

The Multiple Dimensions of Multiple Determination

Abstract *Multiple determination* is the epistemic strategy of establishing the same result by means of multiple, independent procedures. It is an important strategy praised by both philosophers of science and practicing scientists. Despite the heavy appeal to multiple determination, little analysis has been provided regarding the specific grounds upon which its epistemic virtues rest. In this article I distinguish between the various dimensions of multiple determination and show how they can be used to evaluate the epistemic force of the strategy in particular cases. Distinguishing between the various dimensions helps to resolve the disagreements regarding the relevance and epistemic import of multiple determination for scientific research. It also suggests a fruitful mode of interaction between the philosophy of science and historical accounts of scientific practice.

1. Introduction

Multiple determination is the epistemic strategy of establishing the same result by means of independent determination procedures.¹ It is widely regarded as one of the most important strategies that scientists use to establish the validity of their results (Wimsatt 1981; Hacking 1981, 1983; Cartwright 1983, 1991; Franklin 1986, 1998; Bechtel 1990, 2002; Culp 1994, 1995; Burian 1997; Chalmers 2003; Nederbragt 2003; Weber 2005; Woodward 2006; Soler et al. 2012;

¹ Although I will be using the terms ‘determination procedure’ and ‘experimental procedure’ interchangeably, the first term should be interpreted in a wide sense to include different practices: experimental procedures, observational procedures, reasoning from empirical facts and, in general, any procedure that can be used to establish the validity of a result. The term ‘result’ should also be interpreted in a wide sense. It usually refers to a claim about the world, but it can also refer to something non-translatable into an explicit claim such as instrument readings, complex visual displays and images, and raw data; it can also refer to claims about determination procedures.

Schickore and Coko 2013; Kuorikoski and Marchionni 2016).² In general, it is claimed that determining a result with multiple procedures is better than doing the same with a single procedure. Despite the heavy appeal to multiple determination, little analysis has been provided regarding the specific grounds upon which its epistemic virtues rest.³ Traditionally, the rationale underlying the epistemic import of the strategy has been presented to be a no-coincidence argument. Namely, that it would be an improbable coincidence for multiple independent procedures to determine the same result and yet for the result to be incorrect or an artifact of the determination procedures. One (contrived) example that is often used to illustrate this *blunt*

² Multiple determination is not the only epistemic strategy for establishing the validity of results.

Experimentalists seem to have a rather large repertoire of strategies for establishing the reliability of experimental results and procedures (Franklin 1986; 2002).

³ The epistemic import of *variety-of-evidence* arguments has been discussed from a formal, Bayesian perspective (Bovens and Hartmann 2003; Claveau 2013; Stegenga and Menon 2017, Claveau and Grenier 2019). Although somewhat clarificatory with respect to our normative intuitions, the Bayesian approach is not reflective of real scientific practices. Therefore, it is not helpful for understanding the epistemic role of multiple determination in actual scientific research. Thus, most of the relevant Bayesian accounts aim to quantify the degree of confirmation of a general theoretical hypothesis (H), by diverse bodies of evidence (E_1, E_2), without devoting much attention to the questions of ‘what constitutes a body of evidence?’ and ‘how is the validity of a body of evidence established in scientific practice?’. This paper focuses on the latter question. In addition, according to the Bayesian approach there is an algorithmic way to decide the probability of H , given E_1, E_2 , which is shared by all the members of the scientific community. The account presented here, however, shows that this is not the case in real, ongoing scientific research. Finally, most of the Bayesian accounts focus on the various ways that evidence is independent and neglect the importance of other aspects. As this paper aims to show, however, any account of multiple determination that disregards the contribution of the different dimensions of the strategy upon its epistemic force is necessarily incomplete.

rationale is that of independent witnesses: if several witnesses testify that a certain event occurred, and we can be certain that the witnesses' testimonies are independent—that is, if we can be certain that the witnesses did not base their testimonies on one another, that they were not coached, that they were not bribed by some villain, and so on—we can safely conclude that the event did occur. It would be an improbable coincidence for multiple witnesses to have fabricated the same story independently of one another.

Although intuitively correct, the traditional rationale is too coarse-grained. It is not helpful for understanding how the multiple determination strategy is supposed to work in scientific research. For example, the blunt rationale cannot tell us if a result that is determined by several independent procedures should be regarded as more credible than a result that is determined by a single procedure; especially in the case when a result that is determined by a very reliable procedure—the latter in the sense of being based on well tested, well understood, and widely accepted principles—is compared with a result that is determined by multiple procedures that are not considered to be reliable (Hudson 1999, 2014; Stegenga 2009, 2012). In addition, relatively recently, there have emerged many case studies that investigate the employment of multiple determination in scientific practice. The general conclusion from these case studies is that scientific practice is far more complex and messier than implied by the blunt rationale.

In the literature, one can distinguish two main reactions to the inability of the blunt rationale to accommodate the complexity of scientific practice. The first doubts the epistemic value and relevance of multiple determination for scientific research (Hudson 1999, 2003, 2014; Stegenga 2009, 2012; Soler 2012a; Boon 2012). The second accepts that multiple determination is an important strategy for scientific research but rejects that its underlying rationale is a no-coincidence argument (Staley 2004; Kuorikoski and Marchionni 2016; Basso 2017; Schupbach 2018).

In this article I propose a framework for understanding the use, structure, and epistemic import of multiple determination in scientific practice.⁴ On the one hand, this framework accepts that the blunt rationale is not an adequate description of scientific research. On the other hand, however, it recognizes that the blunt rationale captures something essential about the multiple determination strategy. Namely, that its epistemic import is indeed based on a no-coincidence argument. Not all cases of multiple determination carry the same epistemic force, however. The important thing, therefore, is to be able to assess the epistemic force of concrete applications of

⁴ The limited space does not allow to argue extensively against the alternative accounts of multiple determination here. Furthermore, whether multiple determination is an important strategy in scientific research and whether its underlying rationale is a no-coincidence argument are empirical questions, which should be answered by ongoing research: by studying a large variety of cases of multiple determination and by paying attention to the historical emergence and development of the strategy. Nevertheless, the framework presented here, provides some indirect arguments against the alternative accounts. First, the various examples of multiple determination presented in section 2 show that multiple determination as a no-coincidence argument is not a philosopher's invention, but a strategy employed by scientific practitioners themselves. Second, by solving some of the objections against the epistemic value of multiple determination, the dimensions framework undercuts some of the reasons that motivate the alternative accounts. For example, Kuorikoski and Marchionni (2016) deny that the strategy's rationale is a no-coincidence argument, because the latter, they claim, cannot identify when confidence in a result is justifiably increased. Third, the dimensions framework is able to accommodate a variety of cases of multiple determination, including what have been considered to be paradigmatic cases: Jean Perrin's argument for molecular reality (section 2.2) and Ian Hacking's description of the discovery of 'dense bodies' in red blood cells (section 2.1). Because of their difficulty to accommodate such cases, the alternative accounts present them as misleading and/or not representative of the strategy (Kuorikoski and Marchionni 2016; Basso 2017). By identifying the elements responsible for the epistemic strength of multiple determination, the dimensions framework shows what made these cases so successful.

multiple determination in scientific research. To conduct such an assessment, I distinguish between the various dimensions of multiple determination. These are the elements of the strategy that are responsible for the epistemic force of the no-coincidence argument as it is made in a particular case. Distinguishing among the various dimensions is akin to “looking under the hood” of multiple determination and seeing how the no-coincidence argument emerges from more elementary aspects. For example, one dimension of multiple determination is the number of determinations. The relation between the number of determinations and the strength of the no-coincidence argument seems intuitively clear: the greater is the number of determination procedures that establish the same result, the stronger is the no-coincidence argument. The final evaluation of the epistemic force of the argument, however, needs to consider also the contributions made by the other dimensions. Distinguishing among the various dimensions can help solve most of the existing disagreements regarding the relevance and epistemic import of multiple determination for scientific research.

The paper is structured as follows. In section 2, I present several examples of multiple determination that indicate various ways in which concrete applications of the strategy can differ from the (ideal) epistemic situation implied in the blunt rationale. In section 3, I introduce the dimensions framework. In Section 4, I discuss the dimension of *independence*. I distinguish three kinds of independence: *theoretical*, *genetic*, and *historical*. In sections 5 and 6, I present the *reliability* and *the clarity* of a single determination, respectively. In section 7, I show how the *quality of the convergence of independent determinations* affects the no-coincidence argument. I distinguish between *convergence of raw data* and *convergence of theoretically interpreted results*. In section 8, I discuss the “*plasticity*” of *inferences and determination procedures*. In section 9, I discuss the complexity of the result determined by independent procedures. In section 10, I show how the *existence of discordant results* affects the no-coincidence argument. I argue that discordance is not a “hard-problem” for multiple determination. In section 11, I discuss the relation between the *number of determinations* and the epistemic force of the multiple

determination strategy. I show how the dimensions framework answers the question(s) of whether (and why) is it better to establish a result with multiple procedures. In section 12, I conclude by showing how the framework proposed here paves the way for a mutually beneficial relationship between the philosophy of science and (historical) accounts of scientific practice.

2. Examples of multiple determination

2.1 The multiple determination of “dense bodies” in red blood cells

A paradigmatic example of multiple determination was presented in Ian Hacking’s (1981) essay “Do we see through a microscope?”, which became the eleventh chapter in his (1983) influential book *Representing and Intervening*. Hacking’s account is inspired by the experiments that established the existence of dense bodies in red blood cells. In these experiments low powered electron microscopy had revealed the existence of “small round structures” in red blood cells. These structures were called “dense bodies” (a term which simply denoted the fact that these structures were relatively impermeable to the electron beam of the microscope). Based on the movement and density of these bodies during the various stages of cell development or disease it was hypothesized that they were real structures in the cell. There was, however, still the possibility that they might be artifacts of the electron microscope. One test was considered crucial for deciding the question: can one see the same structures by using a different experimental technique? Indeed, the dense bodies were revealed to exist by fluorescent staining and subsequent observation with a fluorescent microscope. The situation gave rise to a no-coincidence argument for the existence (non-artificiality) of the dense bodies.⁵

⁵ “Two physical processes—electron transmission and fluorescent re-emission—are used to detect the bodies. These processes have virtually nothing in common between them. They are essentially *unrelated chunks of physics*. It would be a preposterous coincidence if, time and again, two completely physical processes

Contrary to what is usually assumed in the literature, it was not simply the fact that visual images obtained with electron and fluorescence microscopy contained small round dots which gave rise to the no-coincidence argument. Although implausible, it is not highly improbable for visual images obtained with independent microscopy techniques to contain small round dots that are, nevertheless, artefacts commonly produced by the two methods (see the mesosomes example in section 2.3 below). Furthermore, the visual displays produced with electron and fluorescence microscopy are not literally the same. An important element that strengthened the force of the no-coincidence argument was the identical local arrangements of the small round dots (dense bodies) in the visual displays produced by the two experimental techniques—that is, the fact that the small round dots appeared in the same positions in both visual displays. This point is important, because it concerns an often-neglected element of multiple determination: the complexity of the result that is established by independent procedures (see section 8).

Hacking is explicit about the epistemic import of the no-coincidence argument: determining the same local arrangement of small round dots with independent microscopy techniques established that they were not artifacts of the techniques, but real structures in the cell. He is, however, also explicit about what the no-coincidence argument does not tell us: it does not tell us what the things out there in the cell are.⁶ We cannot say anything theoretically important about these structures because, according to Hacking, the no-coincidence argument does not make use

produced identical visual configurations which were, however, artefacts of the physical processes rather than real structures in the cell” (Hacking 1983, 201, my emphasis).

⁶ “Note also that no one need have any ideas what the dense bodies *are*. All we know is that there are some structural features of the cell rendered visible by several techniques. Microscopy itself will never tell all about these bodies (if indeed there is anything important to tell). Biochemistry must be called in.” (Hacking 1983, 201).

of a theory of the cell and/or the microscopy techniques. It is an argument based on the multiple determination of raw (i.e., theoretically uninterpreted) data.

2.2 Jean Perrin and the multiple determination of Avogadro's number

Another paradigmatic case of multiple determination is the description of thirteen different procedures for determining Avogadro's number (N)—the number of molecules contained in a gram-mole of a substance under the same conditions of pressure and temperature—by the French physicist Jean Perrin (1909, 1913). The different procedures included Perrin's own three, which were based on the experimental study of the vertical distribution, mean displacement, and mean rotation of Brownian particles, respectively. The multiple determination of N is widely considered to be the scientific development that ended the long nineteenth century debates regarding the existence of unobservable atoms and molecules (Brush 1968; Nye 1972; Chalmers 2009).

The chart in *Fig.1* appeared in the concluding section of Perrin's (1909) influential monograph *Mouvement Brownien et Réalité Moléculaire*. In this monograph, Perrin presented for the first time to a wide readership his experimental work on Brownian movement and its relationship with the existing evidence for the atomic-molecular conception of matter.

MOUVEMENT BROWNIEN ET RÉALITÉ MOLÉCULAIRE. 111		
Phénomènes étudiés.		$N \cdot 10^{-22}$
<i>Viscosité des gaz</i> , en tenant compte	du volume à l'état liquide....	> 45
	du pouvoir diélectrique des gaz.	< 200
	de la loi exacte de compressibilité.	60
Mouvement brownien	Répartition d'une émulsion uniforme...	70,5
	Déplacement moyen en un temps donné.	71,5
	Rotation moyenne en un temps donné..	65
Diffusion des corps dissous.....		40 à 90
Mobilité des ions dans l'eau.....		60 à 150
<i>Éclat du ciel bleu</i>		30 à 150
<i>Mesure directe de la charge atomique.</i>	Gouttelettes condensées sur les ions.	60 à 90
	Ions collés sur fines poussières....	64
<i>Émission de projectiles α.</i>	Charge totale rayonnée.....	62
	Constante de temps du radium...	70,5
	Hélium produit par le radium....	71
Énergie du spectre infra-rouge.....		60 à 80

Fig.1

The chart summarizes the values for N determined from the independent consideration of different phenomena such as viscosity of gases, Brownian movement, diffusion of dissolved

bodies, alpha particle decay, charge of the electron. Despite the slight variations and approximations, the different determinations seemed to agree with one another and, according to Perrin, suggested as the most probable value for N the number $70,5 \cdot 10^{22}$ —which was also the value obtained in his own experimental study of the vertical distribution of Brownian particles. Perrin argued that this numerical agreement gave rise to a no-coincidence argument which left no reasonable doubt about the validity of the atomic-molecular hypothesis.⁷

Elsewhere (Coko 2019; 2020) I examine the historical complexity of Perrin's experimental work and the role that the multiple determination strategy played in the determination of N . Here I only focus on one important difference between Perrin's and Hacking's examples. Perrin's three determinations of N did not produce the same result at the level of raw data. Using complex and theory-dependent experimental procedures, Perrin produced a disparate body of raw data which, based on the auxiliary assumptions underlying the procedures, were used to calculate the number of Brownian particles at a specific level of an emulsion, the average mean displacement of a Brownian particle, and the average mean rotation of a Brownian particle. Each one of the numerical values found had to undergo additional complex stages of processing, refining, and theoretical interpretation before it could be claimed that they established the ("same") value for

⁷ "I believe it impossible that a mind clear of all preconception can reflect on the extreme diversity of the phenomena which thus converge towards the same result, without experiencing a very strong impression, and I think that from now on it will be difficult to defend by rational arguments a hostile attitude towards molecular hypotheses" (Perrin 1909, 111). "One is seized with admiration in front of the miracle of so precise concordances, starting from phenomena so different. That one finds the same value within each method, when varying as much as possible the parameters of the experiment, and that the numbers thus calculated without ambiguity by these diverse methods coincide, gives to molecular reality as much certainty as the principles of Thermodynamics" (Perrin 1912, 249.).

the number of molecules contained in a gram-mole of a substance. Perrin's is an example of multiple determination of (theoretically interpreted) experimental results.

2.3 The case of the bacterial mesosome

Perhaps the most controversial case of multiple determination is the one concerning the rise and fall of the bacterial mesosome. Mesosomes were first observed using electron microscopy in the mid-1950s, and for more than fifteen years they were believed to be real cell structures. This belief was partially justified because of their appearance in visual imaging obtained with different microscopy techniques on specimens prepared with different fixation techniques. Finally, in the early 1970s—after more than one hundred fifty scientific papers detailing their structure and biological function were published—mesosomes were deemed to be artifacts of the fixation techniques.

Scholars disagree about the role that multiple determination played (or did not play) in the rise and then fall of mesosomes (Rasmussen 1993, 2001; Culp 1994; Hudson 1999, 2003; Nederbragt 2003; Weber 2005). I will not focus on these disagreements here.⁸ I will only point out how the mesosomes case differs from the epistemic situation portrayed in Hacking's example on the multiple determination of raw data. In the mesosomes case, the visual displays obtained with independent microscopy techniques on specimens prepared with different fixation techniques, were not literally the same. In addition, the supposed convergence of the raw data was only one part of the argument made for the existence (and then the non-existence) of the mesosomes. Considerations regarding the theory of the source (cell) and the theory of the experimental techniques, were crucial for deciding the existential status of mesosomes. The most

⁸ The framework proposed here can resolve most of the disagreements. There is so much that can be accomplished in a single paper, however. The application of the dimensions framework to the mesosome case has to wait another occasion.

dramatic divergence of this case from the dense bodies example, however, is the fact that the mesosomes turned out to be artifacts, even though it was claimed that they had been detected with independent microscopy techniques.

2.4 Multiple determination of visual images and the structure of scientific papers

Another example indicating how actual scientific practice diverges from the epistemic situation implied in the blunt rationale comes from an analysis of the argumentative structure of a very influential and frequently cited radio-astronomy paper (Allamel-Raffin and Gangloff 2012). The paper in question is titled “The Milky Way in molecular clouds: a new complete CO survey”, was published in 2001, and concerns the distribution of CO in the Milky Way Galaxy. Its distinguishing feature is that, although only ten pages long, it contains twenty visual displays of various kinds (images, maps, diagrams) that aim to visually portray the distribution of CO in our galaxy. The authors of the radio-astronomy paper, similarly to Hacking, use the structural similarities between non-verbal entities such as images and maps constructed with data obtained with different kinds of telescopes (radio, infrared, and optic) to argue for the accuracy of the content of the visual displays and the reliability of the instruments and techniques used to obtain them. We should not mistake this case as one of multiple determination of raw data, however. The structurally similar visual images are not raw data—i.e., the direct observational output of telescopes. They are the result of theoretical processing and synthesis of a large body of visual and non-visual data obtained with different kinds of telescopes, and which are afterwards presented in a visual or chart format. It seems that emphasizing that the same result is established by multiple independent means is an effective and widely used principle of argumentation in scientific papers (Allamel-Raffin and Gangloff 2012, 140).

2.5 DNA sequencing

In the 1970s, different techniques had become available to obtain information about an unobservable phenomenon, the sequence of nucleotides on segments of DNA. One of these

techniques was known as the ‘chemical technique’; another became known as the ‘chain-termination technique’ (Culp 1995). Both techniques produced similar raw data, in the sense of similar ‘visual configurations’: both techniques produced X-ray films showing similar patterns of dark bands. The similarity of raw data, however, was not enough to establish that they were produced by the nucleotides. The theory underlying the functioning of the techniques as well as of the composition of DNA was needed to show that the similar visual configurations referred to nucleotides. A close look at all these theories involved in DNA sequencing revealed, however, that the theories informing the different techniques were only partially independent from one another. Moreover, they were not completely independent of the theory of the object under investigation (Culp 1995, 454).

2.6 Clathrin-mediated endocytosis

A relatively recent study on how researchers in cell biology are trying to establish the correctness of claims regarding clathrin-mediated endocytosis (CME) provides yet another interesting departure of concrete applications of the multiple determination strategy from the ideal epistemic situation (Trizio 2012). Endocytosis is the process by which nutrients in the form of ions and molecules are internalized in the cell. CME is the internalization of nutrients realized when clathrin coated vesicles are formed via the invagination of the cell membrane. Until recently it was believed that CME occurs only when the size of the internalized nutrients does not exceed 150 nm and that it plays no role in the internalization of bigger size nutrients. A study concerning the invasion of mammalian cells by *Listeria* (a bacillus about twenty times in size the accepted limit for CME), however, came in conflict with these entrenched views. The study was conducted using the technique of fluorescence microscopy, which allowed filming the different phases of the invasion of mammalian cells by *Listeria* and suggested that the latter entered the cell in a clathrin mediated way. Further, it was established that it is possible to induce a measurable decrease in bacterial entry by reducing the ability of the cell to produce clathrin. This latter evidence relied on

independent theoretical principles from those of fluorescence microscopy. However, it suggested only indirectly that *Listeria* invade cells in a clathrin-mediated way (Trizio 2012, 109-110).

These findings were received with skepticism by the scientific community. There were two interrelated reasons for this skepticism. The first was the findings' conflict with the entrenched views on endocytosis. The second was that fluorescence microscopy does not provide enough details about the role of clathrin in cases of CME. Despite the (indirect) independent support from biochemistry, it was deemed that the evidence provided was not conclusive enough to justify a revision of the entrenched views. What is interesting in this case is that the scientific community *explicitly* demanded further theoretically independent evidence (in particular, evidence from electron microscopy) for establishing the validity of the controversial findings (Trizio 2012, 110).

Another interesting aspect of this case is the way in which the different independent procedures (fluorescence microscopy, biochemistry, electron microscopy) called to establish the validity of the controversial findings diverge from the ideal epistemic situation. Not only the different techniques yield different outputs at the level of raw data, but they do not even focus on the same aspects of the inquiry. Thus, even when the raw data are theoretically interpreted, they do not provide evidence for the same claim. On the one hand, fluorescence microscopy (due to its limited resolution power) can only reveal the presence of clathrin where and when bacteria invasion takes place but can say little about the role of clathrin in the causal mechanism of endocytosis. On the other hand, biochemical evidence provides information about the causal role of clathrin in this type of endocytosis but fails to give information about its size and shape. Finally, electron microscopy, because of its ability to produce high resolution images, if successfully implemented in this case, would provide information for both the localization and the shape of clathrin structure during this kind of endocytosis (Trizio 2012, 115-16).

Despite the complexity of the ECM case and its divergence from the ideal epistemic situation, it seems that the latter plays a very important role in regulating the researchers work. The ideal epistemic situation plays the role of a “methodological attractor”—i.e., the role of “an

idealized logical scheme guiding the efforts of experimenters and technicians, and at the same time, playing a crucial role in the acceptance of a result by the scientific community” (Trizio 2012, 120). As the ECM case shows, the requirement for multiple independent determinations becomes explicit when a new result challenges an entrenched body of knowledge, and it seems to be proportional to the extent of the revision required.⁹

2.7 Electron microscopy and the discovery of dislocations of molecular lattices in crystals

The early use of electron microscopy to observe dislocations of molecular lattices in crystals, during the 1950s, provides another interesting example of multiple determination (Chalmers 2003). In this case, the evidence for the existence of molecular dislocations obtained through electron microscopy was heavily dependent on the theory underlying the operation of the electron microscope. Even more crucially, the justification of some key data was dependent on some untested and speculative theories about the electron microscope and its interaction with the crystal molecules. Nevertheless, the researchers were able to argue strongly, and simultaneously, not only for the validity of the experimental results obtained (i.e., the existence of dislocations), but also for the novel and speculative theories of electron microscopy that were used to interpret raw imaging data as evidence for molecular dislocations (Chalmers 2003, 493). What made this apparent circularity non-vicious was the fact that a multitude of experimental techniques and arguments were used to support the existence of dislocations in crystals. These included: experimental data dependent on well-established parts of the theory of the electron microscope, experimental data from X-ray diffraction, arguments from the motion of dislocations, and quantum mechanical explanations of the interaction of the electron microscope with the specimen. The rationale used was that it would be an improbable coincidence for the result produced with the electron microscope to match the evidence produced by other (theoretically) independent procedures, and yet for the theory used to interpret that result to be largely incorrect.

⁹ See also Nederbragt (2003)

2.8 The discovery of weak neutral currents

Léna Soler has presented another important way in which the use of multiple determination in scientific research diverges from the ideal epistemic situation (Soler 2012b). Soler's analysis is based on a historical episode known as the 'discovery of weak neutral currents' from the 1970s. Soler argues that although the existence of weak neutral currents is traditionally considered to have been established by independent experimental procedures, the historical situation was different. The result establishing the existence of weak neutral currents was in reality the conclusion of a process of *mutual adjustment* of initially divergent results obtained by different experimental groups using different experimental procedures. Note that, for Soler, the initial results were different from the result finally accepted not only at the level of raw data, but also (and most importantly) at the level of theoretically interpreted data. Soler argues that in this case the no-coincidence argument is only apparent and has no persuasive power, because of a lack of independence. Not of traditional theory independence, however: the initially divergent experimental results that were mutually adjusted to support the same final conclusion may as well have been based on independent theoretical principles. What was lacking in this case is *genetic independence* (Soler 2012a, 2012b).

3. The dimensions of multiple determination

These examples reveal some of the most important ways in which concrete applications of the multiple determination strategy can differ from the ideal epistemic situation implied in the blunt rationale. The complexity of actual scientific research, however, should not be a reason to abandon the effort for a unified assessment of the epistemic import of multiple determination. We can account for this complexity within a single framework in which the ideal epistemic situation plays an essential role. The previous examples show that—even though the ideal epistemic situation is rarely achieved in actual scientific practice—multiple determination is not a

philosophers' invention, but an important strategy employed and appealed to by the researchers themselves. The clathrin-mediated endocytosis example, in particular, presents an important way in which the ideal epistemic situation is important for scientific research. Namely, it plays the role of a *methodological attractor*: an (not so easy to achieve) ideal goal that serves to organize and guide the efforts of the experimentalists. I claim that, in addition, the ideal epistemic situation can play an *evaluative* role—it can help to “measure” the divergence of a concrete case of multiple determination from the desired ideal. In the rest of the paper, I provide a concrete way by which to conduct such an evaluation. More specifically, I distinguish among the dimensions of multiple determination: the elements of the multiple determination strategy that are responsible for the force of the no-coincidence argument (the distinguishing feature of the multiple determination strategy). Assessing how much the instantiation of these dimensions in a particular case differs from their ideal form, helps in evaluating the epistemic force of the no-coincidence argument made in that particular case.

Before proceeding with the presentation of the various dimensions, two clarificatory comments are in place. First, the evaluation of concrete cases of multiple determination is not meant to work as an algorithmic procedure that will offer a quantitative measure of their epistemic force—each evaluation would be qualitative rather than quantitative in nature. Second, the list of the various dimensions upon which the evaluation of concrete cases is to be made (as it is presented here) is open-ended. It is not meant to be exhaustive and should not be taken as the final word on the topic. It can be extended, elaborated, and revised as we look at additional cases of multiple determination.

4. Independence of determinations

A first dimension of multiple determination is the independence of the determinations that establish the same result. The ability of a procedure to establish a certain result with respect to a phenomenon of interest should not depend on the ability of another procedure to do the same.

Genuine independence is essential for the very possibility of the strategy: if the determinations are not (somewhat) independent, they would not count as different (i.e., multiple) determinations. Independence is essential for the emergence of the no-coincidence argument encountered in the blunt rationale.¹⁰

If we look at actual scientific research, however, the situation is more complex. The main problem is establishing that the procedures determining the same result are indeed independent. Before we can even attempt to solve this practical problem, however, we need to be clear on what is meant by ‘independence’ of the determination procedures. It seems that scientists work with a rather intuitive notion of independence. Even in philosophical treatments, the meaning of independence is rarely analyzed explicitly. The historical and philosophical analysis of the multiple determination strategy reveal at least three kinds of independence: *theoretical independence*, *genetic independence*, and *historical independence*.

4.1. *Theoretical independence*

At least since Pierre Duhem (1894), most philosophers of science accept that (most) experimental evidence is theory dependent.¹¹ They are, therefore, familiar with a particular notion of theoretical

¹⁰ The Bayesian analyses of the variety-of-evidence thesis have discussed various formal notions of evidential independence (Bovens and Hartmann 2003; Claveau 2013; Stegenga and Menon 2017; Claveau and Grenier 2019).

¹¹ Duhem argued that an experiment in physics is not simply the observation of raw data (concrete facts, instrument readings, etc.), but also the theoretical interpretation of these raw data. The process of theoretical interpretation replaces the raw data gathered from observation “*with abstract and symbolic representations that correspond to them in virtue of the physical theories admitted by the observer*” (Duhem 1894, 182, emphasis in the original). Whereas the observation of raw data is available to the theoretically uninitiated observer, the process of theoretical interpretation requires a deep knowledge of physical theory

independence: to avoid the circularity implied by the so-called Duhem-thesis, whenever one sets up an experiment to investigate a phenomenon or test a hypothesis of interest, the theoretical hypotheses underlying the experimental procedure ought to be independent from the theory of the item under investigation.¹² In cases of multiple determination, besides the independence of the theory doing the testing from the theory being tested we are also (in fact, we are mostly) concerned with the independence between the theories underlying the different experimental procedures that establish the same result. The theories underlying the different procedures ought to be two-by-two independent.

The most cogent way to deal with this form of theoretical independence has been proposed by Peter Kosso (1988, 1989; see also Culp 1995). According to Kosso, for two procedures to be theoretically independent they must rely on independent background theories: the background theoretical auxiliary assumptions that underlie the procedures' functioning and interaction with the object under investigation (which assumptions are used to interpret and transform the initial raw data produced by the procedures into experimental evidence) ought to be two-by-two independent. And the latter in the sense that the validity (or acceptance) of the auxiliary assumptions underlying one procedure should not depend on the validity (or acceptance) of the auxiliary assumptions underlying another procedure.

The problem of the theoretical independence of two determination procedures is thus reduced to that of examining the independence between the two sets of auxiliary assumptions that are responsible for the claim that the results obtained by the two procedures are "the same". To apply this suggestion concretely, however, we need to make clear what kind of independence is required: total or partial? Ought all the auxiliary assumptions that underly one procedure to be

¹² The theoretical hypotheses (auxiliary assumptions) underlying the experimental procedure are assumed to be valid and are not questioned during the experiment, but they can be the subject of independent experimental testing if doubts about their validity arise.

independent from all the auxiliary assumptions that underly the other procedure? This seems to be a very stringent requirement. As the DNA sequencing example shows, theoretical independence is often a question of degree. One practical problem, therefore, is that of determining the degree of theoretical independence that obtains between two procedures and assessing its impact on the no-coincidence argument. Another related problem is that of identifying all the auxiliary assumptions that underly a specific procedure, given that not all of them may be explicit. A final source of complication has to do with whether the determination of theoretical independence is itself theory dependent: that is, whether the decision to regard two sets of theoretical assumptions—for example, the set of theoretical assumptions underlying the working of fluorescence microscopy and the set of theoretical assumptions underlying the working of electron microscopy—as theory independent may itself depend on background theoretical knowledge. The response to this problem is not very different from the response to the Duhem thesis: the (second order) theoretical dependence of the assessment of (the first order) theoretical independence is not problematic as long as the background theory doing the assessment is different from the theories being assessed.

This analysis of theoretical independence provides the conceptual tools for assessing the level of theoretical independence that obtains in concrete cases of multiple determination. To determine the degree of independence that obtains in a concrete case and assess its impact on the no-coincidence argument, it is necessary to pay attention to the details of the case. The importance of the degree of theoretical independence for the force of the no-coincidence argument is straightforward: the more theoretically independent are the determination procedures that establish the same result, the stronger is the argument.

4.2 Genetic independence of the multiply determined results

Genetic independence is rarely discussed in philosophical analyses of the multiple determination strategy. Léna Soler is an author who discusses this notion explicitly (Soler 2012a). Even within

Soler's discussion, however, we can distinguish two different kinds of genetic independence that are not explicitly distinguished in her analysis. First, we have the genetic independence of results. To have this kind of independence, initially discordant results obtained by theoretically independent procedures should not be transformed, mutually adjusted, or in any other way changes so as to converge on a single result which is then, retrospectively, considered to have been the result of multiple determination. This is the version of genetic (in)dependence at play in the 'discovery of weak neutral currents' example above. It is also the notion of independence we have been concerning ourselves all along in the witnessing example. In the latter, we are not concerned with the theoretical independence of the witnesses. In fact, we assume that the witnesses are identical with respect to their perceptual and cognitive abilities.¹³ We are concerned with the possibility that the witnesses might have changed their original discordant testimonies to achieve agreement. In scientific research, however, we are concerned with both the theoretical and the genetic independence of experimental results. The no-coincidence argument cannot be made if researchers mutually adjust or calibrate theoretically independent discordant results in order to achieve agreement.

Given this version of genetic independence, the important question is: how can we go about ascertaining its existence in a concrete case? Although establishing the existence of genetic independence may not always be an easy task, the best way for doing so is by looking at concrete cases of multiple determination from a temporal perspective, i.e., by looking at the establishment of concordance in real time, as opposed to only relying on retrospective analyses. This is a strong argument why the temporal dimension of science is important for philosophy of science (at least for a philosophy of science that is concerned with the structure and epistemic import of scientific arguments).

¹³ Additional evidence from CCTV footage, or DNA analysis, for example, would constitute evidence that is theoretically independent from the witnesses' testimonies.

4.3 *Historical independence of the determination procedures*

There is, however, another kind of genetic independence. The historical independence of the determination procedures that are employed in a case of multiple determination.¹⁴ To have this kind of independence the reliability of a determination procedure, during its historical development, should not have been exclusively based on its ability to produce results concordant results with those obtained by a determination procedure that was judged to be very reliable. For example, if electron and fluorescence microscopy were historically mutually adjusted or calibrated on one another so that they would constantly produce similar visual displays when investigating the same phenomena, there would not be much of a no-coincidence argument to be made if, when used to investigate a new phenomenon, they continued to produce concordant results. In broad terms, the concern is that at some point in the development of science a determination procedure or body of knowledge becomes so entrenched, that the reliability of a new experimental procedure is judged exclusively on how well it agrees with the entrenched item (Wimsatt 2007, 133-58). Note that the concern does not arise from the entrenchment *per se*, but from the claim that the processes by which determination procedures or bodies of knowledge become entrenched are historically contingent (Hacking 1992; Pickering 1995, 2012; Wimsatt 2007; Soler 2012b; Boon 2012).

¹⁴ Genetic and historical independence do not depend on one another. That is, one can have the one without the other. They are closely related, however. Their failure to obtain ultimately rest on processes of mutual adjustment and/or calibration of results obtained by different procedures. The difference is that in the case of genetic dependence of the results these processes occur within the context of the multiple determination case at hand, whereas in the case of historical dependence of the procedures the processes of calibration and/or mutual adjustment have occurred (over longer periods of time) outside the context of the multiple determination case at hand.

Like genetic independence, historical independence is also unrelated to whether the experimental procedures are theoretically independent, although (as I argue below) they both ultimately rely on the assumption that theoretically independent results and procedures are “plastic”; that starting with the same raw materials we can arrive at radically different conclusions. To establish the historical independence of two determination procedures employed in a concrete case of multiple determination we need to consider their historical development. This is an even more important argument for the importance of the historical development of science for the philosophy of science. The relationship between genetic and historical independence and the force of the no-coincidence argument is straightforward: the more the results and the determination procedures are judged to be genetically and historically independent, the stronger is the force of the argument.

5. Reliability of a single determination

A second dimension is the reliability of a single determination. This dimension refers to how reliable a determination procedure is in general, independently of how clearly it suggests a specific result. For example, fluorescence microscopy is regarded to be a reliable method—it is well-tested and well-understood. Nevertheless (as the CME example above also shows), because of its limited power of discrimination often it does not provide very clear imaging of the specimen under investigation. The relationship between reliability and the no-coincidence argument is straightforward: the more reliable are the determination procedures that establish the same result, the stronger is the no-coincidence argument.

Note that in the witnessing example no differentiation is made between the various testimonies with respect to their reliability. In actual cases of multiple determination, however, is common for some procedures to be considered more reliable than others. This situation has given rise to claims that what actually happens in cases of multiple determination is a process of

calibration or mutual adjustment where results obtained with procedures considered to be less reliable are adjusted to results obtained with procedures that are considered to be more reliable. The differential reliability of the procedures, however, by itself, is not sufficient for the existence of genetic dependence. On the other hand, the problem of the historical dependence of the determination procedures suggests that the reliability of a procedure depends on a diachronic process of calibration, during which experimental procedures are rejected or adjusted depending on how they compare with already entrenched procedures. It seems, however, that experimentalists have a rather large repertoire of independent general epistemic strategies for establishing the reliability of experimental procedures. Allan Franklin (1986; 2002) has developed what he calls an ‘epistemology of experiment’, a set of strategies that provide justification for rational belief in the reliability of an experimental procedure (and, consequently, in the validity of a result).¹⁵ In section 7, I argue that the multiple determination strategy itself can

¹⁵ We can regard the simultaneous employment of several of Franklin’s epistemic strategies in a concrete case, as a kind of *second order multiple determination*. In contrast with (first order) multiple determination (where independent procedures are employed to establish the same result), in second order multiple determination different *general epistemic strategies* are used to establish the reliability of an experimental procedure (and consequently the validity of the experimental result obtained by that procedure). Gandenberger’s (2010) account of the introduction of the cathode-ray oscillograph into electrophysiology during the 1920s, and Bechtel’s (1990) analysis of how electron microscopy and cell fractionation were employed to investigate the structure and function of the cell, can be construed as examples of second order multiple determination. Although first and second order multiple determination are not mutually exclusive (they can be combined in a concrete case to generate an even stronger no coincidence argument), they usually play different roles in scientific practice. Second order multiple determination is a very helpful strategy in situation where (first order) multiple determination is not available (Franklin 1997). Second order multiple determination is also a preferred strategy in cases where independent experimental

be used to argue about the validity of the procedures that establish the same result.

6. Clarity of a single determination

A third dimension is the clarity of a single determination. Clarity refers to how clearly the determination procedure suggests a specific result, independently of whether we have good reasons to believe in the reliability of the procedure itself. As we saw (in section 2.7) above, electron microscopy in its early uses in experimental investigations during the 1950s was able to produce clearer imaging than the other microscopy methods available. Nevertheless, it was not considered to be a very reliable procedure because its use was based on auxiliary assumptions that were considered speculative, and its inner workings were not well understood at the time.

A determination procedure can produce a clear result by producing as an outcome a very clear and detailed visual pattern—if the procedure is a microscopy or telescopic method, for example (see sections 2.1; 2.4; 2.7). A determination procedure can also produce a clear result by providing a precise numerical value for a magnitude being measured (as opposed to only providing upper or lower limits, or a range of values). This was the case in Jean Perrin’s determination of Avogadro’s number that was based on the experimental study of the height distribution of Brownian particles. One of the reasons why this specific determination was considered important for establishing the existence of molecules, despite there being other available determinations, was the precise determination it allowed for the value of N (Coko 2019, 198). The clarity dimension is important for assessing the quality of convergence of the independently established results.

procedures lead to discordant results and there is the need to assess the reliability of each one of the procedures separately (see the discussion of discordance in section 10 below).

7. Quality of convergence of the independently established results: multiple determination of (theoretically interpreted) results and multiple determination of (theoretically uninterpreted) raw data

The degree to which results obtained with independent procedures agree with one another gives rise to a fourth dimension of multiple determination: the quality of convergence of the independently established results. This dimension refers to how outcomes obtained with independent experimental procedures compare to one another and the degree to which they can be said to have established the same result. The relationship with the no-coincidence argument is clear: the more the different procedures are considered to have established the same result, the stronger is the argument.

In Perrin's example, the quality of the convergence was judged to be extremely high at the time, especially if one considered the possible values for N that were possible in each one of the various determinations.¹⁶ The high degree of convergence between the (theoretically and genetically) independent determinations of Avogadro's number played a crucial role in convincing the scientific community for the reality of atoms and molecules (Nye 1972, Coko forthcoming).

The quality of convergence is a dimension that has been the source of much misunderstanding. A main reason for this is the way in which multiple determination in actual research differs from the blunt rationale. Let us consider again the witnessing example. The no-coincidence argument emerges because different witnesses—with identical perceptual and cognitive systems, and who are considered to be equally reliable—independently of one another,

¹⁶ According to Perrin, the values that were possible for N in his own experimental determinations extended from zero to infinity. The fact that his determinations, independently of one another, from all the possible values for N converged on one specific value, made the argument for molecular reality “bordering on certainty” (Perrin 1909, 111).

testify that a certain event occurred. With respect to the question of the event's occurrence, the witnesses' testimonies are identical (or sufficiently similar). The relevant notion of independence in the witnessing example is that of genetic independence. In actual scientific research, however, the primary notion of independence is that of theoretical independence. Theoretically independent procedures, however, almost by necessity of being theoretically independent, often yield completely different outputs at the level of raw data. Further, if we accept that: (a) the theoretical interpretation of the raw data is ubiquitous for experimental results to be epistemically relevant, and (b) for two experimental procedures to be independent they must use independent sets of theoretical auxiliary assumptions to interpret the raw data, it follows that it is only at the level of the theoretically interpreted data that we can say that two procedures establish "the same" result—which is often the same with respect to some local hypothesis of interest.

The fact that theoretically independent determination procedures do not yield literally the same outputs, however, has been claimed to be problematic for the multiple determination strategy.¹⁷ It is at the heart of Jacob Stegenga's notion of "incongruity", which he considers to be

¹⁷ A main source of misunderstanding is the consideration of multiple determination as a *robustness* variant. In most of the recent literature, multiple determination is referred to as 'experimental robustness', 'measurement robustness', 'robust detection', or simply 'robustness'. The problem with this terminology is that the term 'robust'—whenever it is used rigorously, and not simply as a synonym for 'solid' or 'strong'—refers to something that remains invariant in face of perturbations. Indeed, in the context of the robustness literature, the result that is determined by independent procedures is considered to remain invariant (i.e., literally the same) across the different determination procedures. This, however, is rarely the case. Considering the multiple determination strategy as some sort of invariance to perturbations misdiagnoses its structure and it is not helpful for understanding its epistemic import. The essential characteristic of the multiple determination strategy is its ability to support a no-coincidence argument; no such argument can be made from invariance to perturbations. In addition, considering multiple determination as some sort of invariance to perturbations blurs a crucial distinction in experimental

a “hard-problem” for multiple determination (Stegenga 2009, 2012).¹⁸ Similar skeptical conclusions have been derived from other philosophers (Soler 2012b; Boon 2012; Pickering 2012). The fact that the convergence of experimental results is achieved only after the raw data have been (independently) theoretically interpreted, however, is not necessarily a hard problem for multiple determination. It only means that the concordance of the results is not immediately given as such. Additional steps are required to establish the existence of agreement. It is certainly true that these additional steps may make the concordance claim a bit more uncertain; researchers may be mistaken about their theoretical auxiliary assumptions, or about the manner in which they use them to interpret the raw data; they may disagree about how to theoretically interpret the raw data; and there is always the danger of the genetic dependence of the (independently) theory interpreted data. However, this is a problem only if we are under the misconception that results

research. Researchers often use the sensitivity of a result to variations of the experimental parameters *within sufficiently similar experimental procedures* to establish causal dependencies. This commonly used strategy, however, is considered (both historically and conceptually) to be different from the ability to establish “the same” result by means of independent experimental procedures. Schupbach (2018) is a recent example where the robustness (invariance) of the phenomenon of Brownian movement to the experimental manipulation of the various suspected causal factors, during the course of the nineteenth century, is considered to have the same epistemic rationale as Jean Perrin’s description of thirteen different procedures for determining Avogadro’s number. Elsewhere (Coko 2015), I provide a historical account of the distinction between these two strategies.

¹⁸ For Stegenga, incongruity occurs because evidence produced by theoretically independent procedures is written in different theoretical languages: procedure *A* establishes *x*, procedure *B* suggests *y*, procedure *C* claims *z*, and so on.

produced by different procedures should be literally the same for us to be able to say that they establish the same result. But this occurs only in cases of multiple determination of raw data.¹⁹

Multiple determination of raw data offers a strong no-coincidence argument for the non-artificiality of the data. The strength of the argument comes from the fact that it is not based on comprehensive theoretical considerations regarding the entities being investigated, the experimental techniques used, and the interaction between the two. This strength, however, comes with a price. Exactly because of the limited role played by theoretical considerations, multiple determination of raw data plays only a limited role in scientific research. It is better suited for establishing that the raw data are not artifacts of the experimental procedures employed (assuming that the procedures are independent). Multiple determination of raw data is not sufficient for establishing the theoretical identity of the raw data. In addition, genuine cases of multiple determination of raw data are rare in experimental research. Even when they occur some knowledge of the underlying theoretical principles is needed to establish the sameness of the result and/or the theoretical independence of the determination procedures.²⁰

¹⁹ Even in such cases the outputs are not literally the same. In the dense bodies example (section 2.1), the colored film realized with fluorescence microscopy was not identical with the black and white picture obtained with electron microscopy. It was the structural similarities between the two visual displays that gave rise to the no-coincidence argument.

²⁰ In Hacking's example, it is doubtful whether the no-coincidence argument establishes that the dense bodies are cell structures. Without employing any theoretical considerations, we can only conclude that the dense bodies are not artifacts of the techniques—that they are really out there “in the world”. But this is not the same as establishing that they are cell structures. They might be artifacts of another nature. For instance (as the mesosomes example shows), they might be artifacts of the fixing and staining techniques used to prepare the specimens for microscopical observation. In addition, even in cases of multiple determination of raw data the cogency of the argument depends on the assumption that the determination procedures used are (at least) theoretically independent. But how can one establish that the determination procedures are

To summarize, theoretically independent experimental procedures, exactly because they are based on different “chunks of physics” (to use Hacking’s phrase), do not generate, literally, the same experimental outputs. Theoretical considerations play a crucial role for establishing the epistemic identity of the experimental result as well as it being “the same” with results obtained with independent experimental procedures. For experimental results to be epistemically relevant, they need to be mediated by what we already know. Given the ubiquity of the theory dependence of experimental results, what matters about the quality of the convergence of results obtained with theoretically independent procedures is not whether the experimental outputs at the level of raw data are the same, but whether the auxiliary assumptions underlying the interpretation of the raw data are reliable. Seen in this manner, the problem of incongruity appears to be just another form of the Duhem-thesis.

There is already a large body of literature on how to respond to the Duhem-thesis.²¹ Here I will focus on how the multiple determination strategy itself can be used to argue about the reliability of the assumptions underlying the independent procedures that establish the same result. When the dimensions give rise to a *strong* no-coincidence argument, the multiple determination strategy can be used to argue, not only for the validity of the result, but also for the reliability of the auxiliary assumptions underlying the procedures. This reasoning is not necessarily troublesome or vicious. Since the emergence of the problem of the *experimenters’ regress*, philosophers have been aware of the interdependency between the correctness of an experimental result and the reliability of the experimental procedure used to obtain that result: a

theoretically independent without any knowledge of the theoretical assumptions underlying their operation and interaction with the object under investigation? Contrary to Kuorikoski and Marchioni (2016), it seems that the evidential status of data requires knowledge of the theoretical principles underlying the data generating processes.

²¹ See, for instance, Laudan and Leplin (1991), Psillos (1999), and Zammito (2004).

correct result is generally considered to be one produced with a reliable and properly functioning experimental apparatus, but a reliable and properly functioning experimental apparatus is the one that produces the correct result, and so on. In other words, we don't know if we have obtained the correct result unless we have used a reliable procedure, and we don't know if we have used a reliable procedure unless we have obtained the correct result (Collins 1985).

It has been argued that the experimenter's regress raises general concerns about experimental evidence and its role in theory evaluation (Collins 1985). It seems, however, that the problem is more serious in cases where a new and untested experimental procedure is used to investigate a novel phenomenon, and where no independent means of investigation are available. Insofar as the multiple determination strategy helps to break the regress by arguing about the validity of the result, it can also be used to argue about the reliability of the procedures. This form of reasoning is already implicit in the witnesses example and in cases of multiple determination of raw data.²² The ability of the no-coincidence argument to support both the validity of the result and the validity of the auxiliary assumptions underlying the various independent procedures is of special importance in cases of multiple determination of theoretically interpreted results. The early use of electron microscopy to establish the existence of dislocations of molecular lattices in crystals (section 2.7) and the multiple determination of visual images (section 2.4) are relevant examples. This same form of reasoning was also heavily used in Jean Perrin's experimental

²² In the witnessing example we infer from the convergence of the testimonies not only (*the ontological claim*) that the event about which the independent witnesses report occurred, but also (*the epistemological claim*) that the witnesses are telling the truth (that they are reliable)—it would be an improbable coincidence for the witnesses to have independently concocted the same story. Similarly, Hacking's no-coincidence argument argues simultaneously for both the reality of the "dense bodies" and the validity of the microscopy techniques: the techniques ought to be flawed in a very remarkable way to produce the same error independently of one another.

investigation of Brownian movement. For example, Perrin used the agreement between the numerical value for N obtained in his height distribution experiments with the value for N theoretically inferred in the kinetic theory of gases to argue, not only for the correctness of the result, but (perhaps even more importantly) also for the validity of some doubtful theoretical and experimental assumptions upon which his result was based.²³

8. “Plasticity” of results and determination procedures

A fifth dimension of multiple determination concerns the “plasticity” of results and determination procedures. Given that most determination procedures involve a process of theoretical transformation of raw data into epistemically relevant experimental results, an important question is: how “plastic” or malleable are the raw data? How plausible is that beginning with the same experimental output at the level of raw data, researchers can infer different results. This dimension of multiple determination is important because of the problem of the genetic dependence of experimental results. The latter ultimately rests on the assumption that raw data are malleable and amenable to different theoretical interpretations. One reason why the (theoretically independent) multiple determination of raw data provides such a powerful no-coincidence argument is because the experimental outputs are relatively rigid: what you “see” is what you get. There is no argument to be made that the raw data were transformed so they would agree with one another. The no-coincidence argument from the concordance of raw data is still vulnerable to the problem of the historical dependence of experimental procedures. The latter, however, is also ultimately based on the claim that theoretically independent experimental procedures are relatively “plastic” and can be easily adjusted to produce concordant raw data. The

²³ In fact, it can be shown that Perrin’s goal was to counteract Duhem’s skeptical argument that experiments in physics (exactly because they are theory-dependent) cannot establish the validity of claims about unobservable entities (Coko 2019).

relationship of this dimension with the no-coincidence argument is straightforward: the more rigid and standardized are the independent inferential procedures that are used to establish “the same” result, the stronger is the no-coincidence argument.

9. Complexity of the independently established result

A sixth dimension of multiple determination is the complexity of the independently established result. This dimension concerns the issue of how complex is the result that is multiply determined. The no-coincidence argument is more forceful, the more complex or intricate is the result determined by independent procedures.²⁴ Going back to the witnessing example, consider the case in which two independent witnesses are asked to report on the result of a basketball game. Even if the witnesses’ reports agree on which team won the game, we still would not have a strong no-coincidence argument for the validity of their report. There is a roughly fifty percent chance the witnesses’ reports would agree even if none of them knew the game’s score—and this assuming that the game was one between two equally strong teams. On the other hand, consider the situation where the witnesses’ reports, besides agreeing on which team won, also agreed on the game’s final score, on the name of the player who scored the winning basket three seconds from the end of the game, and other details. The more detailed is the report upon which the independent witnesses agree, the more improbable is the possibility that they independently made-up the report.²⁵

²⁴ The complexity or intricacy of the result produced by a determination procedure can be used to argue about the non-artificiality of the result even when corroborating evidence from independent procedures is lacking (Bechtel 1990; Franklin 2002). Being able to produce such complex patterns with independent procedures highly increases the force of the no-coincidence argument.

²⁵ This aspect of multiple determination was present in the early writings on the multiple determination of psychological test results. Campbell and Fiske (1959), for example, besides what they called “convergent

This point is readily applicable to cases of multiple determination of raw data where, insofar as the outputs are structurally similar, the similarity extends also to their complexity. But is there an analogous argument from the convergence of complex results to be made in cases of theoretically interpreted results? In the latter case, the initial experimental outputs at the level of raw data need to be theoretically interpreted before one can claim that they establish the same result. Can the “sameness” in this case refer to an independently determined complex result, as opposed to the various determinations simply providing a “yes” or “no” verdict with respect to the validity of a hypothesis of interest? I claim that there is. And it is to be found in the quantitative determination of an experimental result. For example, what made Perrin’s argument for molecular reality so convincing, was the accuracy by which all the independent procedures (of all the possible outcomes) converged at similar values for N . With respect to the possible values for N that were possible *a priori*, the actual value upon which all the various determinations converged constituted an extremely complex result (Coko forthcoming).

10. Lack of discordant results and compatibility with already accepted knowledge

A seventh dimension of multiple determination is the existence of conflicting results among independent investigations of the same phenomenon. In ongoing scientific research, scientists are often faced with both concordant and discordant results. According to Stegenga, who is also the main exponent of the “discordance problem”, more often than not, evidence generated by independent procedures is discordant: whereas procedure A indicates h , procedure B indicates $\neg h$ (Stegenga 2009, 2012). For Stegenga, discordance is another “hard-problem” for multiple determination. As he correctly points out, discordance is not visible in retrospective readings of past science, which often produce an artificial picture of concordance by “forgetting” or

validation” (i.e., different tests converging on the same result), required also “discriminant validation” (i.e., the ability to detect spurious convergence).

dismissing all the results that were in conflict with the result finally accepted. To illustrate the situation, Stegenga employs a case-study, which concerns the scientific investigation regarding the manner in which influenza virus is transmitted from one person to another. Two hypotheses have been proposed: influenza virus is transmitted via an airborne route or influenza virus is transmitted by contact. Empirical evidence is called to decide between the two alternatives. The evidence generated, however, is discordant. Some scientists argue (using mathematical models and animal experiments) that influenza is transmitted via an airborne route, whereas others (based on clinical experience and observational studies) argue that influenza is transmitted via a contact route.

Discordance is somewhat different from the other dimensions. To assess the effect of conflicting results on the no-coincidence argument is not sufficient to weigh the positive *vs.* the negative results, and then decide which alternative to accept. The existence of conflicting results is something that requires an explanation—the conflict needs to be resolved before the no-coincidence argument can even be made. If the existence of conflicting results is a widespread phenomenon, and attempts to resolve the conflict are often unsuccessful, then discordance would indeed set serious limitations to the epistemic relevance of multiple determination for scientific research.

Discordance can be a hard problem for multiple determination, but not a fatal one. First, it is worth pointing that the problem of conflicting results is in tension with the problem of genetic dependence and the mutual adjustment of experimental results. If genetic dependence and mutual adjustment of experimental results is so widespread and easy to achieve, as some critics of multiple determination claim, then the existence of conflicting results “should not be a problem”.²⁶ If, on the other hand, the existence of conflicting results is widespread, then it seems that mutual adjustment of conflicting experimental results may not be so easy to achieve after all.

²⁶ We would have a rampant problem of genetic and historical dependence instead.

Second, the question of whether the results produced by independent procedures are frequently conflicting is an empirical one, to be decided by looking at a large number of cases. The fact that the existence of conflicting results is often invisible in retrospective reconstructions of scientific research gives an additional reason for following a historical approach when investigating cases of multiple determination. We need to look at these cases from a temporal perspective—look at the achievement of concordance in real-time, as opposed to only looking at the final reconstruction of relevant cases as they may appear in a published book or research article (this is the approach that Trizio follows in the CME example above). Third, cases where evidence from independent procedures remains conflicting or unresolvable for a long time, rather than showing the inadequacy and limitations of the multiple determination strategy, often indicate some deficiency in the way that the empirical question to be decided by that evidence is posed. This seems to be the case with the influenza transmission example. As a recent reinterpretation of the case indicates, for the influenza example to work it is necessary that the airborne and contact transmission alternatives are mutually exclusive. This requirement, however, seems to have been long abandoned—both the US and European centers for disease control explicitly state that influenza can be spread in either fashion (Hey 2015). Thus, a change in the way that the problem of influenza transmission is conceived, shows that the evidence was not conflicting after all. Rather than being an example against the relevance and adequacy of multiple determination, this case shows that the strategy functions exactly as it should. Further, it suggests that the claim that incongruous or conflicting independent results can be easily changed or adjusted to support a particular claim, is highly implausible. Rather than a hard problem for the multiple determination strategy, the protracted and unresolvable existence of conflicting results is often the starting point of conceptual revolutions.

11. Number of Determinations

An eighth dimension of multiple determination is the number of determinations that establish the same result. Intuitively, the relationship between the number of determinations and the force of the no-coincidence argument is straightforward: the greater is the number of determinations that establish the same result, the stronger is the argument. However, there is still room to make this intuition clearer by looking at the role this dimension plays in actual research. Intuitively three determinations are better than two, and thirteen are better than twelve, but where should we stop? How many independent determinations are sufficient to establish the validity of a result?²⁷

Using the dimensions framework to approach this question, makes it clear that the number of determinations judged to be sufficient for establishing a result depends on the contributions that the other dimensions make to the no-coincidence argument. The less the various determinations are judged to be independent (theoretically or genetically), the greater is the number of determinations needed. The more reliable the determinations, the lower the number of determinations needed. The more the result challenges accepted knowledge, the greater is the number of determinations needed, and so on. For example, in the molecular reality case, Jean Perrin (at least initially) did not really need thirteen different determinations to establish the validity of the kinetic molecular explanation of Brownian movement. The (rough) agreement between the value for N determined in his height distribution experiments with the value for N inferred independently in the kinetic theory of gases, was judged to be sufficient—at least in his own eyes. What made this agreement sufficient for Perrin was the fact that it was a surprisingly good numerical (quantitative) agreement, which was obtained by (genetically and theoretically) independent determination procedures, and which was extremely striking if one considered all the

²⁷ There appears to be a tension between the epistemic relevance of the multiple determination strategy and the number of independent determinations required to establish the validity of a result: the greater the number of determinations needed to establish the validity of a result is, the less relevant multiple determination seems to be for scientific research. The framework proposed here resolves this tension.

values for N that were possible in his height distribution experiments. Insofar as Perrin felt the need to provide further independent determinations for Avogadro's number, it was to counteract the skepticism regarding the molecular theory of Brownian movement that emerged when other independent experimental efforts to determine N , which were based on Einstein's mathematical treatment of Brownian motion, gave discordant results with those obtained in his experiments (Perrin 1909; Coko forthcoming).

It goes almost without saying that there is no algorithmic procedure for establishing how many determinations are sufficient in a concrete case. The number of determinations deemed to be sufficient will depend on the contributions that the other dimensions make to the no-coincidence argument. But researchers may disagree on how the different dimensions affect the no-coincidence argument: they may disagree on whether the reliability of a determination is more important than its independence from other determinations; whether the complexity of the multiply determined result is more important than the reliability of the determinations, and so on. In addition, they may disagree on whether a determination procedure is sufficiently reliable, independent, clear, and so on. These disagreements, however, become less acute the greater the number of determinations establishing the same result is.

Is this malleability in evaluating the validity of a result a problem for multiple determination? As Kuorikoski and Marchionni (2016) point out, many critics of multiple determination sometimes argue as if the strategy has little or no epistemic value, because it does not always deliver the truth. Against this line of argument, Kuorikoski and Marchionni claim that the value of the multiple determination lies not in its being a foolproof strategy, but in its ability to increase the reliability of inferences that are based on fallible procedures: "the challenge is to identify the circumstances in which confidence in the result rationally increases on the basis of the concordance between independent methods of determination" (Kuorikoski and Marchionni 2016, 230). I think this is a correct assessment. By identifying the elements that are responsible for the epistemic force of multiple determination, the dimensions framework succeeds exactly in

this task. Multiple determination, however, is not a panacea that solves all the traditional philosophical problems surrounding the validity of experimental results. In many concrete cases there might still be room for disagreement. But as pointed out in the previous section, cases when multiple determination fails to resolve the disagreements over long periods of time, rather than showing the limits of the strategy, indicate the existence of a deficiency in the way the problem to be solved by multiple determination has been posed, and are often the starting point of conceptual overhauls.

The dimensions framework is also helpful for answering the important question(s): namely, whether (and why) determining a result with several procedures is better than determining it with a single procedure? Why should we regard a result established with a reliable experimental procedure as less certain than a result established by several mediocre experimental procedures about the reliability of which we are not certain? Using the dimensions framework, the answer is clear: if we have a very strong determination of a result, from a procedure that is well-tested and relies on strong auxiliary assumptions, there is less need for independent determinations.²⁸ This claim is strengthened by what happens in mathematics. Because mathematical proofs are considered to be infallible (the result is deductively derived from accepted premises) there is no need for multiple proofs, as long as we are convinced for the validity of our deductive steps. It would be a complete waste of time with respect to establishing the validity of the proof. In empirical science, however, contrary to what happens in mathematics, one is never certain about the validity of the many auxiliary assumptions upon which the validity of an experimental determination relies.²⁹ This is why the need for independent determinations is ubiquitous, and

²⁸ See also Cartwright (1991).

²⁹ This problem lies at the heart of the Duhem thesis and the problem of the experimenter's regress. There may also be doubts about the implementation of a reliable procedure in a particular case.

why it is almost always the case that establishing a result by independent procedures is better than establishing it only by a single procedure.

It is worth mentioning, however, that the view of mathematical proofs as infallible and radically different from the inductive inferences made in the empirical sciences has been recently challenged. Mathematicians often offer independent proofs for the same theorem (Krömer 2012). This claim does not conflict with the argument above; if anything, it makes it stronger. Insofar as mathematicians feel the need to offer independent proofs for the same result, it is because of the possibility of errors creeping in undetected even in the most rigorous of deductive proofs (Krömer 2012). Therefore, the need for multiple independent determinations must be higher in experimental science where, because of the complexities involved, the possibility for error is also much higher.³⁰ What the case of mathematics shows, is exactly that the need for more independent determinations is less, the more the determinations are judged to be certain, and *vice versa*.

To sum up, the dilemma one *vs.* multiple determination is a false one. The important question in scientific research is not whether it is better to have one or multiple determinations of the same result. Rather, it is what to do in cases when a single determination is judged to be insufficient. This is where the multiple determination strategy enters. Take, for example, the case of molecular reality and the determination of N . Because of the large amount of auxiliary assumptions involved in determining the number of (unobservable) particles contained in a unit of a substance, no theoretical or experimental determination, *by itself*, could ever be sufficient to establish both the validity of the result and the validity of the determination procedure.³¹ Only the exceptionally strong no-coincidence argument that emerged when independently theory-

³⁰ Hon (1989) provides a catalogue of possible sources of error when procuring evidence by means of scientific instruments.

³¹ One would be faced with the problem of the experimenters' regress.

dependent procedures converged on the same value for N could be used to argue both for the validity of the value found and the validity of the auxiliary assumptions underlying the various determination procedures. Finally, even in the case when researchers use only a single determination procedure, this is because the multiple determination strategy either has been applied at an earlier stage or it has being applied at a meta-level: that is, researchers either have established the validity of the auxiliary assumptions underlying the determination procedure by independent means, or they have established the validity of the determination procedure by using other general epistemic strategies.³²

12. Conclusion

Recent studies of concrete cases of multiple determination have revealed that the traditional blunt rationale offered to explain the epistemic value of the strategy does not capture the complexity of scientific research. Researchers often do not have independent procedures to investigate the same phenomenon or establish the same result. Even when multiple procedures are available, they may not be entirely independent from one another or from the theory of the phenomenon under investigation, they may not be genetically independent, they may not be entirely reliable, they may differ in the strength they suggest the same result, they may not point to the same result (both at the level of raw data and of theoretically interpreted data), they may lead to discordant results, and so on. The divergence of concrete cases from the ideal epistemic situation implied in the blunt rationale is the main source of philosophical disagreements regarding the relevance and epistemic import of the multiple determination strategy for scientific research.

I have argued that the complexity of actual scientific research should not be a reason for abandoning the effort for a unified assessment of the structure and epistemic import of the multiple determination strategy. We can account for this complexity within a framework in which the no-coincidence argument encountered in the blunt rationale plays an essential role. There is a

³² See the description of second-order multiple determination in footnote 15.

large number of examples showing that multiple determination as a no-coincidence argument is not a philosopher's invention, but a strategy that plays an important role in scientific practice: as a way of securing the validity of results and methods, as a way of justifying them in the eyes of the scientific community, or as a regulative ideal organizing and directing the researchers' efforts. Distinguishing between the various dimensions of multiple determination helps in evaluating the epistemic force of concrete implementations of the strategy. We should not expect, however, this evaluation to be algorithmic offering a quantitative measurements of the strategy's epistemic force. Even though the framework suggested here can only offer qualitative assessments of concrete cases, distinguishing between the various dimensions is useful for several reasons. First, it can help dissolve much of the philosophical disagreements regarding the epistemic role and import of multiple determination, in general. Second, it can help resolve the historical and philosophical disagreements regarding the employment of multiple determination in concrete cases. For instance, it can explain why in some cases (such as Perrin's) the employment of the strategy was successful, whereas in other cases (such as that of the bacterial mesosome) it was not.

Third, and finally, this approach is useful for construing the relationship between the philosophy of science and concrete historical accounts of scientific practice as a two-way, mutually beneficial, relationship. The effort to understand the structure and epistemic import of the multiple determination strategy is open-ended: it aims to accommodate a large variety of concrete cases but is also sensitive to changes occurring within the strategy over time and across disciplines. This paper is part of this open-ended effort. We began with the blunt rationale for multiple determination. By examining concrete examples of the strategy, we arrived at a more refined and nuanced framework for understanding its structure and epistemic import. On the one hand, this framework can be used to achieve a better understanding of other—or even of the same—cases of multiple determination. On the other hand, the consideration of additional cases can help to further refine and sharpen our initial framework, and so on and so forth. This approach can

solve some of the problems surrounding the role of concrete case-studies within a philosophy of science that aims to arrive at general conclusions (Pinnick and Gale 2000; Pitt 2001), and paves the way for a mutually beneficial interaction between the history of science and the philosophy of science.³³ Thus, it is not only at the level of scientific content and methodology, but also at the (meta-) level of philosophical understanding that we make progress by building on what we already know.

References

- Allamel-Raffin, Catherine and Jean-Luc Gangloff. 2012. "Scientific Images and Robustness." Pp.169-188 in *Characterizing the Robustness of Science: After the Practice Turn in Philosophy of Science*. Edited by Léna Soler et al. Dordrecht: Springer.
- Aratzis, Theodore. 2012. "Hidden Entities and Experimental Practice: towards a two-way traffic between history and philosophy of science." Pp. 125-138 in *Integrating History and Philosophy of Science: Problems and Prospects*. Edited by Seymour Mauskopf and Tad Schmaltz. Dordrecht: Springer.
- Aratzis, Theodore and Jutta Schickore. 2012. "Ways of Integrating History and Philosophy of Science." *Perspectives on Science*, 20:395-408.
- Basso, Alessandra. 2017. "The Appeal to Robustness in Measurement Practice." *Studies in History and Philosophy of Science, Part A* 65-66:57-66.
- Bechtel, William. 1990. "Scientific Evidence: Creating and Evaluating Experimental Instruments and Research Techniques." *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association* 559-572.

³³ For more on this integrated approach to history and philosophy of science, see Aratzis (2012); Chang (2012); Schickore (2011); Aratzis and Schickore (2012); and Coko (2019).

- Bechtel, William. 2002. "Aligning Multiple Research Techniques in Cognitive Neuroscience: Why Is It Important?" *Philosophy of Science* 69: S48-S58.
- Boon, Mieke. 2012. "Understanding Scientific Practices: The Role of Robustness Notions." Pp. 289-316 in *Characterizing the Robustness of Science: After the Practice Turn in Philosophy of Science*. Edited by Léna Soler et al. Dordrecht: Springer.
- Bovens, Luc and Stephan Hartmann. 2003. *Bayesian Epistemology*. New York: Oxford University Press.
- Brush, Stephen G. 1968. "A History of Random Processes: I. From Brown to Perrin." *Archive for History of Exact Sciences* 5:1-36.
- Burian, Richard M. 1997. "Exploratory Experimentation and the Role of Histochemical Techniques in the work of Jean Brachet, 1938-1952." *History and Philosophy of the Life Sciences* 19:27-45.
- Campbell, Donald T., and Donald W. Fiske. 1959. "Convergent and Discriminant Validation by the Multitrait-multimethod Matrix." *Psychological Bulletin* 65:81-105.
- Cartwright, Nancy. 1983. *How the Laws of Physics Lie*. Oxford University Press.
- Cartwright, Nancy. 1991. "Replicability, Reproducibility and Robustness: Comments on Harry Collins." *History of Political Economy* 23:143-155.
- Chalmers, Alan. 2003. "The Theory-Dependence of the Use of Instruments in Science." *Philosophy of Science* 70:493-509.
- Chalmers, Alan. 2009. *The Scientist's Atom and the Philosopher's Stone: How Science Succeeded and Philosophy Failed to Gain Knowledge of Atoms*, *Boston Studies in the Philosophy of Science* 279. Springer.
- Chang, Hasok. 2012. "Beyond Case Studies: History as Philosophy." Pp. 109-124 in *Integrating History and Philosophy of Science*. Edited by Seymour Mauskopf and Tad Schmaltz. Dordrecht: Springer.

- Claveau, François. 2013. "The Independence Condition in the Variety-of-Evidence Thesis." *Philosophy of Science* 80:94-118.
- Claveau, François and Olivier Grenier. 2019. "The variety-of-evidence thesis: a Bayesian exploration of its surprising failures." *Synthese* 196:3001-3028.
- Collins, Harry. 1985. *Changing Order: Replication and Induction in Scientific Practice*. Chicago: The University of Chicago Press.
- Coko, Klodian. 2015. *The Structure and Epistemic Import of Empirical Multiple Determination in Scientific Practice*. Dissertation. Indiana University.
- Coko, Klodian. 2019. "Towards a Mutually Beneficial Integration of History and Philosophy of Science: The case of Jean Perrin." Pp. 186-209 in *The Past, Present, and Future of Integrated History and Philosophy of Science*. Edited by Emily Herring et al. New York and London: Routledge.
- Coko, Klodian. forthcoming. "Jean Perrin and the Philosophers' Stories: The role of multiple determination in determining Avogadro's number." *HOPOS: The Journal of the International Society for the History of Philosophy of Science*.
- Culp, Sylvia. 1994. "Defending Robustness: The Bacterial Mesosome as a Test Case." *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association* 46-57.
- Culp, Sylvia. 1995. "Objectivity in Experimental Inquiry: Breaking Data-Technique Circles." *Philosophy of Science* 62:438-458.
- Duhem, Pierre. 1894. "Quelques réflexions au sujet de la physique expérimentale." *Revue des questions scientifiques* 36:179-229.
- Franklin, Allan. 1986. *The Neglect of Experiment*. Cambridge University Press.
- Franklin, Allan. 1998. "Experiments in Physics." In *Stanford Encyclopedia of Philosophy*, <http://plato.stanford.edu/entries/physics-experiment/>.
- Franklin, Allan. 2002. *Selectivity and Discord: Two Problems of Experiment*. Pittsburgh: University of Pittsburgh Press.

- Gandenberger, Stephen. 2010. "Producing a Robust Body of Data with a Single Technique." *Philosophy of Science* 77:381-399.
- Hacking, Ian. 1981. "Do We See Through a Microscope?" *Pacific Philosophical Quarterly* 63:305–322.
- Hacking, Ian. 1983. *Representing and Intervening: Introductory Topics in the Philosophy of Natural Science*. Cambridge University Press.
- Hacking, Ian. 1992. "The Self-Vindication of the Laboratory Sciences." Pp. 29-64 in *Science as Practice and Culture*. Edited by Andrew Pickering. Chicago: University of Chicago Press.
- Hey, Spencer P. 2015. "Robust and Discordant Evidence: Methodological Lessons from Clinical Research." *Philosophy of Science* 82:55-75.
- Hon, Giora. 1989. "Towards a Typology of Experimental Errors: An Epistemological View." *Studies in History and Philosophy of Science* 20:469-504.
- Hudson, Robert G. 1999. "Mesosomes: A Study in the Nature of Experimental Reasoning." *Philosophy of Science* 66:289-309.
- Hudson, Robert G. 2003. "Mesosomes and Scientific Methodology." *History and Philosophy of the Life Sciences* 25:167-191.
- Hudson, Robert G. 2014. *Seeing Things: The Philosophy of Reliable Observation*. Oxford: Oxford University Press.
- Kosso, Peter. 1988. "Dimensions of Observability." *The British Journal for the Philosophy of Science* 39:449-467.
- Kosso, Peter. 1989. "Science and Objectivity." *The Journal of Philosophy* 86:245-257.
- Krömer, Ralf. 2012. "Are we Still Babylonians? The Structure of the Foundations of Mathematics from a Wimsattian Perspective." Pp. 189-206 in *Characterizing the Robustness of Science: After the Practice Turn in Philosophy of Science*. Edited by Léna Soler et al. Dordrecht: Springer.
- Kuorikoski, Jaakko and Caterina Marchionni. 2016. "Evidential Diversity and the Triangulation of Phenomena." *Philosophy of Science* 83:227-247.

- Laudan, Larry and Jarrett Leplin. 1991. "Empirical Equivalence and Underdetermination." *The Journal of Philosophy* 88:449-472.
- Nederbragt, Hubertus. 2003. "Strategies to improve the reliability of a theory: the experiment of bacterial invasion into cultured epithelial cells." *Studies in History and Philosophy of Biological and Biomedical Sciences* 34:593-614.
- Nye, Mary J. 1972. *Molecular Reality: A Perspective on the Scientific Work of Jean Perrin*. New York: American Elsevier Company.
- Perrin, Jean. 1909. "Mouvement Brownien et Réalité Moléculaire." *Annales de Chimie et de Physique* 18:1-114.
- Perrin, Jean. 1912. "Les Preuves de la Réalité Moléculaire." Pp. 153-251 in *La Théorie du Rayonnement et les Quanta: Rapports et discussions de la Réunion tenue à Bruxelles, du 30 Octobre au 3 Novembre 1911*. Edited by Paul Langevin and Maurice de Broglie. Paris: Gauthiers-Villars.
- Perrin, Jean. 1913. *Les Atomes*. Paris: Libraire Félix Alcan.
- Pickering, Andrew. 1995. *The Mangle of Practice: Time, Agency, and Science*. Chicago: University of Chicago Press.
- Pickering, Andrew. 2012. "The Robustness of Science and the Dance of Agency." Pp. 317-28 in *Characterizing the Robustness of Science: After the Practice Turn in Philosophy of Science*. Edited by Léna Soler et al. Dordrecht: Springer.
- Pinnick, Cassandra and George Gale. 2000. "Philosophy of Science and History of Science: A Troubling Interaction." *Journal for General Philosophy of Science* 31:109-125.
- Pitt, Joseph C. 2001. "The Dilemma of Case-Studies: Toward a Heraclitian Philosophy of Science." *Perspectives on Science* 9:373-382.
- Psillos, Stathis. 1999. *Scientific Realism: How Science Tracks Truth*. Routledge.
- Rasmussen, Nicolas. 1993. "Facts, Artifacts, and Mesosomes: Practicing Epistemology with the Electron Microscope." *Studies in History and Philosophy of Science* 24:227-265.

- Rasmussen, Nicolas. 2001. "Evolving Scientific Epistemologies and the Artifacts of Empirical Philosophy of Science: A Reply Concerning Mesosomes." *Biology and Philosophy* 16:629-654.
- Schickore, Jutta. 2011. "More Thoughts on HPS: Another 20 Years Later." *Perspectives on Science* 19:453-481.
- Schickore, Jutta, and Klodian Coko. 2013. "Using Multiple Means of Determination." *International Studies in the Philosophy of Science* 27 (3):293-311.
- Schupbach, Jonah. 2018. "Robustness Analysis as Explanatory Reasoning." *The British Journal for the Philosophy of Science* 69:275-300.
- Soler, Léna. 2012a. "Introduction: The Solidity of Scientific Achievements: Structure of the Problem, Difficulties, Philosophical Implications." Pp. 1-60 in *Characterizing the Robustness of Science: After the Practice Turn in Philosophy of Science*. Edited by Léna Soler et al. Dordrecht: Springer.
- Soler, Léna. 2012b. "Robustness of Results and Robustness of Derivations: The Internal Architecture of a Solid Experimental Proof." Pp. 227-266 in *Characterizing the Robustness of Science: After the Practice Turn in Philosophy of Science*. Edited by Léna Soler et al. Dordrecht: Springer.
- Soler, Léna et al. eds. 2012. *Characterizing the Robustness of Science: After the Practice Turn in Philosophy of Science*. Dordrecht: Springer.
- Staley, Kent W. 2004. "Robust Evidence and Secure Evidence Claims." *Philosophy of Science* 71:467-488.
- Stegenga, Jacob. 2009. "Robustness, Discordance, and Relevance." *Philosophy of Science* 76:650-661.
- Stegenga, Jacob. 2012. "Rerum Concordia Discors: Robustness and Discordant Multimodal Evidence." Pp. 207-226 in *Characterizing the Robustness of Science: After the Practice Turn in Philosophy of Science*. Edited by Léna Soler et al. Dordrecht: Springer.

- Stegenga, Jacob and Tarun Menon. 2017. "Robustness and Independent Evidence." *Philosophy of Science* 84: 414-435.
- Trizio, Emiliano. 2012. "Achieving Robustness to Confirm Controversial Hypotheses: A Case Study in Experimental Biology." Pp. 105-120 in *Characterizing the Robustness of Science: After the Practice Turn in Philosophy of Science*. Edited by Léna Soler et al. Dordrecht: Springer.
- Weber, Marcel. 2005. *Philosophy of Experimental Biology*. Cambridge: Cambridge University Press.
- Wimsatt, William C. 1981. "Robustness, Reliability, and Overdetermination." Pp. 124-163 in *Scientific Inquiry and the Social Sciences*. Edited by M.B. Brewer and B.E. Collins. San Francisco, CA: Jossey-Bass.
- Wimsatt, William C. 2007. *Re-engineering Philosophy for Limited Beings*. Cambridge, MA: Harvard University Press.
- Woodward, Jim. 2006. "Some Varieties of Robustness." *Journal of Economic Methodology* 13 (2):219-240.
- Zammito, John H. 2004. *A Nice Derangement of Epistemes: Post Positivism in the Study of Science from Quine to Latour*. Chicago: The University of Chicago Press.

