

What exactly is stabilized when phenomena are stabilized?

Uljana Feest

Received: 20 May 2009 / Accepted: 3 June 2009 / Published online: 2 July 2009
© Springer Science+Business Media B.V. 2009

Abstract The last two decades have seen a rising interest in (a) the notion of a scientific phenomenon as distinct from theories and data, and (b) the intricacies of experimentally producing and stabilizing phenomena. This paper develops an analysis of the stabilization of phenomena that integrates two aspects that have largely been treated separately in the literature: one concerns the skills required for empirical work; the other concerns the strategies by which claims about phenomena are validated. I argue that in order to make sense of the process of stabilization, we need to distinguish between two types of phenomena: phenomena as patterns in the data (“surface regularities”) and phenomena as underlying (or “hidden”) regularities. I show that the epistemic relationships that data bear to each of these types of phenomena are different: Data patterns are *instantiated* by individual data, whereas underlying regularities are *indicated* by individual data, insofar as they instantiate a data pattern. Drawing on an example from memory research, I argue that neither of these two kinds of phenomenon can be stabilized in isolation. I conclude that what is stabilized when phenomena are stabilized is the fit between surface regularities and hidden regularities.

Keywords Data · Phenomena · Stabilization · Validation · Bogen/Woodward · Hacking · Epistemology of experimentation

1 Introduction

In the philosophy of science literature of the past 20-some years or so, the term “phenomenon” has been used by the proponents of two, partially overlapping, bodies of literature: One goes back to [Bogen and Woodward \(1988\)](#) paper “Saving the

U. Feest (✉)
Institut für Philosophie, Wissenschaftstheorie, Wissenschafts- und Technikgeschichte, Technische
Universität Berlin, Straße des 17. Juni 135, Sekr. H72, 10623 Berlin, Germany
e-mail: uljana.feest@tu-berlin.de

Phenomena”. In this article, the authors introduced the notion of phenomenon as distinct from theories and data, thereby questioning the traditional dichotomy between theory and observation as well as the evidential and explanatory relations usually thought to hold between them (Bogen and Woodward 1988). The other body of literature, mainly associated with the history and philosophy of experimentation, has begun to take a much more detailed look at the physical and cognitive intricacies of identifying and “stabilizing” phenomena in experimental contexts (see Hacking 1983, p. 239). While there are obvious affinities between the philosophical intuitions expressed by these two approaches, their respective insights and suggestions have not, to my knowledge, been explicitly brought to bear upon each other. In this article, I provide such an analysis. In doing so I show that a careful analysis of the notion of *stabilization* can be utilized to fine-tune the notion of *phenomenon* by differentiating between two kinds of phenomena that have—in my view—not been sufficiently distinguished in the literature: surface phenomena and hidden phenomena. I conclude by arguing that what is stabilized when phenomena are stabilized is the mutual fit between these two kinds of phenomena, and I emphasize the substantial conceptual work that is required in the process.

I begin by providing some background for my usage of the key terms of this article: “phenomena” and “stabilization”, drawing an analytical distinction between two aspects of stabilization that have been emphasized by different contributors to the literature, respectively: a skill- aspect and a validational aspect (Sect. 2). It is then argued that a philosophically satisfactory account of stabilization has to integrate these two aspects. This claim is made plausible by refuting two arguments that might be put forward in favor of treating skill and validation separately (Sect. 3). I then go on to develop a sketch of a positive account of the stabilization of phenomena. This account draws on two crucial points developed in the previous section, i.e., (a) a distinction between two types of phenomena, and (b) an emphasis on the conceptual prerequisites that are needed to identify data points as instantiating or indicating phenomena (Sect. 4). This analysis is followed by a discussion of an example from memory research (Sect. 5). I conclude with some reflections on the scope of my analysis.

2 The key terms explained

The ways in which the notion of *phenomenon* is used in the above-mentioned literature self-consciously positions itself vis-à-vis traditional philosophy in two ways: (1) In contrast to traditional epistemological and phenomenological usages, where phenomena are usually associated with fleeting sensory experiences, or with the irreducible subjectivity of such experiences, phenomena are here taken to be “relatively stable and general features of the world” (Woodward 1989, p. 393). Hacking in particular makes it quite clear that his usage of the term “phenomenon” is not informed by philosophical tradition, but rather follows that of physics, where phenomenology is a branch of research that attempts to provide detailed calculations and descriptions of a given object of research, especially in an experimental context (Hacking 1983, Chap. 13). While Bogen and Woodward are less explicit about this point, their usage is also clearly inspired by scientific, rather than philosophical, usage. This choice of terminology reflects their intention to do justice to scientific practice. (2) By using the

term “phenomenon” in the way they do, Bogen and Woodward, like Hacking, not only distance themselves from certain usages in traditional epistemology, but also express a critique of a widespread assumption within twentieth-century philosophy of science, according to which the notion of phenomenon is closely tied to an idea of unproblematic observability. Countering this assumption, Bogen and Woodward emphasize a distinction between that which is actually empirically registered by scientists (the data), and that which constitutes the objects of scientific research (the phenomena). Their claim is that scientists do not directly observe phenomena, but rather use data to gain evidence for claims about phenomena. Hacking is less concerned with the question of whether phenomena are observable in principle. His point rather is that even if we grant that phenomena are observable, it is not at all a trivial matter to actually observe them. He thereby draws attention to the fact that empirical work in the sciences is not a matter of passively registering facts about the world, but that a good deal of empirical work in the sciences consists on efforts *to make* phenomena observable.¹

Prima facie it may appear that [Bogen and Woodward \(1988\)](#) on the one hand and [Hacking \(1983\)](#) on the other are directly contradicting each other, since Hacking appears to affirm what Bogen and Woodward want to deny, i.e., that phenomena are observable. However, as I hope to show in this paper, the two analyses are in fact compatible. In order to appreciate this, we have to make slight amendments to both accounts. First, Bogen and Woodward are right to emphasize that phenomena are not observable. Second, Hacking is right to draw attention to the intricacies of empirical work, though he is probably ill-advised to use the term “observation” in this context, especially since he, like Bogen and Woodward, wants to distance himself from much of the philosophical baggage that comes with the term. These pitfalls and potential misunderstandings can be avoided if we assert that what Hacking really means when talking about the complexities of making a phenomenon observable is what he, in other places, refers to as “stabilization”. It is therefore this concept that I want to turn to next.

The notion of *stabilization* is found in some of the literature on scientific experimentation, though its scope is not restricted to experimental sciences, but also encompasses other forms of empirical work. Roughly, this notion is taken to refer to the processes whereby scientists (a) empirically identify a given phenomenon or entity, and (b) gradually come to agree that the phenomenon is indeed a stable and robust feature of the world, rather than being (say) an artifact of any particular experiment or instrument. Philosophers tend to think of this process as one that is rationally guided (or can at least be rationally reconstructed), whereas proponents of the social studies of science have focused more on other (e.g., social) mechanisms that contribute to scientific consensus formation. However, even within the philosophical community, it is generally taken for granted that the process of stabilization has two aspects: One has to do with the ability to make empirical results physically stable, i.e., to “make an experiment work”,

¹ Of course there is a finer distinction to be made here, regarding the question of whether Hacking means that certain phenomena literally did not exist before they were produced in the laboratory, or whether he merely means that the laboratory sciences make stable phenomena empirically accessible by producing data that instantiate them. As will become apparent, it is this latter reading that I endorse.

or make an instrument reliably reproduce particular results, where this clearly involves the ability to recognize that the experiment/instrument did in fact work. The other has to do with strategies of establishing that the phenomenon is genuine, or “robust”. Versions of the first aspects were touched on by Kuhn (1962) and Polanyi (1966), with their emphasis on tacit knowledge in the use of scientific concepts. It is this aspect that Hacking has in mind when he talks about Caroline Herschel’s remarkable ability to identify and classify planets, or about the ability of certain lab assistants to read X-ray pictures (Hacking 1983, Chap. 10). The focus here is on the question of how scientists (individually or collectively) come to converge in their delineation and classification of relevant phenomena. Versions of the second aspect of stabilization are treated in the literature on the calibration of instruments and the validation of claims about phenomena. Keywords here are “robustness” (e.g., Wimsatt 1981; Hacking 1981; Culp 1994, 1995) and the experimental strategies scientists draw on in establishing the validity of a finding. Such strategies include, amongst others, that of “multiple determination”, i.e., the identification of a phenomenon by means of more than one instrument (e.g., Franklin 1993; Woodward 1989).

In the following, the two aspects of stabilization outlined in the previous paragraph shall be referred to as the “skill”- and the “validation”-aspect of stabilization. Let me clarify the distinction by responding to a potential objection. The objection states that if the skill aspect of stabilization is connected to some form of tacit *knowledge* then this implies that the results delivered by a person who possesses the skill would thereby automatically also be validated. In response I would like to suggest that the practical knowledge required to run an experiment or to use an instrument has to be distinguished from the kinds of arguments required to back up the claim that the results of the experiments or instrument readings indeed indicate the type of thing the scientist take it to be. For example, while Caroline Herschel may have been better than anybody else at using the light telescope to identify and classify a particular kind of structure, it is still conceivable that this structure might have turned out to be an artifact of the instrument rather than indicating a heavenly body, let alone a planet. This would have become apparent later, when different kinds of telescopes were used to replicate her findings. In such a case we could still credit her with a particular kind of skill, while maintaining that her claim to have identified planets was not validated. In such a case we might also conclude that her skill has turned out to not be terribly relevant to any question of current scientific interest.

3 Two readings of the skill/validation distinction—and their problems

Having drawn an analytical distinction between two aspects of the stabilization of phenomena, let me point out that there are no attempts in the literature to provide an account of stabilization that integrates them. Instead, scholars have focused either on the one or the other aspect. The question I want to raise here is whether a philosophically satisfactory account of stabilization can afford to neglect one of these two aspects at the expense of the other, simply treating them as two processes that can be analyzed separately. My thesis is that it cannot (i.e., the answer is “no”), and that an analysis of stabilization is needed that integrates the two aspects. My argument for this thesis comes in two parts, one positive and one negative. In Sects. 3.1 and 3.2 my focus will

be on the negative arguments. I proceed by discussing—and rejecting—two possible rationales for *not* integrating the two aspects of stabilization outlined above.

In the process of discussing and criticizing these rationales, two insights will emerge that form the cornerstones of my positive account of the stabilization of phenomena. The two insights are that if we want to understand the ways in which phenomena are *stabilized*, we need (1) a more fine-grained analysis of the very notion of *phenomenon*, and (b) an appreciation of the central ability that underlies (and thereby integrates) both aspects of stabilization.

3.1 Skills versus validation: different aspects of stabilizing the same phenomena?

Of the two notions of stabilization of phenomena I have delineated, one focuses on the skills that enable scientists to empirically classify phenomena, whereas the other focuses on methodologies that enable them to validate claims about (previously classified) phenomena. One way of justifying separate philosophical treatments of those two issues might be to say that while each of those two aspects is involved in the stabilization of any given phenomenon, the distinction between them can be construed along the lines of that between discovery and justification, understood in a particular way, i.e., as a distinction between (a) the skills required for empirical work, and (b) the ways in which the results can be justified (or the justification can be rationally reconstructed). Of course, this is only one of several readings of the dichotomy between discovery/justification, and it has not remained uncontested (e.g., [Hoyningen-Huene 1987](#)). Nonetheless, there is a tendency within philosophy of science to assume such a dividing line. Accordingly, the question of the craft aspect of experimentation is often (implicitly or explicitly) relegated to the realm of psychology or sociology of science, whereas philosophers prefer to focus on the rational criteria that can be appealed to in validating claims about the robustness of phenomena. Of course, I do not mean to suggest that philosophers of science who have analyzed epistemic strategies used in experimental science accept the distinction between discovery and justification as just described. My point, rather, is that they might appeal to something like this distinction when explaining their neglect to provide philosophical analyses of the first type of stabilization in favor of the second type of stabilization. The underlying rationale would be that philosophers and sociologists/psychologists of science ask different kinds of questions about the stabilization of phenomena, and the results of the two complement each other rather than needing to be integrated.

While this reasoning has a certain plausibility, it is worth taking a closer look at what exactly it claims. One possible interpretation, which I shall refer to as the “genetic” reading of the skill/validation distinction, holds that the skill- and validation aspects of stabilization reflect the actual order in which research proceeds, such that scientific research would first converge on a particular classification of a phenomenon and then proceed to show that the classification is robust. But this reading is suspect in several ways. Firstly, even if research really did proceed in this sequence, it would be hard to align it with the distinction between discovery and justification. That is to say, it would be rather implausible to claim that the skillful design and execution of empirical work involves no elements of rational justification, and—conversely—it is not so clear that the application of validation strategies cannot constitute part of the discovery

process. Hence, an argument for a genetic reading of the skill/validation distinction that appeals to the distinction between discovery and justification is questionable. Secondly, the very premise of the genetic reading of the skill/validation distinction, i.e., that stabilization occurs in two consecutive stages, may be doubted. While it is perhaps the case that the skillful design and execution of empirical work about phenomena does not necessarily aim at the validation of claims about those phenomena, the converse is surely not true: any and all attempts to empirically validate claims about the stability of a given concept are going to draw on precisely the skills that—on the genetic reading of the skill/validation distinction—are supposed to precede it. It is impossible to empirically investigate of the question of whether a given phenomenon is robust without empirically identifying instances of it. This means that at least the validation aspect of stabilization is not temporally distinct from the skill aspect of stabilization.

Now, this leaves open a second reading of the skill/validation distinction. I shall refer to it as the “systematic” reading. According to it, the practice of validation may well coincide with the skills required to engage in it, but it does not follow that they are not logically distinct.² Therefore, it might be argued that a philosophical treatment of validation can take the skill aspect of stabilization for granted, leaving it to others to give a theoretical account of it. I would like to offer two critical responses to this argument, one weak and one strong. The weak argument grants that it is possible to give a satisfactory account of *validation* that exclusively focuses on reasoning strategies employed to demonstrate the robustness and stability of a previously classified phenomenon. But if our aim is to give an account of *stabilization*, we cannot restrict our attention to a purely logical evaluation of reasoning strategies, since stabilization is a process in time, which involves two aspects (skill and validation). An account of stabilization, therefore, needs to say something about how they work together. While I embrace this argument, I would like to go further and put forward a stronger thesis, which states that a distinct treatment of the skill and validation aspects of stabilization not only fails to provide an adequate analysis of stabilization, but also of *validation*.³

To provide some background for this thesis, let us return to the question of what precisely constitutes the “skill aspect” of stabilization. Above I have characterized it as the ability to empirically identify and classify phenomena, where this ability was further differentiated into (a) an element of physical craftsmanship (being able to make an experiment or instrument work) and (b) an element of cognitive judgment (being able to recognize that an experiment or instrument in fact works). I argue that this latter, *conceptual*, skill should take center stage in an account of stabilization that integrates notions of skill and validation, since it is not only required in the process of identifying potentially interesting empirical phenomena, but it also has to be drawn on in attempts to validate claims about those phenomena. As I will argue more fully

² Indeed, I made a similar point above when I argued that the two aspects of stabilization are analytically distinct.

³ I am here talking about the problem of providing a descriptively adequate account of validation. This is to be distinguished from the normative question of what it would take to actually validate a given finding. My overall focus in this article is on descriptive questions, but we will return to the normative question in the last section of this paper.

in Sects. 4 and 5 below, however, it is not merely the case that conceptual skills are required to even start inquiring about the validity of specific claims about phenomena. The reverse is also true: the results of validational procedures will have an effect on our judgment of what are relevant conceptual skills. Hence, I suggest that the stabilization of phenomena, on the account provided here, has to be viewed as a process of *mutual adjustment* between notions of what are stable phenomena and notions of what are relevant conceptual skills.

3.2 Skills versus validation: stabilizing different kinds of phenomena?

There is a second way in which one might attempt to construe the distinction between the two notions of stabilization outlined at the outset, namely as applying to two *different types of phenomena*, where the first notion of phenomenon is equated with empirical data patterns that are either found in the world or created in the lab, whereas the second notion of phenomenon is more removed from particular regularities, i.e., is, in some sense, *unobservable*. To avoid the problematic language of “observable” and “unobservable” phenomena”, I will refer to these two types of phenomena as “surface phenomena” and “hidden phenomena”. To stabilize the first type of phenomenon, then, would mean to identify some empirical data pattern and to provide evidence for its robustness. To stabilize the second type of phenomenon would mean to provide evidence for, and validate claims about, some regularity that is more removed from empirical data, though the evidence would presumably be provided by means of data.

It might be suggested that data patterns, especially when they are experimentally produced, are not really phenomena in Bogen and Woodward’s sense. I believe that this suggestion is wrong, and I would like to argue for this by way of a discussion of Ian Hacking’s notion of a *phenomenon*. According to Hacking: “To experiment is to create, produce, refine and stabilize phenomena” (Hacking 1983, p. 239). Phenomena, on this reading, are empirical regularities that can be (though are not necessarily) created in the lab. Hacking specifically talks about experimental effects, such as the Hall effect. In contrast, Bogen and Woodward downplay the importance of experimental production, suggesting instead that phenomena are general features in the world, as opposed to the idiosyncratic features of data production in experiments. This prompts them to argue that “[the] features which Hacking ascribes to phenomena are more characteristic of data” (Bogen and Woodward 1988, p. 306, footnote 6). Should we then conclude that when Hacking uses the term “phenomenon” he is really talking about *data* in Bogen and Woodward’s sense, since he is specifically interested in effects that are created under the idiosyncratic conditions of a given experiment? There are two reasons why I want to resist this conclusion. Firstly, some of the examples of phenomena presented by Bogen and Woodward are very similar to the kinds of experimental effects reported by Hacking (e.g., the chunking effect, to which I will return in Sect. 5 below). More importantly, Hacking’s own definition of the term “phenomena” as “noteworthy discernible regularities” (Hacking 1983, p. 225) suggests that he does not think (or at any rate *should* not think) that any one experiment literally creates a phenomenon, since the data that constitute the

outcome of any one individual experiment cannot display a regularity. Regularities, on my understanding, are events that occur regularly. Experiments create data. Data may be classified as instantiating patterns of what I have called surface phenomena.

This way of paraphrasing Hacking's notion of *phenomenon* has two interesting consequences. First, it implies the recognition that phenomena—even in the sense of “mere” empirical regularities—transcend the idiosyncratic features of any one experiment or data point. If we talk about the results two different experiments in which the Hall effect was demonstrated, there is a sense in which we are talking about something that is very idiosyncratic to the specific conditions of the two experiments, respectively. But of course we can only compare them if we assume that both results are instances of the same empirical regularity. In other words, empirical regularities meet Bogen and Woodward's criterion of phenomenonhood. Second, if individual experimental data can at best instantiate (rather than constitute) phenomena, the question arises, what is required to classify data as instantiating a particular phenomenon? It should be clear that we are here touching on an issue that is closely related to a topic that was discussed in the previous subsection, namely, the question of the conceptual skills required to identify a phenomenon. We will return to this in the next section.

Having argued that surface regularities and hidden regularities both qualify as phenomena, we can now return to the main question of this section, i.e., whether the skill and validation aspects of the *stabilization* might be two modes of stabilization that apply to these two different kinds of phenomena, respectively. This suggestion is implausible in several ways. First, it suggests that claims about surface regularities do not need to be validated, since they are delivered to us as self-evident by means of our conceptual skills. Second, it makes the assumption that the validation of hidden regularities does not (or at least not to any significant extent) require conceptual skills. I would like to question both assumptions. The first assumption implies (a) that surface regularities are directly observable, and (b) that these observations are veridical. But, of course direct observability is precisely what we denied by granting surface regularities (data patterns) the status of *phenomena* in the sense of Bogen and Woodward. Furthermore, the mere fact that a given empirical result is classified as instantiating a surface regularity does not mean that there really is a stable regularity. In other words, claims about the robustness of data patterns need to be validated, just as claims about hidden regularities need to be validated. Now, with respect to the second assumption, it is indeed hard to see at first glance how conceptual skills might be involved in the empirical identification and validation of hidden phenomena, since the latter are, after all, more removed from the realm of the empirical than are the surface phenomena. If hidden phenomena are inferred from data, as suggested by Bogen and Woodward, then conceptual skills do not seem to enter. However, we need to be careful here: while it is correct that the raw material of inferences from data to phenomena are individual data points, I find it less obvious that individual data points, as such, qualify as an inductive basis for such inferences. They have to be classified in a particular way,

which in turn will determine the kinds of inferences that are made from them.⁴ I would like to suggest that individual data points are treated as providing an inductive basis for inferences from data to phenomena *insofar* as they are classified as instantiating a particular surface regularity. As I argued above, the identification and validation of surface regularities requires conceptual skills. Hence, the validation of claims about hidden regularities indirectly also requires conceptual skills, because it relies on the identification of surface regularities. If these arguments are correct, then the skill and the validation aspect of stabilization each play a role in the stabilization of both types of phenomena delineated above: surface phenomena and hidden phenomena. Thus, they are not different modes of stabilization that are specific to different kinds of phenomena.

4 Towards a positive account of the stabilization of phenomena

We have come to an end of our discussion of whether there are any defensible reasons for not integrating aspects of skill and validation in an account of stabilization. I conclude that the two reasons discussed in Sects. 3.1 and 3.2 do not hold up. It is conceivable that there might be other reasons, but for now I will shift the burden of proof to those who are still inclined to disagree with my thesis.

Of course, arguing that an analysis of stabilization requires an integrated account of the ways in which skills and validation processes work together is one thing. Providing such an account is quite another. I do not pretend that I have yet done so. However, in the course of the previous section, some insights about the nature of phenomena and of stabilization have emerged that we can now draw on in an effort to construct such a positive account. The first important insight concerns our analysis of the very notion of *phenomenon*. I agree with Bogen and Woodward's general characterization of phenomena as stable features of the world, distinct from the empirical data that are used to test claims about them. However, I argue that if we are interested in the processes of *stabilization* (i.e., of the ways in which scientists come to identify phenomena and validate claims about them by empirical means), a more fine-grained analysis of the notion of phenomena and their epistemic relationship to data is required. Cornerstones of such a more fine-grained analysis will be (a) my distinction between surface phenomena and hidden phenomena, and (b) my thesis that data can only play the role of evidence for phenomena by virtue of being classified as either instantiating a surface phenomenon, or indicating a hidden phenomenon. In turn, data can only be treated as indicating a hidden phenomenon insofar as they are also viewed as instantiating a surface phenomenon. (To put this differently, scientists will not be inclined to treat an individual data point as indicative of a hidden phenomenon as long as they do not

⁴ I take it to be the central insight of Goodman's (1955) new riddle of induction to have pointed out that there is nothing in the raw data that privileges one classification over another. This, then, gives rise to the descriptive question of how we arrive at classifications and inferences that seem "natural", as well as to the normative question of whether we have any reason to believe that what appears natural is in fact correct. As indicated in the previous footnote, the bulk of this paper is concerned with descriptive question here, though some implications for normative questions will be discussed below.

think that the data point is at least in principle replicable, and can therefore be treated as instantiating a data pattern or surface regularity.)

My emphasis on the importance of *treating data as instantiating or indicating phenomena* once again draws our attention to the conceptual prerequisites that must be in place if we want to even begin to inquire about phenomena and their evidential relations to data. In previous sections I have referred to this ability to identify surface phenomena as a *conceptual skill* to distinguish it from the more “crafty” ability to physically create data that instantiate or indicate phenomena. The term “conceptual skill” will do if our aim is to indicate a specific kind of ability, though the ability itself remains rather mysterious. It is something like this ability that Thomas Kuhn attributed to the narrower of his two notions of *paradigm* (Kuhn 1962), later attempting to provide a semi-naturalistic account of it (Kuhn 1974). It is not my ambition here to prove an account of the conceptual skill itself. My interest lies rather in analyzing the contribution made by this skill to the process of attempting to stabilize a phenomenon. One interesting feature of this skill is that on the one hand it seems to be possible to attribute a conceptual skill to someone even if they are mistaken about the nature of the objects they are able to identify empirically. For an example, recall the fictitious scenario of what would have happened if it had turned out that the perceptual objects Caroline Herschel was so good at identifying were not planets, but some systematic artifact of her instrument. I maintain that we would still attribute some kind of conceptual skill to her, whereas we would not do so if her classifications had been entirely random. On the other hand, it is clear that what counts as a *scientifically relevant* conceptual skill will depend heavily on the current state of knowledge at a given time. Staying with the fictitious example: assuming that at a later point in time we had a better understanding of what planets are, we would then also view a different set of skills as relevant to the task of empirically classifying planets.

The reason why I am emphasizing the close connection between knowledge and what counts as a relevant conceptual skill is that it seems to me to point to an important dynamic: On the one hand (as I suggested in Sect. 3.1 above) we cannot even begin to empirically investigate claims about the robustness and stability of phenomena (say, the orbits of planets) unless we assume that we have a way of identifying the objects/phenomena in question (say by means of a specific instrument that provides us with data that we treat as instantiating surface phenomena, which we in turn treat as indicating the phenomena in question). Differently put, our ontology of surface phenomena will inform our ontology of (presumed) hidden phenomena. On the other hand, our attempts to validate claims about hidden phenomena may result in the finding that the surface phenomena we used as indicators of the hidden phenomena were in fact quite inadequate for the purpose. Thus, our ontology of hidden regularities may well alter our ontology of surface phenomena.

If this analysis is correct, what follows from it for the question of this essay (“What exactly is stabilized when phenomena are stabilized?”)? The answer—in a nutshell—is that neither surface phenomena nor hidden phenomena can be stabilized in isolation. They stabilize each other. In other words, what is stabilized

when phenomena are stabilized is the fit between surface phenomena and hidden phenomena.

5 Memory phenomena: surface and hidden

I would like to claim some generality for the account that I have developed so far. However, even if something like the distinction between surface phenomena and hidden phenomena, and the epistemic relationship that I have claimed holds between them, are general features of empirical research, the details of how this relationship plays out in the process of stabilization will be quite different for different fields of research. Hence, the following case study will hardly constitute conclusive evidence for my analysis, but is intended to (a) point to the relevance of the issues addressed by my account, and (b) make my account more tangible by linking it to a specific example. Memory research provides us with a nice example of the distinction between surface phenomena and hidden phenomena, since in the literature we find references to phenomena like encoding and long-term memory, but also to phenomena like chunking effects or ceiling effects, where the latter are obviously very closely tied to particular experimentally generated data patterns, whereas the former are not.

In the following, I will argue that the identity conditions of a surface regularity like a memory effect cannot be established conclusively without drawing on additional assumptions about hidden memory phenomena that are causally responsible for instances of the memory effects. This will raise the question of how the identity conditions of hidden memory phenomena are established.

5.1 Chunking, recency, priming, and other memory effects

As already mentioned, Bogen and Woodward list among their examples of phenomena some findings from experimental psychology, commonly called “effects”, such as the chunking effect, and the recency effect. Other effects one might wish to add are the primacy effect, the ceiling effect, and the priming effect. Our focus here will be on the chunking effect. This expression refers to the fact that in memory experiments human subjects can recall more items when they are instructed to “chunk” them than without such an instruction. For example, it is easier to remember a phone number when it is chunked, then to remember every single digit individually (6-6-1-8-4-0-0 vs. 661-8-400). There is an intuitive sense in which this is an empirical regularity. Moreover, it is one that is thought to be well-confirmed. But what would it take to show that it is a genuine, robust phenomenon? Presumably one would need to show that the effect can be replicated. But here we face a problem: On the one hand it seems obvious that the conditions of the experiment need to be varied in order to rule out that the data in question are specific to one very specific experimental set-up, thereby demonstrating the robustness of the phenomenon. For example, one has to show that instances of the chunking effect hold for one subject in different situations, that it holds for different subjects, that it holds with different materials (different digits, words instead of digits, nonsense-syllables instead of digits, etc...), and that it holds regardless of the type of memory test that is used. But on the other hand, every variation of the conditions of the

experiment raises the question of whether the experiment in question is really an apt instrument for triggering instantiations of the chunking effect. The question is on what grounds a given experimental result can be classified as instantiating one phenomenon, as opposed to some other, entirely different phenomenon, or as not instantiating an interesting phenomenon at all.

A version of this question was asked by Collins (1985) when he raised the issue of how one tells the difference between a replication of an experiment with a negative result, and a failure to replicate the experiment. While we do not have to agree with his account of how this issue is resolved (i.e., that there are no identity criteria that can be explicated and that the question is ultimately resolved by way of power relations), he puts his finger on an important question, namely by what criteria data of different experiments can be classified as instantiating the same empirical regularity. For example, how might we decide that the results of chunking experiments, when replicated with words, are really instances of the chunking effect, rather than, say of a depth-of-processing effect, or an effect of some other brain process? Or even more generally, on what grounds do we classify the outcomes of chunking experiments as memory phenomena rather than, say, attention phenomena? On what grounds do we decide that a given test used to explore the nature of chunking is even a memory test?

One possibility would be to say that some psychologists just happen to have the conceptual skill to hone in on the kinds of experimental data that happen to instantiate the chunking effect. But this is not very satisfactory, since it does not provide us with any criteria for distinguishing between people who have this skill and people who do not. Nor does it tell us how it might be determined that the conceptual skill in question is one of any particular scientific relevance. The questions I am raising here are (a) how the identity conditions of surface phenomena like the chunking effect can be established, and (b) how claims about them can be validated? I argue that neither question has an answer if we stay strictly on the level of surface phenomena. That is, our classifications of surface phenomena, no matter how self evident they may appear, will not be able to shake an air of arbitrariness. My thesis, then, is that even though surface regularities like the memory effects in psychology can be objects of study in their own right, attempts to stabilize them will ultimately turn to something other than surface regularities. More precisely, in their search for criteria that determine the identity conditions of surface regularities, scientists assume that such identity conditions are determined by hidden phenomena that stand in a relevant explanatory relationship to the data that instantiate the surface regularities. In the case at hand: if the chunking effect is a memory phenomenon, its identity conditions are presumably determined by a brain process that is associated with memory (i.e., with a hidden memory phenomenon), where that process is causally responsible for all the data that instantiate the chunking effect.

5.2 Memory phenomena as (hidden) physical regularities

But of course the thesis of the previous section (that the identity conditions of surface phenomena are determined by hidden regularities) immediately raises the question of how the identity conditions of hidden regularities are determined. In this case, if we

assume—as was done for many decades within the memory literature—that the chunking effect is to be explained by reference to the phenomenon of short term memory (or by some cognitive processes that takes place within short term memory) then the issue is how the identity conditions of this phenomenon might be established.

In the philosophical literature this kind of question has been addressed by proponents of experimentalism, who have argued that it is possible to validate claims about what I here call hidden phenomena by drawing on methodological strategies that will determine whether a given phenomenon is robust. Such strategies include controlling for known factors that might interfere between phenomenon and data, etc. (Franklin 1993; see also Woodward 1989). These methodological strategies will then allow for the generation of data, which in turn will provide evidence for the robustness and stability of the phenomenon in question. With respect to the example, let us assume that we tentatively explain the chunking effect by reference to a chunking mechanism that is located in working memory. Evidence for the existence of such a chunking mechanism might be provided by controlling for other known mental operations, by testing for correlations and/or dissociations between different ways of testing for chunking, and by causally manipulating the brain area where the chunking mechanism is thought to be located. These all appear to be promising ways to go. The only trepidation one might have is that the data used to validate claims about the hidden phenomenon in question are not neutral. It seems that if the empirical data we use are to have any relevance at all to the hidden chunking phenomenon, they would have to be data from chunking experiments. But if the data produced in chunking experiments are supposed to instantiate the chunking effect, which in turn is supposed to be causally related to the chunking mechanism, this appears to beg the question, since the very issue of what are the identity conditions of this purported empirical regularity was supposed to be settled by reference to its cause, i.e., some yet to be determined hidden regularity. In other words: knowledge of the identity conditions of the mental operation of chunking (the hidden regularity) was supposed to help us delineate the identity conditions of empirical chunking phenomena (the surface regularity). But how can this be achieved if the only way we can gain access to the presumed mental operation of chunking is by way of the already existing criteria of delineating the surface regularity?

What I am pointing to here may at first glance sound a lot like the type of argument that has come to be known by names such as the “experimenter’s regress”. However, is the threat of circularity real? I do not think so. It is quite conceivable that we use a specific chunking experiment under various different conditions, e.g., with subjects that have specific forms of brain damage with known functional impairments; or under conditions where subjects are given other cognitive tasks to see how they affect their performance on the chunking task. Consider for example the scenario where an attention-demanding task greatly diminishes the subject’s chunking ability. This might leave scientists to suspect that chunking is really a phenomenon of attention, rather than memory.⁵ Or consider the scenario where patients with known memory impairments do well on some chunking tasks, but not on others. This might lead scientists to suspect that what was previously classified as one surface regularity is in fact a

⁵ This is in fact entertained as a life possibility in recent psychological publications (Jonides et al. 2008).

much more complex phenomenon, in that what were previously thought to be different indicators of the same hidden regularity are in fact indicators of different hidden regularities. These are both examples of how the surface regularity is instrumental in an investigation of a hidden regularity, which in turn may change the ways in which the surface regularity is classified: in the former case, the outcome might be that the chunking effect is no longer classified as a memory phenomenon. In the latter case, the outcome might be that we end up distinguishing between different types of chunking effects.

The point of this example is to provide some plausibility to my thesis that neither surface regularities nor hidden regularities can be stabilized individually. Insofar as anything is stabilized here, what is stabilized are not claims about the robustness of specific regularities, but rather claims about the fit between surface regularities and hidden regularities.

6 Conclusion

In this article I have proposed an account of the stabilization of phenomena that integrates two aspects that have previously not been discussed systematically in relation to one another: skill and validation. Central to my analysis was the insight that what I called “conceptual skills”, i.e., the ability to classify data as instantiating or indicating phenomena of interest, are a vital prerequisite for the very possibility of empirical research about the phenomena in question. This kind of skill, then, is a necessary condition for the possibility of validating claims about phenomena. In turn, however, skills do not embody some kind of mysterious and eternal wisdom. Rather, what counts as a skill, and moreover, what counts as a relevant skill, can change as a result of changing ideas about what kinds of phenomena are in fact stable and robust. Hence, I argued, the skill and validation aspects of stabilization are strongly interconnected. Turning then to an analysis of the process of stabilizing phenomena, I presented a fine-tuned version of Bogen and Woodward’s notion of phenomenon, one that distinguishes between what I called “surface regularities” and “hidden regularities.” Both, I suggested, are phenomena in Bogen and Woodward’s sense, and claims about either kind of phenomenon calls for validation. I argued that neither kind of phenomenon can be stabilized and validated on its own. Ultimately, if anything is stabilized, then it will be the fit between those two kinds of phenomena.

The main aim of my analysis has been to give a descriptively adequate account of how the stabilization process proceeds. I thereby responded to what I argued are descriptively inadequate or one-sided treatments of stabilization in the literature. However, given my treatment of processes of validation as constituting one aspect of stabilization, the question still remains whether there is a normative conclusion to be drawn from my analysis. In other words, can we say something about what it takes for a claim about a phenomenon to be validated? In response to this question we need to be clear on what exactly we want to assert when we say that a claim about the existence of a phenomenon has been validated. Some of the authors who have written about robustness have viewed their work as potentially being able to show that there can still be a realm of secure knowledge, even if it is no longer thought to be delivered

by neutral observations. The basic idea, that we find, for example in Hacking's entity realism, is that knowledge about certain robust phenomena can prevail, even if our theories about the phenomena change. We therefore do not have to make a choice between being realists or antirealists about theories. We can simply be realists about the stability of phenomena. By severing the tie between theory and observation and emphasizing the partial autonomy of research on phenomena, Bogen and Woodward, at times also lend themselves to this reading, though it is not clear whether this is in fact what they intended.

Given my own analysis of stabilization, it should be clear that I do not think that any such attempts of isolating a small pool of secure knowledge will work. As was suggested by my example of the chunking effect, claims about phenomena require for their stabilization and validation assumptions about other phenomena and about the ways in which they are causally and systematically related. Hence, my conclusion about the validation of claims about robustness of phenomena is mainly negative, in that I reject philosophical attempts to show that individual phenomena can be validated in isolation. A more positive account might try to argue that once we have a large number of claims about interrelated phenomena, the very fact that they fit so well together might in itself constitute an argument for their validity (see [Hacking 1992](#) for a version of this argument). A discussion of this suggestion will have to be deferred to another paper.

References

- Bogen, J., & Woodward, J. (1988). Saving the phenomena. *The Philosophical Review*, *XCVII*(3), 303–352.
- Collins, H. M. (1985). *Changing order: Replication and induction in scientific practice*. London: Sage Publications.
- Culp, S. (1994). Defending robustness: The bacterial mesosome as a test case. *PSA 1994*, *1*, 46–57.
- Culp, S. (1995). Objectivity in experimental inquiry: Breaking data-technique circles. *Philosophy of Science*, *62*, 430–450.
- Franklin, A. (1993). The epistemology of experiment. In D. Gooding, T. Pinch, & S. Schaffer (Eds.), *The uses of experiment* (pp. 437–460). Cambridge: Cambridge University Press.
- Goodman, N. (1955). *Fact, fiction, and forecast*. Cambridge: Harvard University Press.
- Hacking, I. (1981). Do we see through a microscope? *Pacific Philosophical Quarterly*, *62*, 305–322.
- Hacking, I. (1983). *Representing and intervening. Introductory topics in philosophy of science*. Cambridge: Cambridge University Press.
- Hacking, I. (1992). The self-vindication of the laboratory sciences. In A. Pickering (Ed.), *Science as practice and culture* (pp. 29–64). Chicago: University of Chicago Press.
- Hoyningen-Huene, P. (1987). Context of discovery and context of justification. *Studies in History and Philosophy of Science*, *18*, 501–515.
- Jonides, J., Lewis, R. L., Nee, D. E., Lustig, C. A., Berman, M., & Moore, K. S. (2008). The mind and brain of short-term memory. *Annual Review of Psychology*, *59*, 193–224. Retrieved February, 2008 from <http://psych.annualreviews.org>.
- Kuhn, T. (1962). *The structure of scientific revolutions*. Chicago, IL: University of Chicago Press.
- Kuhn, T. (1974). Second thoughts on paradigms. In T. Kuhn (Ed.), *The essential tension. Selected studies in scientific tradition and change*. (pp. 293–319). Chicago, IL: University of Chicago Press.
- Polanyi, M. (1966). *The tacit dimension*. Gloucester, Mass.: P. Smith.
- Wimsatt, W. (1981). Robustness, reliability, and overdetermination. In R. Brewer & B. Collins (Eds.), *Scientific inquiry and the social sciences* (pp. 124–163). San Francisco: Jossey-Bass.
- Woodward, J. (1989). Data and phenomena. *Synthese*, *79*, 393–472.