A theory of everything. Book review of Richard Dawid: *String Theory and the Scientific Method*

Philosophers of science generally assume that theorizing is boundless in principle, but in fact constrained mostly by the phenomena. Theories which violate empirical constraints are weeded out in the scientific process. String theory research, which unifies the four fundamental forces in nature and postulates strings as the smallest physical unit, is different: in its current form, string theory makes no empirical predictions. Still, string theory research has enjoyed great popularity in the community of high energy physics theoreticians for decades. In fact, string theory is more popular than any other approach that attempts to unify gravity with the principles of quantum mechanics. Given the lack of empirical constraints, this surely seems odd. One way of viewing the situation is it to hold on to the standard view of the relation between theorizing and empirical constraints. The popularity of string theory is then an unhealthy anomaly, which can be explained sociologically as resulting from the influence of certain individuals and herd behavior (for such views see Penrose 2004; Smolin 2007; Woit 2011). Richard Dawid, in his rich book *String Theory and the Scientific Method*, has set himself a much more demanding task. Instead of castigating string theory research, Dawid asks how we could perhaps extend our idea of the scientific method so as to incorporate string theory research.

Dawid's book consists of three parts. In the first part, Dawid argues that theorizing in the context of string theory is not as unconstrained as one might think—even in the absence of any direct empirical tests. In the second part, Dawid suggests that string theory research is not so special after all: it can be viewed as the end point of a historical trend toward an ever greater distance between our observations and our theories. In the third part, Dawid relates his discussion to the realism debate in the philosophy of science. In what folllows I will try to lead the reader through some of Dawid's main arguments, focusing on the first and third parts of the book.

Throughout the book, Dawid frames the question of how theorizing gets constrained non-empirically in terms of the underdetermination of theories by the evidence. In the first part of the book, Dawid is concerned only with what he calls 'scientific underdetermination', which is better known as *transient* underdetermination (Stanford 2006): the evidence cannot tell two theories apart now, but might do so in the future.[1] When physicists opt for string theory, Dawid suggests, they 'assess' whether there is, or rather how likely it is that there is, any transient underdetermination of string theory by the evidence. Dawid's thought is that physicists are justified in embracing string theory when they can convince themselves that string theory is not underdetermined. Given that string theory has no testable consequences, this way of setting up the problem should strike one as rather awkward. Dawid, however, believes that it is warranted because string theory "conceptually relies on the given empirical data" (50). I'm not entirely sure what he means by that, but I think it is clearly too thin for talk about potential underdetermination in any case. For present purposes I will have to let this pass though. Unless otherwise indicated, we will always refer to transient underdetermination.

Dawid argues that physicists possess three ways of "assessing", as Dawid puts it himself, whether there is any underdetermination in the context of string theory: (i) even after trying very hard, they can find no alternatives to string theory, (ii) string theory has unexpected explanatory consequences, and (iii) the research program which string theory is part of, has been very successful in predicting new phenomena in the past. Dawid refers to these arguments in favour of string theory as the No-Alternatives-Argument (NAA), Unexpected Explanatory Coherence (UEA), and the Meta-Inductive Argument (MIA), respectively. Dawid purports that physicists are actually led by these

[1] For reasons that I think are insubstantial, Dawid decides against using the standard terminology (46-7).

arguments to believe in string theory. He provides little (if any evidence) for that though.[2] More importantly, I don't think that any of his arguments are compelling.

First NAA. The failure of theorists to come up with alternatives to string theory, Dawid points out, need not mean that we haven't tried hard enough. It doesn't even have to mean that we are cognitively limited beings. It could also mean that there simply are no alternatives. In the latter case, physicists would have good reasons for believing in string theory despite the lack of evidence. But how can we even make sense of the idea that there are no theoretical alternatives to string theory? The underdetermination of theories by evidence is a logical thesis: no given set of evidence implies any one theory. In order for us to say that the number of theories is highly restricted (or even that there exists only one theory for the evidence), we need further criteria than fit with the evidence. Aesthetic criteria like simplicity and beauty, Dawid rejects explicitly (54-55). Curiously, unification is not even mentioned by Dawid. Instead he believes that MIA can do the constraining (see below). Then again one might wonder what NAA does in the first place. Indeed NAA seems to amount to no more than an *assumption* that our failure to conceive of alternatives is a good sign for the lack of alternatives, or, limitation of underdetermination. NAA, on its own, therefore simply begs the question.

Next UEA. In analogy to the rationale offered by some for the importance of novel success in theory confirmation (Worrall 1989), Dawid suggests that a theory's explanations of phenomena count more in favour of the theory when the theory wasn't constructed in order to explain those phenomena (52). It is unexpected, Dawid alleges, that string theory not only unifies gravity and quantum principles within a gauge field theory framework, but at the same time also explains the scale hierarchy between the Planck scale and the abundance of black hole entropy. Dawid claims that it is unlikely that there is more than one theory that explains all of those phenomena, rather than just one (in an unexpected way). Thus UEA 'constrains' underdetermination. But why is it unlikely? Although it seems plausible that a theory is likely to fit the facts when it was constructed to fit the facts, the reverse does not follow. Moreover, Dawid should be aware that the use-novelty account he appeals to has been shown to be highly problematic (Musgrave 1974). Suppose scientist S1, due to her ignorance of the theory itself or her ignorance of the extant phenomena, did not expect theory T to explain phenomenon P, whereas S2 did. Then, absurdly, T would receive more confirmation from P, when considered by S1, than when considered by S2. Furthermore, why can we not appreciate string theory simply for its unifying power? For reasons that escape me, Dawid is unwilling to base his account on the latter.

Lastly, MIA. MIA is a 'meta-induction' over the past successful predictions of a research program to the likely future predictive success of the research program. Dawid takes his lead from the immense empirical success of the standard model, which correctly predicted a host of fundamental particles. Physicists were thus very confident that standard model's prediction of the Higgs boson was correct, even though they had to await its discovery for several decades (36-7). Dawid claims that string theory is part of the same research program as the standard model and concludes that therefore string theorists are entitled to trust in the future predictive success of string theory on the basis of the standard model's past predictive success. But what are the identity conditions of a research program? What makes string theory part of the same research program as the standard model? Dawid has little to say about these basic questions. Dawid makes another claim: the fewer the (existing) theoretical alternatives, the better one's chances of choosing a theory in the present which will be predictively successful in the future. The track record of predictive success of a research program, accordingly, would indicate that 'underdetermination is limited' (54-5). If that really were the case, NAA would be supported (see above). But again, I can detect little by way of argument for these claims. I cannot

[2] In personal communication, Dawid reports that several string physicists have publicly endorsed some main lines of his arguments. See also Castelvecchi (2015)

even see why this ought to be so, unless one supposes that a theory is more likely to be true, the fewer theoretical alternatives we can think of. Then we would have good grounds for expecting our theory to perform well in the future too. But of course we need to be given reasons for this supposition. Dawid seeks to provide such reasons only later in the book. I will get to that in a moment.

The inference from the ensemble of NAA, UEA, and MIA to a limit on underdetermination, Dawid says, is an inference-to-the-best explanation (IBE). That is, the best explanation of the failure to find alternatives, the unexpected explanatory coherence of diverse phenomena, and the regular predictive success in the research programme is it that there exist few (if any) alternative theories in the relevant context.[3] Dawid in fact holds that this 'meta-level IBE' underlies the standard IBE from empirical data to the truth of hypotheses: we must make sure that a likely hypothesis is among the ones considered (67). Dawid then provides a Bayesian analysis of his 'meta-level IBE', which he developed together with two co-authors elsewhere (Dawid et al. 2014). The upshot is that, even in the absence of direct evidence for H, the lack of alternatives to H *confirms* H, i.e., $P(T|F_A)>P(T)$, where F_A = scientists' failure to find alternatives to T, under the assumption that the failure of scientists to find alternatives to T is not due to scientists' limited cognitive capacities.[4] Dawid believes that MIA can compel us to accept this assumption in a research field with regular predictive success, because, according to him, this is an indication of a lack of alternative theories.

In the third part of the book, Dawid finally addresses a concern, which he flags early on in the book, and which anyone familiar with more recent debates in the philosophy of science, will have. Stanford (2006) has argued very forcefully that our repeated failure in the past to conceive of better theories than we did suggests that we have little reason to think that we are doing better now. Stanford's argument, also known as the problem of unconceived alternatives (PUA), directly threatens NAA and the very idea that our inability to conceive of alternatives actually is good evidence for the number of alternative theories being limited. Dawid's response to this concern is that his strategies for transient underdetermination actually apply also to repeated transient, and even permanent, underdetermination, because of the very nature of string theory (cf. 139 and 169). There are two properties in particular, which Dawid highlights in this context: minimal length scale (133ff.) and structural uniqueness (141ff.). According to so-called T-dualities, and for reasons having to do with the compactification of dimensions, any distance measured below the string length can be translated back to the string length. Any information below the string length therefore becomes 'redundant'. String theory is thus 'final' because it implies that no other theory will give us any greater knowledge about the world at the microphysical level. String theory is structurally unique in that, in contrast to many other physical theories, it does not possess any free parameters. Theories of gravitation, for example, have a gravitational constant that is determined experimentally. The standard model possesses more than twenty free parameters. The more free parameters, the easier it is to fit a theory with the data. Many philosophers consider too much parameter-freedom in a theory a vice (Popper 1959; Forster and Sober 1994; Worrall 2002). Both the minimal length scale and the structural uniqueness of string theory, Dawid believes, justify 'final theory claims' about string theory: regardless of inductive risk, there is unlikely to be any theory in high energy physics that will succeed string theory. I'm not persuaded. At least one of these two ways of justifying final theory claims about string theory presupposes the truth of string theory itself. And the vast number of ground states of string theory (100^{500}) render the theory extremely flexible with regard to a possible fit with the

---

[3] In the third part of the book, Dawid claims that a 'limitation of underdetermination' is in fact a *better* explanation of predictive success than the standard realist no-miracles-argument for the truth of a theory (169-170).

[4] One should note that in the book Dawid frequently speaks about the viability of a theory, and in his discussion of Bayesian non-empirical confirmation, also about the truth of a theory, whereas Dawid et al. (2014) only talk about empirical adequacy.

phenomena.[5] Indeed, as Dawid admits himself, it is even questionable whether string theory is a proper theory in the first place (175). Even if Dawid's arguments for the finality of string theory were successful, they would be very specific. But Dawid's account is supposed to put limits on underdetermination not only in the assessment of string theory, but also in many other contexts: predictive success and IBE (as seen above), theory choice (in favor of string theory and to the disadvantage of cosmic inflation and quantum loop theory) (89f.), inferences to unobservables in physics and in paleontology (103ff.), the anthropic principle (149), and even data analysis (111). Dawid's account is thus a bit of a theory of everything itself. However, in none of these contexts, it seems, are there any theoretical features comparable with string theory's structural uniqueness or length scale. Dawid's strategy against Stanford's PUA must therefore fail in all of these cases.

Although I'm not convinced by many of Dawid's arguments, I find his book valuable for a number of reasons. Dawid does a great job in laying out, in a comprehensive way, some of the core ideas of perhaps the most challenging theory that science knows to date and offers a thought-provoking, genuinely philosophical, perspective on it. I'm highly sympathetic to his taking on what I described above as the hard explanatory task; sociological explanations fall too short in making sense of the popularity of string theory amongst theoretical high energy physicists. I also think that he is spot on in assessing the at time acrimonious debate between friends and foes of string theory. It is one of the few instances, Dawid notes, where the 'expert principle' is violated, i.e., non-expert scientists feel warranted to judge the soundness of an established research field on the basis of their intuitions about the scientific method, with the result that people are talking past each other. This does call for a debate of the scientific method and Dawid deserves respect for accepting the challenge. I strongly recommend Dawid's book for anybody interested in a highly accessible philosophical treatment of string theory and for anybody open to some fresh thoughts on the scientific method.

References

Castelvecchi, D. 2015. Feuding Physicists Turn to Philosophy for Help. *Nature* 528 (7583):446-447.

Dawid, R., S. Hartmann, and J. Sprenger. 2014. The No Alternatives Argument. *The British Journal for the Philosophy of Science*:axt045.

Forster, M., and E. Sober. 1994. How to Tell When Simpler, More Unified, or Less Ad Hoc Theories Will Provide More Accurate Predictions. *The British Journal for the Philosophy of Science* 45 (1):1-35.

Musgrave, A. 1974. Logical Versus Historical Theories of Confirmation. *British Journal for the Philosophy of Science*:1-23.

Penrose, R. 2004. *The Road to Reality: A Complete Guide to the Laws of the Universe*. London: Random House.

Popper, K.R. 1959. *The Logic of Scientific Discovery*. London: Routledge.

Smolin, L. 2007. *The Trouble with Physics: The Rise of String Theory, the Fall of a Science, and What Comes Next*: Houghton Mifflin Harcourt.

Stanford, P.K. 2006. *Exceeding Our Grasp: Science, History, and the Problem of Unconceived Alternatives*. Oxford: Oxford University Press.

Woit, P. 2011. *Not Even Wrong: The Failure of String Theory and the Continuing Challenge to Unify the Laws of Physics*: Random House.

Worrall, J. 1989. Fresnel, Poisson and the 'White Spot': The Role of Successful Prediction in Theory-Acceptance. In *The Uses of Experiment*, edited by D. Gooding, T. Pinch and S. Schaffer. Cambridge: Cambridge University Press.

---

[5] The vast number of ground states of string theory is mentioned by Dawid, although he does not seem to be overly concerned by it. On the contrary, he introduces the term 'highly predictive' for theories—such as string theory—where the number of possible free parameter values exceeds the number of ground states. This strikes me as ad hoc.

———. 2002. New Evidence for Old. In *In the Scope of Logic, Methodology and Philosophy of Science*, edited by P. Gardenfors. Dordrecht: Kluwer