



Against Robustness? Strategies to Support the Reliability of Scientific Results

Léna Soler

To cite this article: Léna Soler (2014) Against Robustness? Strategies to Support the Reliability of Scientific Results, *International Studies in the Philosophy of Science*, 28:2, 203-215, DOI: [10.1080/02698595.2014.932530](https://doi.org/10.1080/02698595.2014.932530)

To link to this article: <http://dx.doi.org/10.1080/02698595.2014.932530>



Published online: 12 Sep 2014.



Submit your article to this journal [↗](#)



Article views: 72



View related articles [↗](#)



View Crossmark data [↗](#)

REVIEW ARTICLE

Against Robustness? Strategies to Support the Reliability of Scientific Results

Léna Soler

Seeing Things: The Philosophy of Reliable Observation

ROBERT HUDSON

New York, Oxford University Press, 2014

xxiv + 274 pp., ISBN 9780199303281, US\$65.00, £41.99 (hardcover)

1. Introduction

In the last few decades, a number of philosophical works have addressed the issue of robustness in science. In the corresponding contexts, ‘robustness’ can be employed in related but different, more or less precisely defined senses. Several typologies, and attempts to distinguish robustness from similar but taken-as-different reasoning, have been proposed (see, in particular, Woodward 2006; Calcott 2010; Nederbragt 2012; Soler 2012a; Stegenga 2012; Wimsatt 2012). As a starting point, intended to be subject to qualifications but taken to capture the common core of most significant uses, I propose the following definition, which I call the ‘robustness scheme’: X is robust = X is invariant under a multiplicity of independent processes Ps . The robustness scheme leaves open the nature of the X —which proves to be diverse in the literature, e.g. a perceptual object, an experimental result, a theoretical result or a ‘theorem’, a natural reality such as some living beings at a given evolutionary stage—and the kinds of processes under which X can be said to be invariant—e.g. multiple sensorial modalities, experiments, inferences derived from various theories or models, biological evolutionary processes, etc.

Many philosophers of science have presented robustness as a prototypical way through which scientists can secure their results. Writers have claimed that robustness

Léna Soler is at the Laboratoire d’Histoire des Sciences et de Philosophie—Archives Henri Poincaré, Université de Lorraine. Correspondence to: Laboratoire d’Histoire des Sciences et de Philosophie—Archives Henri Poincaré, Université de Lorraine, 91 avenue de la Libération, BP 454, F-54001 Nancy, France. E-mail: l_soler@club-internet.fr

is able to achieve multiple highly valued tasks (see e.g. Stegenga 2012 for a repertory): to demarcate experimental artefacts from real phenomena; to block the experimenters' regress; to help to recognize the best hypotheses; to provide a strong argument for the truth of one particular scientific hypothesis or more generally for scientific realism; and so on. Consistent with philosophers' claims, incredibly numerous examples of scientists' accounts, in various contexts and periods, look, at first glance at least, like evidence that robustness is frequently used as the basis of reliability claims about scientific results and, when not instantiated, works as a regulative ideal—or as a 'methodological attractor' (Trizio 2012)—which can be normatively prescribed, for example by referees, as the next strategy to follow.

Robert Hudson's recent book, *Seeing Things*, offers an opportunity to revisit the issue of robustness. The book looks like a determined crusade against robustness, and could have been entitled, adapting a famous title of Paul K. Feyerabend, *Against Robustness*. More precisely, most of the volume attempts to establish what I will call Hudson's 'discovery claim': namely that robustness reasoning, as the author defines it, does not possess the high epistemic value philosophers and scientists usually attribute to it, and that other strategies—presented as forms of 'reliable process reasoning' called 'targeted testing' and 'calibration'—do the job wrongly assigned to robustness. Consequently, such alternative strategies should replace robustness in the prize list of philosophers.

Below I reconstruct, and critically assess, Hudson's crusade against robustness.

2. Contents and Support of Hudson's Discovery Claim

The main object of Hudson's book is the particular case in which the *Ps* are 'observational procedures' and *X* is 'an observed result' (say *R*)—a case Hudson explicitly identifies with what James Woodward (2006) calls 'measurement robustness' and Brett Calcott (2010) 'robust detection', and which corresponds to one of the three sub-cases of 'inferential robustness' as defined by William C. Wimsatt (2012). More precisely, Hudson's 'observational procedures' paradigmatically correspond to *experimental* procedures. They are also frequently illustrated through—what I would call—'purely' observational processes such as unenhanced identification of macroscopic objects or events.

Hudson's interest is, primarily if not exclusively, in the issue of 'truthful results' (26) and 'the epistemic appraisal of experimental set-ups' (27)—i.e. whether the exhibition of a robustness scheme is a sound basis to believe in the truth of *R* and in the reliability of the *Ps*. 'Weaker' alleged virtues of robustness, such as what he calls 'pragmatic reliability' (26), are set aside as 'irrelevant' (27).

Hudson's crusade against robustness develops through two kinds of arguments: systematic philosophical analyses and discussions of historical case studies. On the side of the case studies, five episodes are analysed, four in physics and one in microbiology. They are intended to work as an empirical basis for the philosophical positions defended in the book. Hudson insists that his philosophical theses about robustness are partly suggested by the episodes under scrutiny, which themselves are supposed

to reveal 'how scientists actually go about establishing the reliability of observed results' (51). Thus, Hudson's critique is conceived, and valued by its author, as a historically informed and empirically supported one, in contrast to a—today largely discredited—epistemology that would be elaborated 'exclusively on philosophical, a priori grounds' (xvii).

The five case studies are analysed through the same frame. A reading of the historical episode is proposed, according to which what scientists actually do, when they seem to use robustness reasoning for justifying the truth of an experimental result *R*, is, under examination, not an instance of robustness, but the exemplification of some truly different experimental strategy—according to the case, calibration or targeted testing. Thus, what scientists actually do is not what many philosophers claimed they have done, and possibly not either what scientists themselves claimed to have done (scientists sometimes 'misrepresent' their actions, xix). The conclusion of the case studies is: 'robustness reasoning does not play much of a role in how scientists justify their observed results' (169). Such a conclusion is understood as a significant empirical indication that robustness does not possess the high epistemic value philosophers commonly attribute to it.

The systematic philosophical analyses confirm this. Multiple routes are followed, with the crucial issue of independence of the *Ps* as a central thread.

Hudson first criticizes what he identifies as three philosophical strategies 'to put the legitimacy of robustness reasoning on firm footing' (1). He starts with 'probabilistic approaches' (8) and concludes that 'we need to comprehend this independence in a nonprobabilistic way' (24). I fully agree. 'Pragmatic approaches', of which authors like Wimsatt and Kent Staley are supposed to be representatives, are dismissed as irrelevant to epistemic issues. Relying on 'epistemic independence approaches to robustness' (primarily represented by Peter Kosso), the multiple *Ps* that converge on *R* in a robustness scheme should use different (or not too similar) assumptions. Hudson's punch line here, as far as I understand it, is that we do not need (or perhaps not necessarily need? Or cannot get anyway?) epistemic independence. Scientists can achieve justified, truthful results without epistemic independence (through strategies that differ from robustness which, for its own part, requires epistemic independence). What primarily matters according to Hudson is not epistemic independence, but 'empirical sensitivity' (50). The acceptable, likely-to-be-true results can be generated by, and justified on the basis of, epistemically dependent *Ps*, providing that the corresponding proposals are 'motivated empirically' and that scientists remain 'genuinely open' (49), that is, ready to abandon their favourite assumptions when 'the empirical pressure becomes enough to force a change, at which point it would be a mistake to continue reinterpreting experimental situations in accordance with one's theoretical preconceptions' (51).

Since all the available philosophical 'strategies for defending the value of robustness' (8) identified by Hudson 'have their irremediable weaknesses' (2), the ground for robustness seems to collapse. Hudson's crusade, however, is still not rid of its targeted enemy belief, because behind various more or less technical versions of robustness, Hudson identifies a 'fundamental philosophical insight' (2) which 'drives the

proponents of robustness' (24) and 'that many have found compelling' (2). Hudson calls it 'the "core" argument for robustness' (2), and 'concede[s] its intuitiveness' (24).

The key idea is that when a robustness scheme is instantiated, this legitimately increases our confidence that *R* is more likely to be true and the *Ps* are more likely to be reliable, because providing that the *Ps* are truly independent in the sense of very different, 'it is unlikely that the *same* artifact could be independently produced' (24). In more metaphoric terms, it would be a 'miracle' if the same 'physical (or theoretical) bias' (24) were involved in all the so-different *Ps* at stake, and responsible for the production of the same *R*. A better explanation, if not the only explanation, is that *R* is (at least approximately) true and each *P* is (at least globally) reliable. This "no-miracles" argument for robustness' (1) is obviously a version of the no-miracles argument often considered as the main argument in favour of scientific realism.

Having stressed, as others (e.g. Soler 2012b, section 10.3; Stegenga 2012, 210), that some form of no-miracle reasoning lurks behind robustness and is responsible for its intuitive convincing power, Hudson articulates a systematic philosophical critique of the core argument, following different lines. Robustness is found poorly defined, not operative, 'epistemically limited' (200), and unable to 'ground representational accuracy' (195). Looking back to his historical cases with the aim to examine the attitude of practitioners regarding 'converse robustness' (179), Hudson again argues that scientists 'are not really active proponents of robustness reasoning' (182). A few pages are dedicated to the formal sciences, from which it is concluded without further ado that 'no robustness [is] found in mathematics and logic' (189). The upshot of the philosophical critique is that 'the core argument is ultimately unsuccessful' (174).

Overall, multiple empirical and philosophical arguments converge on the same 'cumulative conclusion' (169), i.e. Hudson's discovery claim (I leave it to the reader to decide whether we have here an instance of robustness reasoning in philosophy): robustness is of 'very little, if any, value' (169). Scientists usually do not appeal to such 'epistemically limited' strategy. 'The wide support for robustness reasoning found in the philosophical literature really is the invention of the philosopher of science' (244).

There is, however, according to Hudson, an explanation for the 'evident popularity' (xx) of robustness, but this is a sociological explanation. From a sociological point of view, robustness reasoning possesses an 'interesting feature' (199) and a 'particular merit' (which is also its 'main drawback' from an epistemic point of view): it is a 'low-threshold strategy' (200), which means that just mentioning the convergence of several independent *Ps* on the same *R*, without needed to discuss the details of the *Ps*, can be convincing. Such a strategy is interesting and used by scientists in one case: when 'a scientist is pressured by a disputational situation to find a justification for her observed results *that extends beyond what the basic empirical findings tell her*' (198, italics added). Thus, there is at least one case in which living scientists actually make extensive use of robustness according to Hudson. In this case, robustness is an easy, efficient source of 'social benefits' (200), to the extent that it often convinces peers and can '(at least temporarily) resolve intellectual disputes simply by virtue of appealing to the "presumed independence"' of the *Ps* (200). This, of course, does

not change the fact that from an epistemic point of view, robustness is more than limited.

3. Critical Comments

Hudson's book has several invaluable merits. First, it shows the instantiation in science of, and characterizes, two recurrent forms of experimental reasoning, calibration, and targeted testing. In this respect, Hudson's project is close in spirit to Allan Franklin's 'epistemology of experiment'. The book can be read as a proposal to extend Franklin's set of experimental 'strategies that provides reasonable belief in experimental results' (Franklin 2012).

Second, the book shows the indispensable character of, and encourages us to undertake, a systematic detailed examination of the exact possible meanings, presuppositions, and legitimate conditions of application and success, of robustness. Anyone who believes that *just* the simple report that several intuitively independent experimental accounts all yield the same experimental result *R is enough* to conclude that *R* is robust and likely to be true, will learn a lot about what is additionally required.

Furthermore, Hudson's complex subject matter is presented in a very pedagogic way. The thread is limpid, and the main conclusions are always summarized at the end of each development. The five case studies, although complicated and technical episodes, are introduced in a way that makes their substance accessible and enables the reader to understand the precise function they serve in the author's arguments.

Still, I have not been converted by Hudson's crusade. Or perhaps it would be equivalent to say that, at a certain level, I was already converted to a number of Hudson's central claims, as a result of my own antecedent research on robustness, but that notwithstanding, I do not think that Hudson's higher-level translation in terms of 'no robustness in science' and 'no epistemic value for robustness' (my formulations) is an appropriate characterization of the situation—and this is not just a matter of words. Below, I indicate why. I start with a general formulation of several problematic points, referenced by letters (a) to (c). Then I discuss some particular forms these problems take in the book, in a way that attempts to provide partial evidence for the soundness of my critique.

- (a) Several lines of arguments seem to be directed against a conception of robustness which is either (a1) hubristic—robustness would be the only way to establish the reliability of experimental conclusions—or (a2) excessively narrow—just to appeal to the fact of the convergence of intuitively independent experiments would be enough. The reader fails to identify those who are supposed to endorse such conceptions, and can suspect the invention of a straw enemy for the sake of discovery claims.
- (b) Another set of arguments, which run down robustness as poorly defined and deprived of operative empirical criteria of application, could be turned as well against Hudson's favourite strategies, and more generally seem to ask for the impossible.

- (c) Alternative accounts of Hudson's case studies can be provided, that make room to robustness while not denying the possible involvement of targeted testing and calibration.

Other claims or assumptions of the book, that cannot be discussed here, could be questioned. Some must at least be mentioned, because they are sometimes implicitly involved in Hudson's positions as reconstructed below. Hudson suggests that it is always better to attempt to improve the reliability of a single procedure than to examine whether or not different procedures converge on the same results (see e.g. 7–8). He also suggests that when two *Ps* converge on one *R*, if one *P* is less reliable than the other, then it counts for nothing (which would imply that in a conclusive robustness scheme, all *Ps* have equal reliability). Though such claims can hold in some circumstances, they are arguably not universally valid.

(a1) *Hubristic conception of robustness*. Hudson rightly stresses that something like a requirement of 'minimal reliability' (18) must apply to each *P*, for a *P* to be legitimately included in a robustness reasoning, and for a robustness reasoning to have a chance to be a valid argument. Several scholars have insisted on this point. As an amusing illustration, Hubertus Nederbragt mentions a situation imagined by Harry M. Collins, in which a scientific hypothesis under discussion would be tested, among other ways including experiments, by examining the entrails of a goat. This 'method' is indeed independent of the recourse to experiments, 'but', Nederbragt urges, 'here we reject immediately and with force the whole body of background knowledge of this method' (Nederbragt 2012, 138; for an analysis of the meaning of minimal reliability-like requirements, and other illustrations, see Soler 2012a).

The problem is that Hudson, rather than viewing the minimal reliability requirement as one condition under which the use of robustness can be epistemically valuable, amazingly turns this requirement against robustness: 'The recognition that robustness reasoning assumes the (at least minimal) reliability of alternate observational routes and that it is ineffective at establishing this reliability to begin with forms a key part of my critique of robustness' (6). But even if one is ready to concede that robustness is necessarily 'ineffective at establishing' the minimal reliability of each *P*, so that other strategies are always needed to achieve this aim, why should this taken against robustness? This is an argument against the epistemic value of robustness, solely under the hubristic contention that robustness is the only means to ensure the reliability of scientific items, and an unconditionally relevant and successful strategy. To my knowledge, nobody has ever defended such a position.

(a2) *Unduly narrow definition of robustness*. Although not explicitly stated in Hudson's initial definitions of robustness and of the core argument, the reader is progressively led to infer from Hudson's text that a robustness reasoning is necessarily 'blind' to the contents of the *Ps*—otherwise, it is not robustness. Put differently, Hudson seems to equate robustness with (what I call) *blind* robustness. '[O]ne should be suspicious', he writes,

of a method that can be applied in complete ignorance of the details of a scientific case, as is true with robustness reasoning where all one presumably needs to know is

that two (or more), minimally reliable observational procedures independently converge on the same result, leaving aside all the technical details of how these procedures arrived at these results. (245; see also 184–185; 198–200)

I fully agree that we should be suspicious about such a method, but why should ignorance of the details of the *Ps* be a definitory feature of robustness?

I easily concede that, in *some* contexts, people—scientists or philosophers of science—sometimes *just claim* that some *R* is reliable (i.e. robust) because multiple independent experimental procedures converge on it, without giving any additional details. For example, retrospective reconstructions, by scientists or analysts of science, of the first historical validation of a scientific result presently taken as firmly established, could be foregrounded as an especially favourable context for, and a possibly contestable and worrying use of, blind robustness.

Hudson, for his part, puts forward scientific discourses addressed to students and non-specialist audiences, and, as far as scientific debates among peers are concerned, cases in which ‘a scientist is pressured by a disputational situation to find a justification for her observed results that extends beyond what the basic empirical findings tell her’ (198). But to him, these cases are not just *some* contexts in which scientists use (blind) robustness. Part of Hudson’s discovery claim contends that such cases *exhaust* the real cases in which scientists actually make extensive use of robustness reasoning. In the other cases—i.e. when ‘the basic empirical findings’ tell enough regarding justification purposes (I leave aside the problematic character of the frontier between ‘what the basic empiric findings tell’ and ‘what is beyond’)—what scientists actually almost always do, according to Hudson, is to scrutinize the details of the available experimental procedures with the aim to discuss/test/improve the reliability of these procedure, so that the reasoning involved is not robustness, but one form or another of ‘reliability process reasoning’.

Taking for granted this reconstruction of Hudson’s position, the conclusion that scientists actually make no, or only peripheral, use of robustness, appears to depend on a very narrow background conception of robustness—i.e. blind robustness. But why should we reduce robustness reasoning to blind robustness? Admittedly,

When asked why an observational process ... [and a related observational result are] reliable, a scientist will need to do much better than *simply cite* the convergence of this process’s results with its results with another (minimally reliable) observational procedure. (200, italics added)

But has any philosopher, including robustness fans, ever denied that? Admittedly—or so shall I concede for the sake of the argument—in some contexts, scientists or philosophers appeal to blind robustness with the hope to convince their colleagues, students, or not-specialist audiences, and this sometimes works—though concerning living debates among contemporary scientific peers, we should carefully distinguish ‘genuinely blind’ robustness from black-boxed robustness, i.e. cases in which details of the *Ps* are not explicitly discussed for contextual reasons, but are known and could be discussed. However, blind robustness cannot be unproblematically identified with the common understanding of robustness. I see indications that it cannot even in Hudson’s text itself. When stating his conclusions, Hudson talks, not of robustness *tout*

court, but of “pure” robustness reasoning (i.e. robustness unaffiliated with either targeted testing or Perrin-style calibration)’ (142, italics added; see also 169). That is, he feels the need to introduce qualifications to specify the target of his critique—which would not be needed if ‘pure’ (i.e. blind) robustness were the usual understanding.

I conclude that blind robustness is an unduly narrow conception of robustness, and that an important part of Hudson’s critique applies only to blind robustness. In particular, what Hudson points to as the ‘main drawback’ (200) of robustness—i.e. robustness is a ‘low-threshold strategy’ that attempts to convince at an epistemologically low cost, just by citing the convergence instead of scrutinizing the details of the *Ps* and discussing their reliability—is the main drawback just of blind robustness. I take as a more appropriate and philosophically fruitful attitude to work with a broader and more usual conception of robustness—according to which robustness reasoning can occur in the background of a *more or less* deep understanding of the contents of the *Ps*, a *more or less* detailed discussion of the reliability of the *Ps*, and a *more or less* resulting high confidence in *R*—and to relativize assessments of the epistemic value of a robustness reasoning to these ‘more or less’ variables.

(b) *Excessively strong and asymmetrically imposed demands on robustness.* One—perhaps *the*—main charge of Hudson against robustness concerns the independence requirement involved in the definition of robustness. The independence condition is indeed a crucial one: if, under examination, the *Ps* prove to be dependent (e.g. all the experimental accounts actually share some questionable assumption), this could be the explanation of their convergence on *R*. Put differently, the situation is not just either a miracle, or the reliability of *R* and the *Ps*. A third possibility is fake independence. All philosophers who have systematically worked on robustness have recognized the need to undertake a thorough investigation of the independence of the *Ps* before positing a verdict of robustness. Wimsatt is one of those who early and strongly insisted on this point. He warned of ‘illusions of robustness’, and identified ‘the failure of the different supposedly independent tests, means of detection, models, or derivations to be truly independent’ as ‘one of the most critical and important problems in the study of robustness analysis’ (Wimsatt [1981] 2012, 82, 83).

Thus, Hudson is perfectly right to point out that a discussion of independence is decisive regarding the aim of assessing robustness. Hudson is equally right to insist that ‘The question of what we mean here by “independent” is a substantive one’ (xiv). Too often the independence issue, when addressed in a non-probabilistic perspective, is equated without discussion to ‘no common hidden hypotheses in the *Ps*’, whereas under examination, the situation is much more complex (see Soler 2012a, section 1.8). Several recent proposals have been made to distinguish types of independence and to build scales of independence in relation to robustness (see, in particular, Nederbragt 2012; Soler 2012a; Stegenga 2012; Trizio 2012, in addition to the references discussed in Hudson’s book).

Beyond discussions about the types of independence that are relevant in relation to robustness (physical? Epistemic? No influence of colleagues’ work?), Hudson’s main critique, which is supposed to apply both to physical and epistemic independence, is this: ‘the core argument for robustness ... is ... found to be questionable

due to our *inability to adequately explain what it means* for two observational processes to be independent of one another *in a way that is informative* (xix, italics added).

To assess the independence of the multiple *Ps* involved in a robustness scheme, we must discuss the differences and similarities between these *Ps*. But, Hudson insists, quoting Steven Woolgar, “sameness” or “difference” is not an inherent property of (sets of) phenomena’ (170). They presuppose evaluations that are ‘not a straightforward matter’ and ‘require in every case a certain degree of theoretical or empirical acumen’ (171). The *Ps* ‘mustn’t be too different’ (172), in particular because they must ‘at least be about the same subject matter’ (172)—otherwise they could not converge on the same *R*. But the *Ps* must be ‘different enough’ (172)—otherwise the convergence could be explained by the common points shared by the *Ps*. In brief, to decide whether or not the independence condition built into the core argument for robustness is satisfied requires ‘an assessment of relevance’ (172) for deciding whether the *Ps* are ‘different in the right degree’ (172). But proponents of robustness, Hudson deplores, do not provide any precise definition or empirical criteria for the ‘right degree’ of difference that is required for a robustness scheme to have justificatory power. ‘[T]here’s no guidance at all on how this medium amount of difference is to be determined’ (172). This is why the core argument fails according to Hudson: because ‘the notion of independence . . . lacks clarity’ (173), and because we are ‘left with the problem of determining when, in fact, processes are independent in a way that is suitable to ground a robustness argument’ (170).

Hudson extends the principle of this critique to two further conditions that should be satisfied, in addition to independence, for a robustness scheme to work as a genuine justification: in my terms, the condition of a ‘sufficient’ number of *Ps*, and the condition that each *P* considered in isolation is ‘sufficiently’ reliable. In substance, Hudson complains that the ‘sufficient’ measure is not specified, and denounces this as one of the ‘characteristic weaknesses of robustness reasoning’ (112).

Overall, Hudson’s point is

that supporters of robustness reasoning say precious little about the details of how this reasoning is to be applied. For example, which of the many possible independent procedures should be utilized . . . [condition of sufficient reliability]? How different should these alternate procedures be [condition of independence], and how many of them should be used . . . [condition of sufficient number]? In the literature, robustness reasoning is often presented in such an abstract form that how to use it effectively in practical terms is left unclear. (xx)

Hudson summarizes the situation by talking of ‘the “excessive abstractness” of robustness reasoning’ (187).

To clarify the conditions involved in the judgement that an instance of robustness reasoning is relevant and valid is indeed an important task. If interpreted in this perspective, Hudson’s above-mentioned points are a valuable contribution. We could actually go even further in this direction, and point to additional conditions involved in the judgement that a robustness scheme can work as a justification (see Soler 2012a for a systematic analysis), that Hudson regrettably does not address frontally. For

example, the *Ps* are supposed to converge on the same *R*, but in practice, the identity is rarely, if ever, given as such. Rather, the *Ps* end in strictly speaking different *Rs*, which are then judged similar enough to be taken as one and the same *R* with respect to some purpose. The sameness condition involved in a robustness scheme is thus itself the product of a judgement of degree. Furthermore, it is important to realize that several of the conditions discussed above are not independent. For example, counting and weighting the *Ps* that are required for a robustness scheme to be effective are not two independent operations: *ceteris paribus*, the more reliable each *P* is, the less a high number of *Ps* is needed. Overall, a robustness assessment can, from an analytical point of view, be decomposed in at least four (not independent) degree judgements: *R* is (more or less) robust = a (greater or smaller) number of (more or less reliable) derivations that are (more or less) independent lead to a result (more or less close) to *R*. And even this is not the end of the story, because when assessing the power of a robustness scheme, we should also take into account what is ‘outside’ it, that is, possible available procedures that would support either some results *R'* in conflict with *R*, or alternative accounts of *R* (point (c) below provides an example of such a configuration).

Hudson is perfectly right that supporters of robustness do not specify the right degree of independence, right number of *Ps*, and so forth. More generally, they do not specify the various ‘more or less’ evaluations in which an assessment of robustness can be analysed, in the sense that no definition or empirical criteria are provided that would enable to indicate a threshold, or even a rough interval, below which *R* would be fragile and above which it would be robust. However, this so-called excessive abstractness cannot be turned against robustness, first, because it applies as well to all the other strategies valued by Hudson, and second, because more fundamentally, it cannot but apply to any philosophical project dedicated to the characterization of schemes of action through which scientists establish reliable results. As twentieth-century meta-inquiries into scientific method in the empirical sciences amply demonstrated, no definition, rule, or criteria can be stated, that would include universal compelling conditions of right application. Pragmatic, context-sensitive judgements are pervasively involved in scientific research, especially when degree and relevance appraisals are in play. If it remains possible to exhibit schemes of ‘experimental strategies’ recurrently used by scientific practitioners (in Franklin’s spirit), it would be asking the impossible to require that robustness proponents provide context-independent operative recipes or criteria regarding the right degree of independence, minimal degree of reliability, or adequate number of *Ps*. (Incidentally, the same remarks apply to Hudson’s ‘point’ at which ‘the empirical pressure’ is supposed to ‘becomes enough to force a change’, cf. the discussion of epistemic independence above).

In some passages, Hudson seems to be aware of the problem (see, in particular, 245). In fact, he even explicitly applies considerations of this kind to his ‘reliable process reasoning’. His definition of the latter (229) is actually even more abstract and unspecified than a robustness scheme. Conscious of this excessive abstractness, Hudson writes: ‘it is left as an open variable what to regard as a reliable process, but that is only because we leave it to the scientists themselves in the context of their respective fields to fill out

these details' (163). What is amazing, though, is that the same logic is not symmetrically applied to robustness. With robustness as well, we leave to scientists in context what to regard as a sufficient number of sufficiently reliable and sufficiently independent P s.

(c) *Questionable mutual exclusiveness between targeted testing and robustness.* Hudson's evidence against robustness based on historical case studies of course depends, as any historically based philosophical proposal, on an interpretation of the particular episodes under scrutiny, and more generally on a historiographical approach oriented by some (possibly more or less explicit) assumptions. There is no question of going into the details of Hudson's accounts of the case studies here. Instead, I shall provide a schematic characterization of Hudson's accounts of the episodes he categorized as cases of targeted testing, and shall explain why I think the historical configurations he deals with can, if not should, be understood as cases in which robustness is instantiated *in conjunction with* targeted testing.

When accounting for episodes categorized as cases of targeted testing, schematically, Hudson tells the story as follows. At a given stage of the studied episode, multiple independent 'observational procedures' (say P_1 and P_2) lead to multiple observations that support the same result R (for example, the existence of dark matter or dark energy). In my words, a robustness scheme is instantiated. However, R is still not considered as firmly established, because alternative competing hypotheses are at play (say R'), which could also account for the multiple observations that support R (for example, the effects of extinction and evolution could 'mimic' the observations that support the existence of dark energy). Given that, astrophysicists 'attempt to resolve the . . . controversy once and for all' (144). An ingenious experiment is conceived, which shows (say through P_3) that, in some relevant situations, whatever the truth or falsity of R' , R is still the case (in Hudson's vocabulary: this experiment 'target tests' R'). Hudson's conclusion is: R has been established by this last observational procedure P_3 alone; the strategy that 'ultimately decide[d] the issue' (168) corresponds to targeted testing, and more generally to a form of reliability process reasoning; robustness has nothing to do with this demonstration. For example: 'In the end it was only the unique evidence provide[d] by . . . [P_3] that settled the matter. This evidence established the existence of dark matter, something a robustness argument could not do' (145; see also 168).

Hudson's story immediately raises a problem. The claim that the last observational procedure P_3 alone is responsible for the validation of R is highly questionable. No argument is provided for this strong claim, and the claim is frankly doubtful. Suppose only the last experiment had been performed: would R be considered firmly established? At least, this would need discussion.

Moreover, relying on the information Hudson himself provides about the episodes under study, the status of the competing hypothesis R' under target test could be specified differently. Admittedly, we can present R' , as Hudson does, as a hypothesis alternative to R . But R' can also be viewed as part of the hypotheses one needs to assume to be entitled to conclude that the P s are sufficiently reliable to corroborate R . For example: P_1 supports the existence of dark energy, *provided that* in the account of the experiment involved, the effects of extinction and evolution have been correctly handled; and *idem* for P_2 .

Such a perspective on R' suggests the following story alternative to Hudson's. At a given stage of the episode studied, P_1 and P_2 lead to observations that support the same result R . However, R is still not considered as firmly established, because P_1 and P_2 share some assumption R' (e.g. extinction and evolution effects are not able to mimic R), which worries some practitioners. In other words, each P considered in isolation is not reliable enough, and P_1 and P_2 are not sufficiently independent to establish R , in the sense that something that they have in common (R') is problematic. R' might be responsible for the convergence, so that the seemingly robustness would be spurious. Consequently, R' must be investigated (target tested). For this purpose, some scientists perform an ingenious new experiment, and conclude (through P_3) that in some relevant situations under inquiry, R is the case whatever the truth-value of R' is. The upshot is that scientists are now left with three observational procedures that converge on R . The reliability of both P_1 and of P_2 is enhanced by P_3 , which precisely aimed to target test one possibly problematic shared hypothesis R' involved in each of these two experimental derivations of R . Consequently, we now have two individually more-reliable-than-before- P_3 procedures, P_1 and P_2 , which converge on R . Concerning the reliability of P_3 , there are two possibilities: either we see P_3 as 'sufficiently' reliable 'in itself', or we consider that confidence in P_3 is partly provided by the fact that P_3 itself supports a result R that already has partial support provided by the two already available experimental procedures. Without being able to discuss such possibilities, I think the second is arguably more plausible, at least in the episodes under examination. Anyway, after P_3 , the situation is that three partially different, taken-to-be sufficiently reliable experimental accounts converge on R . Admittedly, P_3 plays a specific role with respect to P_1 and P_2 , to the extent that in addition to reinforcing R (by comparison with the situation in which only P_1 and P_2 were available), P_3 'target tests' a common hypothesis of P_1 and P_2 . P_3 is a means to increase the degree of reliability of the P s involved in the previously available robustness scheme. But the use of this means along the road does not entail that P_3 cancels the two previously available procedures so that we can conclude that only P_3 is responsible for the validation of R , and hence that no robustness is involved here.

I conclude that targeted testing and robustness are not mutually exclusive strategies. The first might be involved as a means to achieve the satisfaction of some of the conditions that a robustness scheme must meet to work as a sufficient justification.

4. Conclusion

To accept Hudson's conclusion as a discovery claim, we must at least (i) endorse very strong requirements about the power of robustness reasoning, (ii) work with an excessively narrow definition of robustness, and (iii) adhere to some questionable historiographical assumptions. To me, the numerous interesting developments of Hudson's book would have benefited from a disentanglement from the discovery claim, which works as a parasitic noise. Instead of the discovery claim that robustness is not used by scientists and has no epistemic value, I suggest a less radical, but more appropriate and still important message, namely that robustness reasoning has relevance and value *only under some*

conditions. The corresponding conditions can be identified and specified up to a point, but their assessment is a pragmatic matter left to professional scientists in the context of their practice. In this perspective, robustness is no more, but also no less, than *one* particular possible ‘reliable processes reasoning’ *among others*.

References

- Calcott, Brett. 2010. “Wimsatt and the Robustness Family: Review of Wimsatt’s *Re-engineering Philosophy for Limited Beings*.” *Biology and Philosophy* 26: 281–293.
- Franklin, Allan. 2012. “Experiment in Physics.” In *Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta. <http://plato.stanford.edu/archives/win2012/entries/physics-experiment>.
- Nederbragt, Hubertus. 2012. “Multiple Derivability and the Reliability and Stabilization of Theories.” In *Characterizing the Robustness of Science: After the Practice Turn in the Philosophy of Science*, edited by Léna Soler, Emiliano Trizio, Thomas Nickles, and William C. Wimsatt, 121–146. Dordrecht: Springer.
- Soler, Léna. 2012a. “Introduction: The Solidity of Scientific Achievements. Structure of the Problem, Difficulties, Philosophical Implications.” In *Characterizing the Robustness of Science: After the Practice Turn in the Philosophy of Science*, edited by Léna Soler, Emiliano Trizio, Thomas Nickles, and William C. Wimsatt, 1–66. Dordrecht: Springer.
- Soler, Léna. 2012b. “Robustness of Results and Robustness of Derivations: The Internal Architecture of a Solid Experimental Proof.” In *Characterizing the Robustness of Science: After the Practice Turn in the Philosophy of Science*, edited by Léna Soler, Emiliano Trizio, Thomas Nickles, and William C. Wimsatt, 227–266. Dordrecht: Springer.
- Stegenga, Jacob. 2012. “*Rerum Concordia Discors*: Robustness and Discordant Multimodal Evidence.” In *Characterizing the Robustness of Science: After the Practice Turn in the Philosophy of Science*, edited by Léna Soler, Emiliano Trizio, Thomas Nickles, and William C. Wimsatt, 207–226. Dordrecht: Springer.
- Trizio, Emiliano. 2012. “Achieving Robustness to Confirm Controversial Hypotheses: A Case Study in Cell Biology.” In *Characterizing the Robustness of Science: After the Practice Turn in the Philosophy of Science*, edited by Léna Soler, Emiliano Trizio, Thomas Nickles, and William C. Wimsatt, 105–120. Dordrecht: Springer.
- Wimsatt, William C. [1981] 2012. “Robustness, Reliability, and Overdetermination.” Reprinted in *Characterizing the Robustness of Science: After the Practice Turn in the Philosophy of Science*, edited by Léna Soler, Emiliano Trizio, Thomas Nickles, and William C. Wimsatt, 61–88. Dordrecht: Springer.
- Wimsatt, William C. 2012. “Robustness: Material, and Inferential, in the Natural and Human Sciences.” In *Characterizing the Robustness of Science: After the Practice Turn in the Philosophy of Science*, edited by Léna Soler, Emiliano Trizio, Thomas Nickles, and William C. Wimsatt, 89–104. Dordrecht: Springer.
- Woodward, James. 2006. “Some Varieties of Robustness.” *Journal of Economic Methodology* 13: 219–240.