phi$ ’ in ‘ $G(B\phi \leftrightarrow \phi)$ ’ yields a truth”)<sup>13</sup>; in a final step, they *express* this by using the truth predicate, i.e. they endorse the truth norm.<sup>14</sup>

The two alternatives represent possible ways of explaining the genesis of beliefs. By the same token, they represent possible ways of explaining why cognitive agents are justified in having these beliefs. As will become clear in the next section, the important bit is that particular beliefs can serve as an explanatory endpoint, whereas general norms can’t. The idea is that people can’t be described as being justified in having certain beliefs if there is no plausible story about how their beliefs evolved in the first place. I will come back to this further below.

In order to argue for minimalism, I shall now show why the second alternative is the preferred option. The answer is twofold: I should like to start with detailed comments on the ‘order of explanation’ issue, before I will formulate some general remarks on the scope of a theory of truth.

## 2.2 From Individual Beliefs to General Norms

Lynch thinks that the truth norm – “Other things being equal, it is good to believe that p if and only if it is true that p” – explains why people tend to form ‘little belief norms’ that then guide their action.<sup>15</sup> For the moment, we only consider instrumental value. Think of the ‘green pill’ example again. Lynch would be inclined to say that what *eventually* explains our success in becoming immortal is the fact that TN holds. But that is absurd. The reason people become immortal (if they do) is *because green pills make immortal*. It is because of this causal chain between pills and their corresponding effects that people become immortal if they

<sup>11</sup>These alternatives illustrate two kinds of explanantia of one single explanandum. They are both supposed to explain the behaviour of people (‘us’, in Lynch’s terms), which is – by assumption – such that people act *as if* they endorse the truth norm. So, although both ‘directions’ culminate in answering different questions – ‘Why do people accept infinitely many little belief norms?’ vs. ‘Why is believing true propositions instrumentally good?’ – they have in common that as a whole they both explain the same pattern of behaviour. Thanks to an anonymous referee for this journal for pushing me in this direction.

<sup>12</sup>An alternative way of formalisation would be to restrict the scope of “G” such that it only aligns the goodness of beliefs about  $\phi$  with  $\phi$ :  $G(B\phi) \leftrightarrow \phi$ . This will leave unaffected the general line of reasoning that follows below. Both variants are legitimate formalisations that can be read off Lynch’s truth norm.

<sup>13</sup>Note that it already becomes obvious at this stage that the need to *formulate generalisations* is one of the main reasons for introducing the truth predicate in the first place (cf. Horwich 1998: 4, footnote 1, 31–33, and 122–125). The reason is that norm B, as opposed to TN, quantifies over sentences (substitutional quantification) rather than objects, which is unusual. Given only object quantification, the truth norm is the only way to generalise from such individual convictions like, say, that it is instrumentally good to believe that snow is white iff snow is white.

<sup>14</sup>The ‘ $\succ$ ’ indicates an asymmetric relation. Here it represents (from left to right) explanatory priority, i.e. it represents which belief state explains the genesis of which further belief state. Note that the schematic ‘ $G(B\phi \leftrightarrow \phi)$ ’ corresponds roughly to Lynch’s norm B, cited above, which, by applying the equivalence schema for propositions to its right-hand side, may very easily be converted to TN.

<sup>15</sup>Similar to Lynch, I do not mean that subjects literally believe that these norms hold, in the sense that they explicitly subscribe to them. It suffices that they behave in such a way that their actions show certain regular patterns that are as if these subjects would be following the norms in question.

swallow green pills, if anything. Or, to put it differently, the causal chain between pills and immortality – in conjunction with specific beliefs and wishes – explains the *success* of actions.

The corresponding belief norm – “It is instrumentally good to believe that green pills make immortal iff green pills make immortal” – only holds (if it holds) because the belief that green pills make immortal can – together with the wish to become immortal – lead to a particular type of action (e. g., swallowing green pills), which is especially successful in exactly those situations in which swallowing green pills causes immortality. Note also that it is this *particular* causal connection between pills and immortality and the *particular* success of one’s belief about pills and immortality that justifies the further belief that the belief that green pills cause immortality is instrumentally valuable iff green pills cause immortality. One’s success, in this example, consists in becoming immortal. And if it was not, given the assumptions, the *pills* that caused this success rather than the *truth* of ‘that green pills cause immortality’, we would be leaving the grounds of rational discussion.

In a nutshell, what explains instrumental success is, eventually, the contents of our beliefs, and not the truth of the propositions or sentence types used to describe them. Or, more precisely, it is truth that explains success, but in an absolutely innocent way. The truth of “Snow is white” consists solely in snow’s being white. This is what is expressed by the T schema. So, in this way it is the truth of “snow is white” that explains the instrumental value of certain beliefs involving snow and whiteness. But this way of talking about things leaves unaffected our claim that what is causally efficacious when it comes to instrumental success is not truth – neither truth as such nor particular truths – but the contents of our beliefs.<sup>16</sup>

Stated thus, Lynch’s argument seems question-begging. Why should his theory be regarded as an explanatory endpoint? Why not explain the validity of TN in terms of its instances in reverse? To be sure, in terms of generality his explanation is better off. One very general assumption accounts for the vast range of ‘little belief norms’, of which we conceded, if only for the sake of argument, that they are valid. But generality is not everything.

When it comes to norms, Lynch speaks of ‘us’ accepting certain norms and that ‘we’ follow certain rules. ‘We’ is probably ‘we, human beings’ or ‘we, the average human’. That we accept TN – which we take for granted for the moment – must mean that judging from the observable (including verbal) behaviour of humans, one may conclude that humans implicitly accept this rule in that they act accordingly. Descriptively speaking, there is no difference with respect to the observable behaviour of cognitive agents relative to whether they endorse TN itself or its corresponding instances.<sup>17</sup> Hence, *theoretical* considerations decide between these two otherwise equivalent ways of description (armchair reasoning, if you like).

We assume that knowledge of TN is not innate. Now consider the following arguably simplified story of how we come to learn that true beliefs are instrumentally valuable. From a certain age, children start acquiring beliefs about the outer world. From time to time, they decide which action to take according to the beliefs they hold (in order to reach certain of their aims). Now, at some point in their life they start experiencing that not all of their beliefs correspond(ed) to reality (i. e. they were false). Ex post, they begin to realise that their actions (in terms of the desired outcome) have been structurally more successful whenever

<sup>16</sup>Horwich argues convincingly that in the context of instrumental value the sole reason to utilise the truth predicate is its function as a generalisation device (cf. Horwich 2006).

<sup>17</sup>An interesting difference between the individual belief norms and the general norm TN concerns confirmation. The ‘validity’ of individual norms may confirm TN, but not the other way around, as only instances can confirm laws (e. g., Hajek & Joyce 2008). In line with this, I assume that individual belief norms can *explain*—both in terms of genesis as well as in terms of justification—why TN holds (if it holds), but not the other way around.



they ‘correctly’ believed that  $p$  iff  $p$ . After some such experiences, children might begin to think: ‘Uh, maybe this is a general phenomenon, and maybe it’s *always* (or most of the time, at least) the case that I’m going to be more successful if my beliefs are correct’. This is to say, they acquire the generalised belief that, in general, it is (instrumentally) good to believe that  $p$  iff  $p$ . Following common deflationary explanation, they express this (if they do) by using the truth predicate, i. e. they endorse TN.<sup>18</sup> This line of explanation is in accordance with the picture according to which we move from single beliefs to a generalised belief about instrumental value of true beliefs:

$$G(Bp_1 \leftrightarrow p_1), G(Bp_2 \leftrightarrow p_2), \dots, G(Bp_n \leftrightarrow p_n) \succ G(B\phi \leftrightarrow \phi) \succ \text{TN}.$$

Why did we sketch this – arguably highly simplified – picture of how TN might evolve? Because this is what required explanation, given that TN in fact holds (which is the assumption we started with). The important thing to note is that the *basis* of this explanation is individual beliefs, individual actions, and the correlations between successful actions and correct (i. e. true) beliefs that cognitive agents experience.

On Lynch’s account, TN is the basic explanation with which we begin. But unlike individual beliefs, we can’t take TN as a starting point. We assumed that (the belief in) TN is not innate.<sup>19</sup> For beliefs, we have more or less plausible stories of how they evolve.<sup>20</sup> There is no similar story for TN.<sup>21</sup> The alleged ‘advantage’ of TN is that it may explain why all true beliefs are instrumentally valuable. But if we *start* our explanation by postulating the norm, the supposed advantage seems to be outweighed by its disadvantages. The problem is that, leaving innateness aside, we have no idea of how TN might crop up in humans. However, if, alternatively, we *end* our explanation with TN (roughly along the lines just sketched), most ‘advantages’ of the rule remain. The story above ends with the belief that TN holds. The alternative does not even get started because it does not have the resources to explain how TN evolved in the first place.

One might object that the developmental alternatives just sketched do not affect the *justification* of the truth norm and the individual beliefs, respectively.<sup>22</sup> For two reasons this objection does not succeed. Firstly, we have assumed that TN holds. This means that people are right in thinking that true beliefs are instrumentally good. As shown above, truth is not required in order to *explain* instrumental success. (It is only required to express corresponding generalisations.) Hence truth is not required to *justify* beliefs about instrumental success. The ultimate justification for individual beliefs about instrumental success is rooted in the fact that people are able to notice correlations between the truth of their action-guiding beliefs and their success. Noticing these correlations is a plausible reason for people to believe certain things about particular instrumental successes. Asking for further justification would be beside the point. However, there is no such corresponding reason for the explanatory endpoint in Lynch’s scenario, i. e. for TN. Hence, even if the focus is restricted to justification-related issues, the theory starting with individual beliefs and ending with TN is better off since its endpoint (individual beliefs) is well justified.

<sup>18</sup>This way of putting things is pretty much inspired by Horwich’s line of reasoning (1998: 44–46, 2001, 2006, 2010: 57–77). McGrath (2005) also discusses the relationship between TN and B. Similar to the present account, he argues that the instances of norm B are, relative to one’s theory of truth, explanatory basic.

<sup>19</sup>I doubt that Lynch would go as far as denying that TN is not innate in order to save his thesis.

<sup>20</sup>See any good textbook on mental representation.

<sup>21</sup>Apart from innateness, the only possible way out is to assume that the belief that TN holds has proven to be evolutionarily adaptive. Even if that was the case, we would still need to account for *that* by recurring to individual norms that, on an evolutionary time-scale, gave rise to this belief in the first place.

<sup>22</sup>This possible response was pointed out to me by an anonymous referee for this journal.

Secondly, even if Lynch's theory is read as dealing only with justification issues rather than with developmental aspects, the presented sketch excels his theory. Here is why. Both directions of explanation are pretty similar in two important respects. The direction starting from individual beliefs *ends with TN*, which is to say that if TN plays any role in justification, this is in line with this approach. For example, it might be that, once people have acquired TN via the developmental process that I have proposed, they then justify their individual beliefs about instrumental goods by appealing to TN. The present theory sketch explains why they are *prima facie* right in doing so. Moreover, since both directions of explanation are *equal* in that they both posit TN and in that they both are indistinguishable in terms of observable behavioural patterns they imply, something else must decide between the two. One further such 'something else' is that, *taken as theories that are solely concerned with justification*, one is compatible with a plausible developmental story whereas the other one is not.

We have seen that when it comes to instrumental value of true beliefs, the debate actually is about the appropriate order of explanation. Also, we have seen that taking TN as one's basic explanans is rather problematic. There are even more general considerations that undermine Lynch's approach to truth. It is to these that I now turn. Before one starts construing a theory of truth, one should become clear about what the theory is supposed to be about, what it needs to achieve to be 'successful', and, in general, when it can be considered to be adequate.

### 2.3 The Primacy of The Phenomena

A very general condition on theory building is what I dub the primacy of the phenomena. A theory, supposed to deal with a certain specified subject, needs to account for the relevant phenomena. In the case of truth, this means that we would begin construing our theory by asking: what are the truth-related phenomena? First of all, there is 'truth-talk', i. e. the fact that our everyday discourses involve the truth predicate. This, I take it, is *the* truth-related phenomenon.<sup>23</sup> It is legitimate to require of a theory of truth that it covers only those aspects of truth-talk that otherwise could not be accounted for, because that is the sole reason for having a truth theory in the first place. Almost everything involving the truth predicate may be explained without referring to truth. In particular, the success of true beliefs may be accounted for by psychology.

Leaving the status of generalisations aside, there are basically two options: saying that it is instrumentally valuable to believe that, e. g., pigs are animals iff *pigs are animals*, or to say that it is so valuable iff *it is true that* pigs are animals. And here we clearly see which option is preferable: we may explain the instrumental value of this belief (or any other) in terms of a psychological story of how beliefs concerning pigs guard one's behaviour in regards to pigs and how believing they are animals helps interacting with them if they really *are* animals. An accompanying theory of truth is only needed to explain our common way of formulating certain generalisations such as TN – just as the minimalist supposes.

The alternative – explaining success in terms of the truth of beliefs – is simply superfluous, since there is this equally general explanation that does not involve truth. Only if no alternative was available, we would need to use 'truth' in explaining things. It is basically the same line of argument that motivates refuting the redundancy theory of truth (if only the other way around; i. e. the theory covers *less* than required by the phenomena). If we only had the redundancy

<sup>23</sup> Acknowledging this is clearly not tantamount to saying that truth talk is the only relevant phenomenon. It is just that truth talk is the most *obvious* truth-related phenomenon, and truth talk is the only *universally accepted* such phenomenon. Still, minimalism, or deflationism more generally, is no linguistic theory but a proper theory of truth. In particular, it is a *metaphysical* theory. (Minimalism says that truth is only a logical, non-genuine property.)

theory of truth – i. e. only a theory that implies that every sentence of the form “T(‘p’)” is identical in meaning to “p” – then a whole bunch of sentences would remain unexplained. And for just these kinds of sentences – generalisations and blind ascriptions – one needs to ‘extend’ one’s theory, which is to say that one ends up with some variety or other of modern deflationism.

Unlike the arguments before, this argument does not presuppose that Lynch’s account of truth is a false description of what is going on. Rather, the claim is that his approach goes against the principle of primacy of the phenomena. It is a *psychological* phenomenon that true beliefs foster one’s success in reaching aims, but it is a *truth-theoretical* phenomenon that sentences involving the truth predicate are not identical in meaning to their “true”-rectified counterparts. And of course we *could* construe a theory of truth that was general enough as to also cover facts of the former kind. But since in this area we already *have* successful theories, it would be wrong to demand of a theory of truth to account for this aspect as well.

Most people probably tend to think that something like TN holds. Assume it does. How can the deflationary view account for this? We have seen that there are theories that explain the relation between instrumental success and true beliefs (or maybe combinations of theories involving theories of representation, action, and causation). We also have a minimalistic theory of truth that explains all truth-related phenomena. This latter theory does not imply TN. But it is *compatible* with the norm, for it explains why TN – although truth is no genuine property and hence is inefficacious – is the most natural way to *express* the relationship between success and true beliefs, albeit that the *eventual explanation* for this relationship comes from somewhere else (probably from cognitive psychology).

## 2.4 Deep Normative Facts

What about the even stronger norm BI? Remember that this norm says that

(BI) it is more than instrumentally good to believe that p if and only if p.

Lynch claims that here we may see even clearer that minimalism lacks the resources to explicate the value of the ‘cognitive good’, true beliefs. I admit it really does. But unlike TN (and B), in the case of BI it is absolutely controversial whether the norm holds at all. We have already seen in the beginning that TN is formulated quite imprecisely. The CP clause is unspecific; the scope of the norm is vague; and we do not know what “good” is supposed to mean exactly, to name but the most obvious difficulties. The purpose of comparing TN to BI was to get a better grip on the meaning of “good”. Assuming that the phrase “instrumentally good” is comparatively unproblematic, that was a success. (We largely ignored the ‘*ceteris paribus*’ issue because that would have led us too far afield.) With these provisos, TN seems plausible. BI, however, is much too underspecified to be even evaluated.

What does “more than instrumentally good” mean? On the most charitable reading of the argument, it means that people pursue truth for its own sake, i. e. irrespective of the (potential) instrumental value a true belief might have.<sup>24</sup> Lynch cashes out the validity of norms in terms of actual behaviour. That is to say, a given norm is said to hold depending on whether people in fact act in accord with the norm:

[...] (TN) can’t be derived from the purely nonnormative (T). Therefore, contra minimalism, that schema cannot fully capture everything *we believe* is true of truth. It can’t capture all the *facts about truth*. (Lynch 2004: 109, my emphasis)

<sup>24</sup>This is not to say that this is a sufficiently detailed description of what “more than instrumentally good” means. Even with this bit of rephrasing it is still unclear what the norm is about, and whether people believe it.

BI is a norm concerning beliefs; it says that ‘we’ pursue truths (true beliefs) for their own sake. But we do not. There are *unbelievably* more propositions about which most of us never form beliefs – and never want to. Do ever people value the true belief that Kim Kardashian is a superstar? If yes, one may very easily think of further, uncontroversial examples, the number of which will trump by far the number of beliefs people actually value.

This concerns trivial truths as well as non-trivial ones. In effect, everything said in section 1.1 on ‘*ceteris paribus*’ conditions applies, *mutatis mutandis*, to BI as it applied to TN.<sup>25</sup> So I will not repeat those objections here. Rather, I shall evaluate two – arguably still tentative – clarifications by Lynch as to what BI amounts to.

The first argument is this:

One of the lessons that I take from [Harry] Frankfurt’s work is that there are at least two ways something can become important or more than instrumentally good for our lives—and therefore worth caring about. [... The second one] is this: something can become worth caring about for its own sake because the very act of caring about it for its own sake is good. [... Art may well have intrinsic worth; but coming to care about art can make art worth caring about. For caring about art [...] can all by itself imbue human life with meaning. Art is important to us, as we might put it, partly because its being important to us is important for us. [...]

We can show that caring about the truth as such and for its own sake is part of a flourishing life. And that, surely, is enough to make truth worth caring about. (Lynch 2005: 335, emphasis original)

The argument seems to be that believing that truth is (more than instrumentally) good fosters flourishing lives. I shall not argue against this claim. The reason is that Lynch, claims to the contrary notwithstanding, considers the *instrumental* value of beliefs in this passage. Even if what he says in the last two sentences cited would be true (which I have doubts about), this is irrelevant for BI. A flourishing life is an aim that maybe can be reached by pursuing truth for its own sake. But that makes the pursuit of truth for its own sake (only) instrumentally valuable insofar that it helps reaching that aim.

A second argument in favour of non-instrumental value runs as follows:

[...] Consider Russell’s scenario, that without our knowledge, the world began yesterday [...]. If we really lived in the Russell world, as I shall call it, almost all my beliefs about the past would be false. [...] When I now think about the worlds [actual world and Russell world] in so far as they are identical in instrumental value, there is a difference between the two worlds that matters to me. Even when it has no effect on my other preferences, I, and presumably you as well, prefer true beliefs to false ones. (Lynch 2004a: 503, emphasis omitted)

<sup>25</sup> Again, cf. Piller (2009) for some convincing illustrations of situations in which, even if substantive issues are at stake, it might be good (or better) to believe falsehoods than to believe the truth. Note that in addition to ‘situational’ reasons for why believing falsehoods is sometimes preferable there might even be, if you like, ‘structural’ reasons. For example, some people think that ex-post rationalisation is an evolutionary advantage. If that is true, then there are *general*, i.e. evolutionary, reasons for favouring falsehoods over truths in certain circumstances. To be sure, evolutionary advantage is something like an ‘aim’. But if believing the truth is sometimes evolutionary disadvantageous, then it can’t possibly be that believing the truth is more than instrumentally good. (I am assuming that evolutionary disadvantages are bad.)

If this argument is sound, it shows that truth is non-instrumentally good. Moreover, if it is sound, it shows that minimalism about truth is false, for minimalism lacks the resources to explain the non-instrumental value of beliefs.<sup>26</sup> Fortunately, the argument is unsound.

First of all, the alleged fact that Lynch and his readership prefer true beliefs is insufficient reason to assume that true beliefs are more than instrumentally good. But leave that aside. The argument apparently is based on plausibility considerations. Lynch thinks that his readers, like himself, plausibly prefer living in the actual world over living in a Russell world. I doubt that. We live in the actual world, and the actual world, as Lynch assumes, is no Russell world. Even worse, the Russell scenario is so far-fetched that our theory of truth can't possibly be based on intuitions concerning such a world. Lynch's argument, as I see it, is based on the claim that most (all?) people have the intuition that living in a non-Russell world is better than living in a Russell world.

However, our intuitions are shaped by the world we live in. (Let us ignore cultural and social differences for the moment, which would further complicate matters.) In particular, with respect to many of our intuitions we are trained relative to what is possible in our world. Moreover, our intuitions concerning truth are to a large extent shaped by the specific instrumental successes of true beliefs. The Russell world does not show that believing the truth is more than instrumentally good, for, if truth and falsity do not affect instrumental success of beliefs (as, by assumption, they do not), then there is no reason to favour the truth in the first place.

Probably many people would agree with Lynch that living in the actual world is better than living in a Russell world. That is absolutely plausible, given that their intuitions are shaped by experiences in a world where truth and falsity make a difference. In a Russell world, very many of our beliefs are false. Most theories of physics are false in a Russell world: there has never been a Big Bang, for instance. Still, the universe looks exactly as implied by the Big Bang Theory, i. e. the accuracy of the Big Bang theory in regard to the observable phenomena is the same in both worlds. In other words, the relationship between truth and *instrumental* success is completely disentangled in a Russell world. (That is the point of the thought experiment.) Since our intuitions are shaped by the actual world in which truth and falsity make a difference, these intuitions can't be expected to yield any *reliable* results if applied to scenarios in which that connection between truth and instrumental success is lacking.

Note that Lynch's thought experiment has the interesting result that, in a Russell world, the instrumentally successful beliefs are false ones. The successful beliefs are exactly those beliefs that would be true in the actual world (our world). For example, if I want a hard-boiled egg in a Russell world, then the most efficient way to achieve this would be to believe that the egg-cooking theory of the actual world is true.<sup>27</sup> This applies to all beliefs in a Russell world,

<sup>26</sup>This does not undermine the plausibility of minimalism, for Horwich's argument in favour of (something like) BI is unconvincing anyway. He argues for the non-instrumental value of truth by saying (i) that 'there is a widespread sentiment that certain items of knowledge are desirable regardless of any practical use', (ii) that it would otherwise 'be hard to justify our pursuit of truth in fields of inquiry such as ancient history', and (iii) that 'it is surely no less important to pursue truth [...] in normative domains' (Horwich 2010: 65, emphasis omitted). Ad (i): that a *conviction* is widespread does not make it right. Ad (ii): ancient history as a field of inquiry *is* (relatively) hard to justify; justifying its pursuit for truth, on the other hand, is not (in terms of instrumental value). Ad (iii): there is no proof that people do pursue (or should pursue) truths in normative domains *independently from any* instrumental value. Interestingly, Horwich (2013) more recently admits that at least *the belief* that truth is non-instrumentally valuable stems from the frequent instrumental success of true beliefs. So the non-instrumental value of normative truths (if there is such a thing) is not completely independent from the instrumental value of truths more generally.

<sup>27</sup>To be sure, true beliefs would *also* lead to successful actions, but in a less efficient way. That is because each true conviction would be a belief similar to the corresponding belief that is true in the actual world *plus* accompanying

which is to say that one is in general instrumentally more successful if one believes falsehoods. (Although these must be particular falsehoods, of course.)

So, BI and minimalism are, as it is, probably incompatible. (Though Lynch has not presented a proof for this.) But this does not undermine minimalism's plausibility, since (i) the phrase "more than instrumentally good" is, unlike its (instrumental) counterpart, particularly underspecified, and (ii) given that unreliable intuitions about far-away possible worlds is the only support in favour of BI, its justification is highly dubious anyway.

### 3 Conclusion

In the beginning, I promised to do two things. Firstly, to explain how what you – correctly, maybe – believe about why true beliefs are valuable may be accommodated within a minimalistic theory of truth. To this end, I discussed the 'directionality of explanation' in the context of truth norms. I suggested that the advantages of going from individual beliefs to generalisations, rather than the other way around, outweigh the disadvantages (the need to explain general facts in terms of particular facts). I argued that, generally, a theory of truth would be the wrong place to look for an explanation of instrumental success. Secondly, I promised to show that not all truth norms commonly thought to hold are supported by the available evidence. In particular, I argued that although true beliefs might be instrumentally good (a result that is compatible with minimalism), they are probably not good in a non-instrumental sense.

Several minor issues further undermine Lynch's position. My worries concerned mainly the *ceteris-paribus* condition involved. It is simply false that believing true things is good tout court. But adding the proviso 'all else equal' is renaming the problem, not solving it. Other worries concerned the term "good", which is largely unspecified on Lynch's account, so that both its meaning and scope remain obscure.

Truth is, after all, non-instrumentally valueless. That is because truth is no genuine property. All plausible reasons for thinking that true beliefs are good stem from the instrumental success of true beliefs in guiding one's action. However, these norms (if true) can be accounted for by a minimalistic, Horwich-style theory of truth. In order to explain instrumental value, we do not need to postulate that truth is a property (and hence something to associate value with). Besides instrumental success, there is no value in the area of truth.<sup>28</sup>

### References

- Coates, A. (2009), Explaining the Value of Truth, *American Philosophical Quarterly* 46, 105–115.
- David, M. (2005), On 'Truth is Good', *Philosophical Books* 46, 292–301.
- Dummett, M. (1959), Truth, *Proceedings of the Aristotelian Society* 59, 141–62.
- Heal, J. (1987), The Disinterested Search for Truth, *Proceedings of the Aristotelian Society* 88, 97–108.
- Earman, J., Roberts, J. and Smith, S. (2002), *Ceteris paribus* lost, *Erkenntnis* 57, 281–301.
- Hajek, A. & Joyce, J.M (2008), 'Confirmation', in S. Psillos & M. Curd, eds., 'The Routledge Companion to Philosophy of Science', Routledge, Abingdon, pp. 115–128.
- Horwich, P. (1998), *Meaning*, Clarendon Press, Oxford.

background assumptions about the peculiarities of Russell worlds. Hence, the fastest – instrumentally most successful – route to a hard-boiled egg is via the false assumption that the egg-cooking theory holds.

<sup>28</sup>This text benefited a lot from the comments that I got when giving a talk about this topic at Professor Schurz's research seminar at Düsseldorf University in November 2013. I would like to thank everyone for their questions and remarks, in particular Alexander Christian, Christian Feldbacher, and Matthias Unterhuber.

- Horwich, P. (2001), 'Norms of Truth and Meaning', in R. Schantz, ed., 'What is Truth?', de Gruyter, Berlin, pp. 133–145.
- Horwich, P. (2005), *Truth*, Clarendon Press, Oxford.
- Horwich, P. (2006), The Value of Truth, *Nous* **40**, 347–360.
- Horwich, P. (2010), *Truth – Meaning – Reality*, Clarendon Press, Oxford.
- Horwich, P. (2013), 'Belief-Truth Norms', in T. Chan, ed., 'The Aim of Belief', Oxford University Press, Oxford, pp. 17–31.
- Lynch, M. (2004), *True to Life. Why Truth Matters*, MIT Press, Cambridge, MA.
- Lynch, M. (2004a), Minimalism and the Value of Truth, *Philosophical Quarterly* **54**, 497–517.
- Lynch, M. (2005), Replies to Critics, *Philosophical Books* **46**, 331–342.
- Lynch, M. (2009), *Truth as One and Many*, Oxford University Press, Oxford.
- Lynch, M. (2012), The Many Faces of Truth: A Reply to Some Critics, *International Journal of Philosophical Studies* **20**, 255–269.
- McGrath, M. (2005), Lynch on The Value of Truth, *Philosophical Books* **46**, 302–310.
- Schaffer, J. (2008), 'The Metaphysics of Causation', in Edward N. Zalta, ed., 'The Stanford Encyclopedia of Philosophy'. Link: <http://plato.stanford.edu/archives/fall2008/entries/causation-metaphysics/>; accessed 30/10/2013.
- Wright, C. (1994), *Truth and Objectivity*, Harvard University Press, Cambridge, MA.
- Wright, C. (2003), *Saving the Differences*, Harvard University Press, Cambridge, MA.