Book review published in *Dialectica* **67** (2013), pp. 367-372

John Wright: Explaining science's success: Understanding how scientific knowledge works

Durham: Acumen Publishing, 2013, 199 pp., ISBN 978-1-84465-532-8

Demands for explanation are often ambiguous, especially if the *explanandum* is such a multifaceted phenomenon as the success of science. It is therefore appropriate that John Wright opens his book by specifying what kinds of success it attempts to explain. In his view, the surprising thing about science is its *novel predictive success*: Scientific theories have led to correct predictions about phenomena which were unobserved or even observationally inaccessible at the time the theories were first advanced. Furthermore, some scientific theories were postulated on more or less *a priori* grounds, but turned out to make correct empirical predictions. It is this kind of success with which Wright's book is concerned.

But even if it is now clear what the explanandum is, opinion may still be divided as to what exactly it means to explain it. This becomes apparent in chapter 2, where Wright critically discusses various explanations of science's success which have been proposed in the literature. One way to interpret the demand for explanation is to assume that what is sought is an answer to the following question:

(1) Why do our scientific theories tend to be successful?

This question plays an important role in the debate on scientific realism. Broadly speaking, scientific realists reply to it by saying that our theories are (approximately) true, while antirealists have their different ways of denying this. Wright considers realist and antirealist approaches equally unsatisfactory, because they do not address a question which he thinks needs to be addressed if the success of science is to be explained, namely:

(2) How did we manage to hit upon successful scientific theories?

A satisfactory answer to that question should inform us about what Wright calls "property M", that is, a feature by means of which we are able to identify theories that tend to be successful. Spelling out what property M consists in will be one of the main tasks of later chapters, but at this point, all that needs to be said about M is that it must be *accessible* in the following sense: It must be easier for us to tell whether some theory has M than it is to tell whether it will tend to be successful. This immediately shows that the realist's notion of approximate truth does not qualify as a candidate for M.

The task of explaining science's success can then be seen as the task of explicating M and providing answers to the following two questions:

(3) Why do we prefer M-theories?
(4) Why do M-theories tend to be successful?

Wright now claims that previous attempts to explain science's success manage to answer only either (3) or (4), but not both. And if some combination of strategies is employed, he still takes the resulting explanation to be unsatisfactory, because it does not explain why the type of theory we prefer *also* happens to be the one that tends to be successful. Many adherents of the views criticized in this chapter will protest that

Wright here only targets overly simplistic versions of their approaches, but this is perhaps inevitable if one tries to address such a variety of views in just one chapter. A more serious objection is that, as we will see below, the basic criticism Wright advances against extant views also applies to his own approach.

The central chapters 3 to 5 are dedicated to developing this approach. The central idea is that any reasonable explication of property M must somehow capture what we intuitively call a theory's *simplicity* or *lack of ad hocness*. Wright proposes a definition of what he calls "the independence of theory from data" to achieve this aim. Very roughly, the independence of a theory T is defined as the ratio between the amount of (previously obtained) data which T can explain and the number of explanatory components which T needs to postulate in order to do this job. As an illustrative example, one may think of the epicycles postulated by the ptolemaic system in order to account for the astronomical data. Such postulates increase the number of a theory's data-dependent explanatory components and therefore reduce its independence from the data, thus rendering the theory increasingly ad hoc.

Now Wright does not equate this notion of independence with the sought-after property M, but introduces some other success-conducive properties, which are, however, closely related to the independence property just discussed. What results is an interrelated cluster of properties, which, according to Wright, constitutes property M. In chapters 6 to 8, he applies these concepts to three historical cases (Newtonian mechanics, special relativity, and Mendelian genetics), attempting to show that his approach can indeed explain the impressive empirical success of these theories.

Much of what Wright has to say on these issues is formulated with great clarity and supported by well-developed arguments, and the historical case studies provide helpful illustration. Nevertheless, I also see some rather serious flaws in his approach, which I will address in the remainder of this review.

One problem concerns the counting of explanatory components, which is presupposed in Wright's definition of independence. Clearly, if we place no restriction on the language in which a theory is formulated, even the most contrived theory, consisting of a multitude of ad hoc postulates, could be made to look as if it contained just one explanatory component, expressed by a suitably defined disjunctive predicate. Wright (pp. 72-79; 85-86) therefore insists that theories must be couched in "basic natural predicates", and he admits that this threatens the epistemic accessibility of his proposed property M, because it may not be obvious which predicates are to count as "natural". His reply to this worry appeals to David Lewis's (1983) idea that it is the very nature of the reference relation to prefer natural predicates to non-natural ones. It is, however, unclear how this is supposed to help with the epistemic problem, as can be seen from an example which Wright himself introduces (p. 78): Prima facie, "carbon" does not look like a natural predicate, as it applies to such observationally different objects as lumps of coal and diamonds. But of course, it *is* natural to group these things together, by virtue of the chemical elements they consist of. This shows that assessments of naturalness need to be theoretically informed; what we count as basic natural predicates partly depends on the theories we have accepted (modern chemistry in this case). But if such a theory-dependent notion plays an essential role in Wright's proposal to explicate property M, that property's role in theory choice becomes obscure. It seems that we already need to decide whether or not to accept a theory before we can evaluate to what degree it possesses property M.

No one denies that the lack of ad hocness, which Wright's property M is supposed to capture, plays an important role in theory choice, but it seems to me that his approach predominantly addresses cases of theory choice of a rather trivial sort. For example, he extensively discusses (pp. 100-105; 117-120) why we prefer laws according to which some physical quantities are *exactly* conserved, as opposed to laws postulating small changes in those quantities, although both types of laws would be compatible with our previous experience. Or he attempts to justify (pp. 105-106) the appearance of the exponent 2, rather than, say, 2.01, in Coulomb's law, although both options would have been compatible with the data available at Coulomb's time. But these are hardly cases of underdetermination which we would consider as scientifically interesting. I grant that Wright's approach has the virtue of explaining *why* we consider these cases trivial and why we do not lie awake at night pondering why Coulomb put "2" rather than "2.01" in his formula, but then again, this phenomenon may not seem surprising enough to call for an explanation.

When it comes to non-trivial cases of underdetermination, Wright does not seem to have much to say. For example, the chapter on special relativity does not mention the most serious early alternative to Einstein's theory, namely Lorentz's ether theory. This is all the more surprising, because the FitzGerald-Lorentz contraction hypothesis incorporated in Lorentz's theory is generally viewed as a paradigm case of an ad hoc hypothesis, which would make it an ideal test case to see whether Wright's account of ad hocness (or lack thereof) can explain why Einstein's theory was preferred to Lorentz's. Christopher Hunt (2012) has recently reviewed different attempts at characterizing ad hocness in the literature, and concluded that the concept is useless for understanding relevant cases of theory choice (such as the Einstein-Lorentz case). I suspect that this negative judgment also applies to Wright's account, because I do not see how Wright's account is supposed to go beyond the earlier proposals discussed by Hunt. Here, Wright's study could have profited from a critical engagement with his predecessors, such as Popper, Grünbaum, or Leplin.

Hunt's (2012, p. 6) paper also raises an interesting question regarding another historical example Wright (pp. 136-138) discusses: the prediction (made by Adams and LeVerrier on the basis of Newton's theory of gravity) of the existence of the planet Neptune. Let us assume that, as Wright claims, his account can explain the success of this prediction. Is it then not somewhat surprising that LeVerrier a few years later, using exactly the same type of reasoning, arrived at a spectacularly *un*successful prediction, namely the existence of the planet "Vulcan", supposedly responsible for the observed deviations in Mercury's orbit (which really are effects of general relativity)?

This brings me to my final (and probably most fundamental) point of criticism. Wright has a response to my rhetorical question in the previous paragraph, namely that he only aspires to give a *probabilistic* explanation of science's success, and it is not part of this notion of explanation that the explanandum is to be expected, given the explanation (p. 55). In other words, having a probabilistic explanation of LeVerrier's success concerning Neptune is compatible with acknowledging his failure concerning Vulcan. My fundamental disagreement with Wright is that I do not consider this notion of explanation sufficient for giving a satisfactory answer to *both* questions (3) and (4) mentioned above.

Notice first that, by Wright's own admission (pp. 83-84), probabilities by themselves do not explain anything. What plays the explanatory role is a *propensity* that underlies the probabilities. I am not here concerned with general metaphysical objections to propensities.[1] What worries me is that the propensities postulated by Wright's explanation are of a rather peculiar kind. To see this, it is instructive to look at how Wright introduces them. He starts with a case in which I take the appeal to propensities to be unproblematic, namely enumerative induction about concrete objects (pp. 43-56). Having observed a large number of crows and having found that they are all black, it is reasonable to infer that crows have a propensity to be black, which in turn licenses the inference that the next crows to be observed will be black as well. But even in this case, not every appeal to propensities is equally well justified. As Wright himself observes (p. 53), even if we had never seen any black things that were not crows, the observation of many black crows would not license the inference that black things have a propensity to be crows. The reason for this asymmetry is that an object's being a crow plausibly has a causal influence on its being black, but not vice versa (p. 190n8). So whether or not claims about propensities are justified depends on the causal structure of the world. Now let us turn to the kind of propensity used in Wright's explanation of science's success (pp. 66-70). This is a propensity *of some data to exemplify a certain theory*. I find the analogy between *data* and concrete objects like crows difficult to grasp. In particular, nothing in my causal background knowledge tells me whether I should think of data as analogous to crows (which have a propensity to be black), rather than analogous to black things (which do *not* have a propensity to be crows).

But even if we accept the existence of the propensity which Wright attributes to data, and further accept that this propensity is, as he claims, conferred on the data by the corresponding theory's having property M, I doubt that he has given us a satisfactory explanation of science's success. For the statement just given is obviously not an *answer* to question (4), but a restatement of the explanandum contained in (4); to say that the propensity of the data to exemplify an M-theory *explains* the M-theory's tendency to successfully predict the data is like saying that Djokovic's defeat *explains* Murray's victory in the 2013 Wimbledon final.

Wright might again reply that I have mischaracterized what a probabilistic explanation is supposed to do. According to Salmon's (1971) statistical relevance (SR) model of explanation, an explanation just needs to state the facts which are statistically relevant to the explanandum, and Wright's account does this by stating that a theory's possessing property M is statistically relevant to a theory's success. In his concluding remarks (pp. 180-184), Wright admits that we may very well ask for more and thereby enter the debate on scientific realism, but he seems to think that his SR explanation of science's success will be untouched by the outcome of that debate. However, this may not be so: If scientific realism is true, then, plausibly, *both* a theory's possessing property M *and* its empirical success will be consequences of the theory's approximate truth. Put in statistical terms, approximate truth will *screen off* property M from success, just as atmospheric pressure screens off the barometer reading from the occurrence of the storm in the famous example discussed in Salmon (1971, pp. 53-55). In other words, approximate truth renders property M *statistically irrelevant* to a theory's success. But if that is the case, then Wright's explanation of science's success fails, even in the weak sense of the SR model.

---

[1] Let me note in passing that Wright's commitment to propensities may well undermine his claim (p. 180) that his explanation does not "appeal to the reality of any theoretical entities".

These critical remarks about Wright's reply to question (4) should not distract from the fact that his discussion provides a well-argued reply to question (3) and thereby also offers valuable insight regarding question (2). To everyone who is interested in these questions, the book is warmly recommended.

References

Hunt, J. C. (2012). "On Ad Hoc Hypotheses". *Philosophy of Science 79*, 1-14.

Lewis, D. (1983). "New Work for a Theory of Universals". *Australasian Journal of Philosophy 61*, 343-377.

Salmon, W. (ed.) (1971). *Statistical Explanation and Statistical Relevance*. Pittsburgh: University of Pittsburgh Press.

Matthias Egg, University of Lausanne

matthias.egg@unil.ch