DANIEL ANDLER

THE ADVANTAGES OF THEFT OVER HONEST TOIL COMMENTS ON DAVID ATKINSON

David Atkinson asks whether nonempirical constructions can lead to genuine knowledge in science, and answers in the negative. Thought experiments, in his view, are to be commended only insofar as they eventually lead to real experiments. The claim does not rely on a general study, conceptual or historical, of thought experiments as such: the range of the paper is at once narrower and broader. Atkinson views thought experiments as commonly understood as just one kind of episode in the development of physics in which real experimentation is bypassed, and he believes that such episodes are justified only inasmuch as they are transitory stages on the way to genuine empirical inquiry.

Atkinson wants to kill with one arrow what is usually regarded as two different birds: the notion that thought experiments proper can be persuasive in themselves, and the thesis that theories which cannot be brought to the tribunal of experience can nevertheless belong to science. He thus implicitly opposes *three* views which are commonly, albeit not universally, held: (i) some thought experiments are conclusive; (ii) some theories belong to science despite not being evidently and concretely amenable to empirical corroboration; (iii) the two issues are largely independent.

The standpoint from which Atkinson operates is a rather strict form of empiricism, one which relies on a fairly sharp distinction between the conceptual and empirical dimensions of inquiry. My outlook is rather different: the conceptual and empirical seem to me to be intertwined, both conceptually, as suggested by Quine's critique of logical positivism, and empirically, as revealed by the evidence provided by science itself in its daily and historical reality.

Rather than take the high road, I propose to focus first on the critique to which Atkinson subjects Galileo's thought experiment, and the lessons he draws from his analysis. The other two case studies I will not examine individually, due first to insufficient expertise, second to the dialectics of the situation. I have no quarrel with the conclusions which Atkinson draws

M.C. Galavotti (ed.), Observation and Experiment in the Natural and Social Sciences, 1–10.

from the history of EPR: I certainly don't doubt that such happy *dénouements* are productive for science and deeply satisfying from both a historical and an aesthetic perspective.

As for string theory, it seems to me to raise a separate issue. High-level theories are notoriously hard to confront with experience. Evolutionary theory is perhaps the most familiar example; Popper categorized it as a metaphysical research program for precisely that reason. Most philosophers of biology today would be reluctant to set it so sharply apart from more typical scientific theories, such as Newtonian mechanics or molecular genetics. But the important question which Atkinson raises, as I understand it, is whether mathematical physics, which does not deal with one-time only sequences of events and in that sense is, contrary to evolutionary theory, fully theoretical, is not committed to stricter standards of verifiability. It certainly seems that verification procedures which are in principle possible but call for impossible feats of engineering, though not unrelated to the situation of forever lost historical or paleontological archives, create a novel predicament: the impossibility goes deeper, one is inclined to think. However I do not see exactly how one could rule out theoretical developments which would bring the higher-level theory closer to a, surely quite indirect, confrontation with empirical data. I will not attempt to go further into the matter.

In the second part of this comment, I shall attempt to gain, from the consideration of the set of case studies which Atkinson picks, as it were, as his data base, a perspective on his methodological doctrine.

1. Galileo's "conclusive argument"

Atkinson faults Galileo for claiming that logic alone shows the Aristotelian dogma (in either strong or weak form) to be inconsistent (the "destructive aim" of Galileo's argument), and the Galilean doctrine about free fall *ipso facto* correct (the "constructive aim").

Atkinson's argument takes the following form:

(i) Galileo offers the following argument P:

> Let B be the common ground accepted by both thinkers, including the hypotheses concerning the two bodies whose rate of fall is being compared, and A the weak Aristotelian dogma. Then there is an assertion Z such that from B and A jointly follow both Z and not-Z. Therefore, granted the consistency of B, A must be false. Hence not-A.

(ii) There are conditions under which both B and A+, the strong Aristotelian dogma, are true.

(iii) Hence it cannot be the case that A (which is logically weaker than A+) is "internally inconsistent", as Galileo claims: his "destructive aim" is defeated.

(iv) The "new Galilean dogma" G –all bodies fall at the same rate– follows from not-A, according to Galileo. But the inference is unsound: in fact, there are conditions C' under which A is indeed false, but so is G. Thus his "constructive aim" is equally defeated.

(v) The crucial part of P is (S1), the assumption that (given B), natural speeds are mediative. As a matter of fact, there are conditions D under which they are, and conditions D' under which they aren't. Hence, it cannot be the case that S1 is a logical (conceptual) consequence of B, as Galileo maintains.

The complete analysis of Atkinson's argument would require a lot of work, especially since it is divided into a straightforward, non-technical part in the body of the paper, and an "interpretative", technical part in the Appendix. I grant the physics, of course, and find Atkinson's discussion of Gendler's reconstruction of Galileo's argument against Aristotle highly illuminating in its own right.

I do on the other hand have a pair of objections.

There is first a logical point. Atkinson insists on the logical structure of the argument. But in *strictly* logical terms, the new Galilean dogma G *is* (by the double-negation rule) logically equivalent to not-A. So if, as Atkinson argues, it is not the case that not-A is proved, Galileo cannot be faulted for making the *further* logical mistake of positing G on the basis of the rejection of A.

Readers familiar with Gendler 1998 may think that Atkinson has the following issue in mind: rephrasing A as "Rates of fall vary (nontrivially) with weight", not-A comes to "Rates of fall remain constant when weight varies" *but* Galileo's dogma G *further* asserts that rates of fall are insensitive to everything else, such as position with respect to the center of the Earth. *Pace* Gendler, Norton and others, I think this is a red herring: the disagreement between Aristotle and Galileo regards rate of fall as a function of weight, *all else remaining constant.* No doubt there are circumstances which we can imagine under which *color* of the bodies matters, circumstances which neither Aristotle nor Galileo could have imagined. But this presumed fact is neither here nor there: they would have been happy to take this issue on had they come to suspect it had potential interest; as it is, they were concerned with dependency with respect to *weight* not *color.*

What Atkinson actually proves is something different. He shows that (a) there is an *interpretation* J of Galileo's argument, and circumstances C such that A, under J, is actually true; (b) there is an *interpretation* J', and circumstances C' such that although A+, under J', is indeed false, nevertheless G, under J', is also false. (As I understand the arguments, J involves terminal rates of fall in non-vacuous media, while J' involves accelerations *in vacuo*.) This is logically possible because A (hence not-G) does not entail A+. But surely Galileo cannot be faulted for (implicitly) choosing one fixed interpretation in the short excerpt of the dialogue quoted; in fact, had he shifted interpretations half-way in the argument, he could have rightly been faulted for disingenuousness.

Still, one might think, the case against Galileo remains basically sound: Galileo has failed (i) to prove by logical means alone that terminal velocities in non-vacuous media are constant (hence that the weak dogma, thus construed, is false) *and* (ii) to prove by logical means alone that accelerations *in vacuo* are constant (hence in establishing his own "dogma").

However, as "grammar" suggests, one cannot fail unless one has tried. My second and main objection to Atkinson's overall argument is this: Galileo does *not* attempt to provide a *logical proof* of either the falsity of the Aristotelian dogma(s) or the truth of his own theory. The reason is simply that no concept of logical truth in a sense commensurate to our own, and to the one Atkinson relies on, is available to Galileo and his contemporaries. What Galileo is after is an *argument*, which he hopes will be conclusive. Nor is this a mere verbal issue: as his extended discussion in the *Discorsi* shows, he is quite aware of the fact that his conclusions result not from the sole application of logical rules, but implicate a variety of considerations, ranging from interpretation of the terms involved, to acceptability of idealizations, common sense, convergence of arguments, rejection of putative objections, and so forth[1]. There is even an unarticulated consideration which plays a role in the central point's persuasiveness, for Galileo as well as many of his readers, *viz.* the very straightforwardness of the sub-argument that natural speeds are mediative. Finally, Atkinson's hard work is clearly of a nonlogical nature: even if, contrary to my reading, Galileo were committing a logical fallacy, Atkinson would not be meeting him on logical grounds, but on the customary grounds where physicists appraise arguments from *physics*.

Once this is accepted, it remains to get clear on the following two related issues. First, is Atkinson right in faulting Galileo's argument form? Second, what, if anything, is wrong about the central sub-argument?

The first issue raises a host of problems. I will briefly focus on just two, which are related and both concern the vexing phenomenon of defeasability. Arguments, however tight and apparently conclusive, regularly turn out to be defeasible (though thankfully not always defeated): airtighteness is not of this world, it appears, whether in science or in lay reasoning. Atkinson, with laudable fairness to Galileo, does not blame him for not having thought of some of the particular ways in which his arguments could go wrong; but he does blame him for not realizing that this could well happen. What he faults in Galileo is what he sees as his conceptual dogmatism. But is Galileo really dogmatic? Does he really think of his arguments as indefeasible? I have my doubts, nor do I think that we are offered evidence to that effect. Further, does Galileo base his alleged dogmatism on a misplaced trust in conceptual reason, as opposed from appraisal of empirical evidence? Again, I do not presume to decide; but I do have a suspicion that Galileo might have less confidence than Atkinson in the virtues of empirical results, and (as I stated earlier), if this is the case I side with Galileo against Atkinson.

The second problem regards rigor: shouldn't we simply admit that Galileo, with all due reverence, is insufficiently rigorous? After all, he must be, since his argument contains loopholes. But *all* arguments do, and standards of rigor, no matter how stringent, cannot provide complete protection. They constrain only the conditions under which subjective evaluations of the soundness of the ideally completed argument are communicated among scientists. The tighter the standards, the more confident one can be that the *actual* (perforce enthymematic) argument under scrutiny is *essentially* sound. However, rigor has a cost which needs to be justified by the subjectively evaluated risk that a trap may have been laid out by nature. So the standards of rigor change over time, tending towards greater strictness as we find out, by falling into them over and over again, that traps abound. That this tightening of standards occurs even in the history of mathematics tends to show that the conceptual/empirical distinction is irrelevant.

About the sub-argument, Atkinson is correct in pointing out that Galileo's way of settling the question of what would be the rate of fall of the composite object, were the rates of fall of its components different, can strike the modern reader as flippant. However, suppose for a moment that, like Galileo, you believe that in fact natural speeds are invariant in our world (neglecting air resistance). A world in which they are not is therefore, as you see it, imaginary. Reasoning on imaginary worlds is, of course, what thought experiments are all about, and it is a notoriously dangerous exercise, because it involves proceeding with counterfactual premises as if they were true in the real world, or, more precisely, it means operating in a possible world closest to our own in which the premises are true. Galileo had no choice but reason as he did; he used his common sense, i.e. his wordly knowledge (his "naive physics"), to speculate about an otherwordly event. But what else could he go by? Common sense is the guide to follow until reasons to doubt its conclusions come to light:

defeasibility again. As Atkinson implies, other worlds need to be kept under the control of a pre-existing theory. But Galileo has such a theory, albeit in a state of less than perfect scientific crystallization. Maybe we have reason to put in question, with hindsight, the choice of this particular counterfactual situation (for example, because we suspect that it is under-described). But exactly the same fate awaits *any* experiment, real as well as imaginary.

2. Conceptual truths, defeasibility, and corroboration

Seen as a whole, Atkinson's paper appears to rest on three broad assumptions: (i) There is a sharp separation between conclusions based on experience and conclusions based on conceptual analysis; (ii) The essential weakness of the latter is that they are vulnerable to new facts; (iii) Conceptual analysis too easily evades the tribunal of experience. I concede that Atkinson does not take an explicit stand on those general issues, and might well want to deny that his case logically depends on the assumptions as I just stated them. The observations that follow are based on my possibly overly schematic reconstruction of his background doctrine. I will take up the three posits in turn.

(i) Matters of fact *vs* matters of meaning. What Atkinson means by "(pure) logic", I take it, is not really what goes under that label in contemporary philosophy, but rather conceptual analysis. He argues that the resources of conceptual analysis, buttressed, when a non-Platonic realm is being investigated (free fall *vs* numbers for example), by common sense, are too weak to bring about results enjoying the robustness of those procured by empirical means.

The difficulty here is familiar: the conceptual/empirical distinction is unclear. There are two well-trodden ways to see this. The (broadly) conceptual way is to follow Quine's rejection of the analytic/synthetic distinction: whether or not we should take a step back and allow for more of a principled distinction where Quine sees none at all, it remains that we have for now no stable notion of a conceptual truth. The (broadly) empirical way is to consider the actual practice of scientific inquiry, where conceptual and empirical considerations are inextricably intertwined. This is not to deny that there is *some* distinction, which can be usefully drawn, on the fly, in a rough and ready way. In fact, I have intentionally worded my objection so that it involves the very distinction whose status I question. What I challenge is the idea that purely empirical facts, plus logic (in the strict sense), in the absence of any conceptual ingredient, yield robust non trivial results. If one considers the process which includes not only the actual performance of a real experiment, but the work which goes into setting it up and drawing the moral which the scientist draws, one immediately realizes that some thought-experimental procedures are brought in at nearly every step. Although they do not amount to thought experiments standing on their own, they do involve the consideration of counterfactual mini-scenarios, the outcome of which is usually regarded as too obvious to call for a separate checking procedure. Whether or not this shows that the blurring of the analytic/synthetic distinction spills over and messes up the borderline between thought and real experiments seems to me a serious possibility, which I will not explore further. For the purpose at hand, the consequence seems to be that thought experiments, and, more broadly, concept-intensive (or largely armchair-based) inquiries in science are not distinctively more fragile than the more typical empirical inquiries for logical reasons connected with their high conceptual content.

(ii) Defeasibility. Are thought experiments, and other concept-intensive inquiries, vulnerable to defeat by unexpected factors in a sense or in a way in which real experiments are not? Atkinson's argument certainly points in that direction: what's wrong with Galileo's thought experiment, in Atkinson's eyes, is that its conclusions, both destructive and constructive, are defeated in circumstances which Galileo could not imagine. By implication, real experiments are not exposed to the same danger.

As I've stated above, I believe this is wrong: conclusions reached by way of real experiments are just as defeasible. Let's see why. Whenever we draw inferences based on some fact about the real world, we draw on default assumptions; in other words, we operate not in the real world W, but in the world W' which we can imagine to be like the real world in every relevant respect. We perform on W' a thought experiment, using our beliefs about the real world. This is not different from the case of an actual thought experiment, except for the fact that the imaginary world W" there is explicitly posited from the beginning and presumed to be distinct from W: it obeys some condition A which is known, or presumed, to be false in W. This in turn explains why we resort to thought experiments at all: W does not satisfy the preconditions of the experiment.

So when we carry out a thought experiment, we operate on a world which we know (or believe) to be different from our own, yet not to such a degree that we cannot rely on our real-world knowledge to navigate: Superman operates pretty much exactly like Batman, except for the flying part. When we do a real experiment, we navigate in the real world much as the absent-minded and short-sighted Mr. Magoo makes his way across crevices which are temporarily filled by backs of hippos or heads of giraffes; the difference is of degree, not of nature: what makes real experiments work (and the inferences we draw from them sound), when all goes well, are the high probabilities of the default assumptions they unknowingly rely on.

This is not to say, of course, that in respects other than in-principle defeasibility, there are no serious conceptual differences between (typical) real experiments and (typical) thought experiments. It matters a lot that the world W" of a thought experiment is known (or presumed) to differ from the real world W in some respect relevant to the question under investigation. First, as we have just reminded ourselves, it accounts for the fact that we are motivated, and often have no choice but, to perform the thought experiment rather than a real one. Second, the value of the thought experiment resides in the contrast it allows us to discern and make explicit between our world and some neighboring unreal worlds. One family of cases involves imaginary worlds which are *crucially* different from our own: they satisfy a condition A which our purpose is precisely to *prove* that it is false in W: these sorts of thought experiments are the quasi-empirical analogue of formal proofs by *reductio*. Another family involves worlds which are *inessentially* different from our own: this is the case of idealization. The ethereal beauty of Galileo's thought experiment is that it combines both cases, while remaining utterly simple: four worlds are involved, the real one, the one which is like ours only frictionless, the one which is like ours only Aristotelian, and the one in which the experiment is conducted, which is like ours except for being both Aristotelian and frictionless.

Finally, Atkinson's intuition that real experiments are, as a general rule, more reliable than thought experiments, is partly vindicated by the fact that, as science progresses, the evidential network in which a real experiment is performed becomes tighter, and leaves less room for default assumptions being falsified in the real world. By contrast, in a thought experiment involving an imaginary world, the network is by definition pulled apart to make room for the assumption A which is false in the real world, issuing in a less than perfectly controllable loosening of the connections surrounding A.

(iii) Corroboration. Galileo's thought experiment (as opposed to the real experiments he performed with inclined planes and so forth), and string theory (as opposed to conventional theories in fundamental physics), cannot be confronted with experience, and this, Atkinson tells us, is a fault. But what is the confrontation supposed to achieve? Atkinson writes as if the scientists' goal were to identify observable consequences of a theory under assessment such that, if the consequences are actually observed, the theory is thereby vindicated. What's wrong with thought experiments, he thinks, is that there are no facts at all to be had in an imaginary world;

while string theory, according to him, does not proprietarily entail facts which it would in the real world be feasible to check.

But as we all know no amount of corroborated consequences can establish a theory: data underdetermine theories. In fact, the more ambitious the theory, the more it goes beyond its observable consequences. So the difference between acceptable and unacceptable candidates to scientific theoryhood cannot reside in corroborability, in this all-or-nothing construal. It is a matter of domain-specific wisdom. Atkinson is not alone among physicists, it seems, to wonder whether the added intelligibility which string theory may provide balances the increase in indirectness which affects its contact with observable facts. But this, surely, is not a matter of logic nor even methodology in a broad sense.

While we can grant Atkinson that thought experiments are less than fully conclusive; that some theories seem so distant from empirical corroboration as to bring doubts regarding their place in science; and lastly that there is a connection between the two issues, *viz.* they both engage more armchair than lab or field work, we must resist, I suggest, his overall picture, with speculation on one side, together with fragile and temporary results, and possibly a hint of cheapness, and on the other side experiment and confrontation with the real world, hand in hand with robustness, stability, scientific honor and earnestness. Atkinson's distinctions, or so I have argued, are not aligned in the way which would warrant this picture. Science is more of a blend in which the conceptual and empirical dimensions are intertwined, and owe their respective identities more to their interrelations and to the historical and local context than to any durable, intrinsic properties.

Université de Paris-Sorbonne (Paris IV)
Paris, France

NOTES

[1] I am grateful to Marta Spranzi for providing me with expert explication of Galileo's text.

REFERENCES

Sorensen, R. A. 1992. *Thought Experiments*. New York: Oxford University Press.