

This item is the archived peer-reviewed author-version of:

The role of orientation experiments in discovering mechanisms

Reference:

Gervais Raoul, Weber Erik.- The role of orientation experiments in discovering mechanisms
Studies in history and philosophy of science: part A - ISSN 0039-3681 - 54(2015), p. 46-55
Full text (Publishers DOI): <http://dx.doi.org/doi:10.1016/j.shpsa.2015.08.015>

THE ROLE OF ORIENTATION EXPERIMENTS IN DISCOVERING MECHANISMS

1. Introduction

A lot has been written recently about the way scientists discover mechanisms, in particular, about the *reasoning* employed that leads to such discoveries. Underlying such discussions is general tendency to move away from the Popperian focus on the logic of discovery and justification, and towards a view of discovery as a problem-solving activity. According to Darden “Philosophers should move beyond talk of the (lack of) a logic of discovery and a logic of justification to study reasoning strategies for generation, evaluation, and revision in the discovery of mechanisms” (2009 p. 54).

Two dimensions are of importance to this project. First, there is the *epistemological order of the events* in providing mechanistic explanations. Here, Darden and others have developed a quite elaborate framework over the years, according to which this process can be partitioned into phases such as identifying the explanandum-phenomenon, the generation, the evaluation, and the revision of mechanistic hypotheses (Darden 2009). Second, there are the *types of reasoning strategies and experiments* that are carried out during these stages of the investigation. Again, many different types of experiments and reasoning-patterns have been described – of particular importance here are inter-level intervention experiments, as described by Craver (2007). It would seem that in order to make progress, philosophers should concentrate on increasing this stock of strategies for discovery, and show how these strategies fit into the epistemological order (e.g. whether a strategy belongs primarily to the phase of revision, or is particularly useful when attempting to identify the explanandum phenomenon). Regarding to the question as to where we should look, Darden suggests the history of science as a particularly promising source (2002 p. S364).

In this article, we will follow this suggestion. In particular, we will argue that the inter-level experiments as envisaged by Craver come with certain presuppositions regarding the knowledge of the mechanism we must have before we can execute them. However, these presuppositions are not always met. Sometimes, the mechanism is largely unknown, or different mechanistic hypotheses might be competing. Instead of inter-level experiments, in these circumstances scientists perform a different kind of experiment that, to our knowledge, so has not been recognized as such in the literature. We call these experiments *orientation experiments*. In section 3, we will spell out in detail the different structural features orientation experiments have. For now, let us confine ourselves to a rough characterization. Orientation experiments are a special type of intervention experiments used to provide evidence for or against a qualitative characterization of a mechanism. They do not go ‘in depth’, actively identifying entities and activities of the mechanism, but instead remain on the level of the explanandum and the environment. As such, they are epistemologically prior to inter-level experiments, which require more detailed knowledge.

We will argue that orientation experiments serve a number of purposes. In particular, they are important in the discovery of mechanisms, because they guide and constrain future, more detailed investigation into the mechanism, and exclude alternative mechanisms. As such, they can be used to settle the competition between rival hypothetical types of

mechanisms. To give a very brief example (discussed in more detail below), if we wonder how ants are able to find the shortest way between their nests and a food source, an orientation experiment can tell us that the mechanism is chemical rather than visual. Though abstract, such a characterization is by no means trivial, since it has implications for future research.

We will argue that orientation experiments are typical of the early stages of investigation, when an explanandum-phenomenon has only recently been addressed. Thus, there is a stage in between the identification of the explanandum-phenomenon and the discovery of the mechanism, in which orientation experiments are devised and performed. We will call this the *orientation-phase*. To illustrate this, we will consider three case studies of different era's: the capacity of eels to produce numbing sensations (18th century), cadaverous poison as famously discovered by Semmelweis (19th century), and the capacity of pigeons to home (20th century). Besides illustrating the structural features of orientation experiments, the case study also help us to flesh out the different goals these experiments can serve – goals that set them apart from inter-level experiments. Of course, we are not against inter-level experiments: the goals of orientation experiments are complementary to those of inter-level experiments. Ultimately, we defend a complementary view, according to which orientation experiments pave the way for the type of inter-level explanations Craver and others have in mind.

Let us conclude with an overview. In section 2, we will first clarify some terminology and sketch the state of the art in the literature about the discovery of mechanisms, both with respect to the different phases of this research and the experimental strategies used during these phases. In particular, we will focus on Craver's (2007) inter-level experiments. In section 3, we will argue that these experimental techniques presuppose knowledge that is not always present, and that in fact, in between the identification of the explanandum-phenomenon and the constructing of detailed models of the mechanism, lies a phase which we dub the orientation phase, during which what we call orientation experiments are carried out. We spend the remainder of section 3 characterizing the structural features of this type of experiment. In sections 4 to 6 respectively, we illustrate the points developed in section 3, and flesh out the different goals orientation experiments serve, by considering the case studies mentioned above.

2. The literature on mechanisms, their discovery, and inter-level experiments

2.1 Some assumptions

Let us start by explicating some of the assumptions we will make throughout this paper. First, we assume that scientists in the biology (and more generally the life sciences) are in the business of (among other things) explaining the capacity of systems to perform some kind of activity; i.e. they try to answer questions along the lines of "How does system *S* perform capacity *C*?" or: "How do systems $S_1...S_n$ perform *C*?". For example, they might want to explain how an individual organism procreates, or they might want to explain how a species procreates. Second, we make the ontological assumption that there is always *some* mechanism responsible for a given activity, even if we have no or very little knowledge about it. To put it concisely: there is no activity of a system without the presence of some sort of mechanism. Third, we do not deny that mechanistic explanations are important – the

strategies we will describe are complementary to mechanistic explanations, not alternatives to them. Finally, we take the same complementary position regarding inter-level experiments (see the next section).

2.2 Mechanisms, organization and explanation

A decade and a half have passed since Machamer, Darden and Craver's seminal article (2000) sparking the debate about mechanistic explanations, so by now, most readers will be familiar with the most important concepts involved. Therefore, we shall not rehearse this debate in detail, but simply limit ourselves to giving some working definitions of the concepts involved. We adopt these definitions from Illari & Williamson 2012:

A mechanism for a phenomenon consists of entities and activities organized in such a way that they are responsible for the phenomenon (p. 123)

Illari and Williamson present this definition as an expression of the core consensus among philosophers in the mechanistic tradition. Many slightly different characterizations have been given in the last two decades¹ Having compared these different characterizations, Illari & Williamson argue that the definition above is the best one. We think they have done a very good job, so we adopt their definition. In this definition, the term *entities* refers to "the bits and pieces of the mechanism" (p. 125), while the term *activities* refers to "what those bits and pieces do" (ibid.). About *organisation* they write:

Most generally, organization is whatever relations between the entities and activities discovered produce the phenomenon of interest [.] (Ibid., p. 128)

...

[I]n our understanding of organization as when activities and entities each do something and do something together to produce the phenomenon, *whatever* relations amongst the activities and entities produces the phenomenon is the relevant organization. (Ibid., p. 131)

Organisation then, is about the relations between the entities and activities.

Illari & Williamson characterise mechanistic explanations as follows:

All mechanistic explanations begin with (a) the identification of a phenomenon or some phenomena to be explained, (b) proceed by decomposition into entities and activities relevant to the phenomenon, and (c) give the organization of entities and activities by which they produce the phenomenon (Ibid p. 123).

¹ For instance: "A mechanism is a structure performing a function in virtue of its component parts, component operations, and their organization. The orchestrated functioning of the mechanism is responsible for one or more phenomena" (Bechtel & Abrahamsen 2005, p. 423); and "[M]echanisms are entities and activities organized such that they exhibit the explanandum phenomenon" (Craver 2007, p. 6, italics removed).

This is a dynamic characterisation, which tells us how mechanistic explanations are construed. Illari & Williamson do not give a (static) definition of the end product. The following definition is complementary to what they say:

A mechanistic explanation for a phenomenon is a description of the mechanism that produces this phenomenon.

Often, such a description is called a *model* of the mechanism. In any case, this concludes the conceptual apparatus with respect to mechanisms and their explanations.

2.3 Discovering mechanisms

How are mechanisms discovered? As we have said in the introduction, over the years Lindley Darden has developed a rich picture of this process (2009). This picture consists of two dimensions: first, the process of discovery is partitioned up into several phases. Second, different strategies are devised that apply to these phases.

First, there is the identification of the explanandum-phenomenon. During this stage, scientists try to give an accurate description of the activity they are investigating. This description of the explanandum-phenomenon, together with the overall goal of constructing a model of the responsible mechanism, guides the process of discovery; i.e. it puts constraints upon the second stage, that of mechanism schema generation. Ideally, the constraints are such that the entire hypothesis stage can be constructed, but in practice, this is almost never the case, so that evaluation and revision (stages 3 and 4) are needed.²

Next comes evaluation. In evaluating a hypothesized mechanism, one looks for failures, such as incompleteness, where a model contains a black box that is yet to be opened (the entity responsible for some sub-capacity has yet to be specified), or incorrectness, where a hypothesized component does not fit. Here, the kind of inter-level experimental reasoning we will consider in the next subsection applies: one tests a proposed mechanism by means of top-down and bottom up intervention-experiments. Finally, in revision, one tries to resolve the anomalies found during evaluation, distinguishing relevant from irrelevant components. Again, inter-level intervention experiments apply to this stage of the discovery process.

Two issues should be noted here. First, this is not a linear, but a reciprocal or cyclic process. For example, evaluation and generation can be performed simultaneously, and findings in later phases of discovery may prompt researchers to go back one or more steps in the process. Second, the list of reasoning (cf. note 2) and experimental strategies is not exhaustive (Darden 2009 p. 306): it may be that there are additional strategies for revision or

² According to Darden (2009), there are several strategies for hypothesis generation. There is *schema instantiation*, where one provides an abstract type of mechanism-schema that can be specified to apply to particular cases. In generating these schemata, one might reason by analogy (e.g. finding a mechanism in the history of science and then dropping specific details) or sketch hypothetical roles for the components of the mechanism. Another strategy is *modular subassembly*. Here, one reasons about groups of mechanism components, hypothesizing that a mechanism consists of known modules or types of modules. These modules are then put together to construct a hypothetical mechanism. Finally, at a finer grain, one employs *forward/backward chaining*: one reasons about a part of a mechanism on the basis of knowledge of or hypotheses about other parts of the mechanism. In forward chaining, one reasons from the early stages of a mechanism and conjectures about what entities and activities are likely relevant during later stages, in backward chaining one reasons the other way round.

schema generation. Part of the work of philosophers of science is to identify such additional strategies.

2.4 Inter-level experiments

Not all entities or activities in a mechanism are relevant. Following one of Craver's analogies, in trying to explain why a car runs, the hubcaps, mud-flaps and the windshield, though clearly part of the car, are not relevant (2007 p. 140). So how do we sort relevant from irrelevant parts? This is where inter-level experiments come in. They guide the process of discovery by establishing the constitutive relevance of the entities and activities that are contained in the explanans.

The crucial question now is: how can we show that an entity X and its activity ϕ indeed are components of the mechanism of the ψ -ing of an S ? Craver's answer is labelled the *mutual manipulability account*³. According to this account, components are relevant to the behaviour of a mechanism when one can intervene upon that behaviour by intervening upon that component, and vice versa. Thus, the mutual manipulability account gives us two directions in which to perform experiments: bottom-up and top-down. As the experiment includes variables both at the level of the mechanism as a whole, and at the level of components of that mechanism, they are inter-level.

Besides direction however, there are two more dimensions to these interlevel experiments: one can intervene by diminishing, disabling, or destroying something (inhibition), or augmenting or enhancing it (excitation). Together with direction, this gives us a variety of experiments, of which shall discuss three.

First, whether a component is relevant to the behaviour of the mechanism as a whole can be established by performing two types of experiments. First, there are *interference experiments*, which are inhibitory bottom-up experiments:

In interference experiments, one intervenes to diminish, disable, or destroy some putative component in a lower-level mechanism and then detects the results of this intervention for the *explanandum phenomenon*. The assumption is that if X 's ϕ -ing is a component in S 's ψ -ing, then removing X or preventing it from ϕ -ing should have some effect on S 's ability to ψ .

...

Lesion experiments, for example, are interference experiments in which something intervenes to remove a portion of the brain and one then detects the effects of the lesion on task performance. (2007, p. 147)

Second, there are *stimulation experiments*, which are excitatory bottom-up experiments:

In stimulation experiments, one intervenes to excite or intensify some component in a mechanism and then detects the effects of that intervention on the *explanandum phenomenon*. The assumption is that if X 's ϕ -ing is a component in S 's ψ -ing, then one should be able to change or produce S 's ψ -ing by stimulating X .

...

The classic example of stimulation experiments is Gustav Fritsch and Eduard Hitzig's (1870) work on the motor cortex (see Bechtel forthcoming). Fritsch and Hitzig performed a series of experiments on dogs in which they delivered low-grade electrical stimuli to a cortical area now known as the motor strip (see Bechtel forthcoming). Localized stimuli along this area produce regular and repeatable movements in specific muscles, including the legs, the tail, and the facial

³ For an up to date discussion of this account, see Leuridan 2012.

muscles. (2007, p. 149)

Finally, there are *activation experiments*, which are excitatory top-down experiments. Here are Craver's characterization and example:

In activation experiments, one intervenes to activate, trigger, or augment the *explanandum phenomenon* and then detects the properties or activities of one or more putative components of its mechanism. ... The basic assumption behind activation experiments is that if X is a component in S's ψ -ing, then there should be some difference in X depending on whether S is ψ -ing or not.

...

There are several common varieties of activation experiment at all levels in neuroscience. In PET and fMRI studies, one activates a cognitive system by engaging the experimental subject in some task while monitoring the brain for markers of activity, such as blood flow or changes in oxygenation. (2007, p. 151)

The link between the experiments that Craver describes (and for which he uses the general label "interlevel experiments") and mechanistic explanations of capacities is straightforward: scientists who want to give mechanistic explanations of capacities must do bottom-up and top-down experiments.

3. Orientation experiments and the orientation-phase

According to the literature then, the discovery of mechanisms consists of identifying the phenomenon, generating mechanism sketches and then evaluating and revising these sketches by means of inter-level experiments. While we think that all these phases and experimental techniques are valuable, we feel that there is something missing from this picture.

In particular, the kind of inter-level experiments Craver describes, presuppose a detailed knowledge of the mechanism that is not always present. For example, to perform interference and stimulation experiments, one must already have identified and localized at least one component part of the mechanism. Activation experiments presuppose that we know the location of at least one part of the mechanism, so that we can measure or observe it in order to detect any changes to it as a result of our intervention upon the behaviour of the mechanism as a whole. In general, such inter-level intervention experiments require that we have localized at least some part of the underlying mechanism – one cannot intervene upon parts that are not yet localized.

Yet when an explanandum-phenomenon is first identified, clearly such information is not at hand. For example, when we have observed that ants have the capacity to find the shortest way between their nest and food sources, then we can infer (given our ontological assumption of section 2.1) that there must be a mechanism. Suppose we propose a mechanism sketch according to which ants leave chemical trails on their way back from a successful foraging mission. This proposed mechanism sketch leads to certain expectations. For example, from the mere proposal that the mechanism involves the detecting of chemical rather than visual signals, one would expect the capacity to be present in the dark. Furthermore, there must be some organ that can detect chemicals, and one that can leave the trails. All this information constrains the search for component parts of the mechanism, but we are still far from being able to perform any inter-level intervention experiments. For example, in this stage it is impossible to inhibit the production of chemical markers and see

whether this affects the ants' capacity to find the shortest way between their nest and a food source, because we have not localized the component parts. We must first establish in broad lines the qualitative character of the mechanism, in this case that it is chemical rather than visual.

We do this by performing orientation experiments. In the case of the ant, we might first establish whether the capacity under investigation is maintained in the absence of light. If it is, then this is evidence against a visual mechanism. There is also a good, economical reason for this: if one did not devise and perform orientation experiments to rule out certain candidate mechanism sketches, one could waste considerable time and effort trying to localize the parts of a visual trailing mechanism that simply is not there. If a number of simple experiments can exclude possible mechanism sketches, and raise the probability of others, then surely these experiments are worth performing.

Thus, in the process of discovering mechanisms, there seems to be a phase between identifying a phenomenon on the one hand, and the discovery, evaluation and revision of mechanistic explanations on the other hand. Let us call this phase the *orientation-phase*. Schematically, during the orientation-phase, the following happens:

- 1) One or more mechanism sketches about the qualitative character of the mechanism responsible for the explanandum-phenomenon are proposed.
- 2) To gather evidence for or against these mechanism sketches, one devises and performs orientation experiments.

As we will show in detail below, the orientation experiments have the following structural properties:

- a) Like inter-level experiments, they require interventions.
- b) Unlike inter-level experiments, they do not require localization of parts of the mechanism.

Since intervention requires that we have already localized the variable we are about to intervene upon, from b) it follows that the relata of the intervention-relation in orientation experiments cannot be the same as in inter-level experiments. This brings us to another structural feature that sets orientation experiments apart from their inter-level counterparts:

- c) Unlike inter-level experiments, they are intra-level, because the relata of these expected causal relations are the explanandum-phenomenon and its environment.

We believe that orientation experiments are an important tool in the experimental toolkit of scientists. However, by themselves, the structural features mentioned here do not reveal anything about the *use* of such experiments. In what follows, we will show that not only do orientation experiments differ from inter-level experiments on account of their structural features, they also serve different *goals*: scientists use them to attain different types of knowledge about the mechanisms responsible for the behavior they are trying to explain. Remember that one important goal of inter-level experiments is to sort out relevant from irrelevant parts of the mechanism. Again, this presupposes knowledge that we might not have. Orientation experiments on the other hand are useful in situations in which we know

(next to) nothing about the mechanism, which are often (but not exclusively) the crucial early stages of an investigation into a phenomenon, namely the orientation-phase. As such, they act as a prelude to inter-level experiments, effectively constraining the range of possible interference, stimulation, and activation experiments.

In the remainder of this paper we will consider three case studies in which experiments were carried out during the orientation phase to discover the mechanisms behind a particular capacity. For each of these case studies, we will: i) show that the experiments involved do indeed exhibit structural features a) to c), and ii) explicate the various goals these experiments were meant to serve, i.e. the different types of information they provide. It will emerge that unlike the structural features, which are present in every orientation experiment, the goals vary somewhat from context to context.

4. First case study: electric eels

For our first case study⁴, we will consider the investigation in the 17th and 18th centuries into the capacity of certain ‘eels’⁵, newly discovered in South-America at the time, to produce numbing sensations in the hands and arms of anyone who touches them.

4.1 Background

Fish capable of delivering numbing sensations were known to Europeans before, even in antiquity, namely sea torpedoes and the African silurus (catfish). We now know that these fish are capable of delivering electric shocks up to 50 and 300 volts respectively. However, commentators in antiquity, such as Galen, attributed the power of sea torpedoes to numb the hand to poison.⁶ The discovery of a new type of ‘tremble fish’ in exotic places like the river systems of South-America, together with the fact that the numbing sensation produced by this new fish was much more powerful (up to 600 volts, as we know now) than those produced by the torpedo and the silurus, led to a renewed interest.

The first descriptions of the eels were made in the 1500s by Spanish administrators and officials, but in the 17th century, more personal accounts appeared. In the 1670s, the French astronomer Jean Richer encountered the eels while on a visit to Cayenne:

I was even more surprised to see a three to four foot fish, similar to an eel [...] that by simple touch with a finger or the tip of a stick, so numbs the arm and that part of the body closest to it that one remains about 15 minutes without being able to move [...] I was witness to this effect and I felt it, having touched the fish with my finger one day when I met up with some savages (Richer 1679)

⁴ The following section draws upon Koehler, Finger & Piccolino 2009.

⁵ Although they were called eels when they were first discovered (their shapes being similar to the eels known to Europeans), the *Electrophorus Electricus* has since been classified as a member of the knife fish group (gymnotiformes) rather than the eels; hence the parentheses. As it makes no difference to our arguments, we will ignore this taxonomical detail and continue to call them eels.

⁶ It is not always clear which of the two species (the torpedo or the silurus) the commentators in antiquity refer to. The Greek word *νάρκη* (which also means numbness) was used to refer to both, as was the Latin word *torpedo*, so that confusion between the two species was very common. Galen was likely referring to the torpedo rather than the silurus; in any case, what matters for our purposes is the general idea that the numbness was somehow caused by poison.

Thus, the first step is made in this report, namely the identification and description of the explanandum-phenomenon. The explanation-seeking question that needs to be answered is: “How do eels produce a temporary numbing sensation in the hand of anyone touching it?”

4.2 *The orientation phase*

Now that the explanandum-phenomenon has been identified and described, research enters the orientation-phase, as scientist propose qualitative characterizations of the mechanism to explain the capacity of the eel to produce numbing sensations. Here are some characterizations that were considered in the 17th and 18th century:

- 1) **Poison.** As we have seen, the earlier known tremble fish were already described and studied in antiquity. One proposal by the ancient commentators was that touching the torpedo somehow releases poison. Of course, this proposal can also be made with respect to the eels of South America.
- 2) **Animal Spirits.** In the tradition of Galen, some believed that animal spirits flowed through hollow nerves, inflating the fibres of one or more of the eel’s muscles. The numbing sensation then is caused a very rapid contraction of a muscle.
- 3) **Biochemistry.** Another idea was that the sensation was indeed caused by a rapid stiffening of muscles, but that this was due to a chemical process akin to fermentation, rather than animal spirits.
- 4) **Corpuscles.** According to Stefano Lorenzini (1678), the numbing effect was due to the violent mechanical emission of a multitude of corpuscles so small that they could penetrate the skin and interfere with the nerve conduction of the hand. The guiding idea here is that of a coiled up spring which is suddenly unleashed.
- 5) **Percussion/concussion.** Giovanni Borelli (1680-1681) proposed a similar mechanism based upon the force of a coiled up spring, but in his mechanism the corpuscles did not penetrate the skin, but rather cause numbness by percussive/concussive action, hitting the hand at high speed.
- 6) **Shape.** René-Antoine de Réaumur held a lecture in 1714 about the torpedo’s, claiming that the torpedo’s upper surface, normally flat or concave in rest, becomes convex (curves outward) very rapidly. The change in surface is due to elastic actions of minute muscular fibers in specialized organs, which causes numbness in the hand by nerve compression. This hypothesized mechanism was applied to the South-American eels by others (e.g. Allemand 1756).

Notice that although these ideas are vague, they are proposed mechanisms by the standards set in section 2.2: they include reference to entities (poison, fibers, muscles, corpuscles etc.), activities (releasing, inflating, stiffening, penetrating the skin, etc) and give a rudimentary kind of organization.

Of course, not all proposals were studied with the same vigour (for example, the second proposed mechanism would presumably not have been widely studied, as by the 17th century, the idea of animal spirits was already on its way out). Some are compatible or show a partial overlap, some are mutually exclusive.

With hindsight, we know that none of these six hypotheses is correct, and that the mechanism in fact employs electricity. But how did the electrical hypothesis come to be proposed?

When the reports about the fish made their way back across the ocean, they were read by a community of scholars in Europe that was increasingly interested in electricity. In particular, the invention of the Leyden Jar, an early kind of capacitor that was small enough to carry around and could be used to produce small shocks in a controlled way, greatly enhanced the possibility to perform experiments with electricity. Nevertheless, between the community of European researchers experimenting with electricity, and the naturalists describing the wildlife of South America, there was only contact by letter, and it was very difficult to get hold of a prepared specimen of the eel, let alone a live one. The point is that the electrical hypothesis is thought of when someone has felt *both* the shocking power of the Leyden Jar, *and* the numbing sensation caused by the eel. In 1754, in a letter to his friend, the physicist Jean Allamand, a Dutch official working in Guyana named Laurens Storm van 's Gravesande wrote (as quoted in Allamand 1756):

The experiment was done with an eel called a tremble fish ... It produces the same effect as the electricity that I felt with you, while holding in a hand a bottle [the Leyden Jar] that was connected to an electrified tube by an iron wire.

As far as hypothesis-generation goes then, it seems that sometimes mechanisms are proposed on the basis of very subjective evidence indeed, in this case the perceived similarity of the felt quality of two numbing sensations.

In any case, the electrical hypothesis soon had its supporters, among them Allamand and the anatomist John Hunter, and was competing with the other six hypotheses. The question now is: which mechanism is correct? Can we rule out any of the candidates? Note that at this stage in the investigation, the kind of inter-level experiments described in section 2.4 are simply impossible. For example, it is impossible to inhibit corpuscles from moving, or the animal spirits from flowing, and then observing whether the power induce numbing sensations diminishes or vanishes; nor is it possible to somehow augment this power and observe changes in the fermentation processes or the electric currents. Such experiments would presuppose that we already know what type of mechanism we are dealing with (which is precisely what is at stake) and that we have already localized at least one part as we cannot intervene upon hypothetical parts.

One way to gather evidence for and against hypothesized mechanisms that *is* available at this stage is to perform orientation experiments. That is, from a hypothesis about the qualitative character of the mechanism, we infer that certain causal relation will be present or absent.

In the 1750s, another eminent Dutch naturalist, Laurens Theodore Gronov, became interested in the proposed electrical hypothesis. Because it was still very difficult to obtain live specimens in Europe, he wrote a letter to van 's Gravesande, still in the colonies, asking him a total of 25 questions about the eel (Gronov 1758). In particular, he asked van 's Gravesande to perform a number of orientation experiments. For example, Gronov reasoned that if the mechanism is electric, then:

- a) Touching the eel with a wooden rod would result in a numbing sensation that is less severe compared to direct contact,
- b) If you insert a metal piece into the tip of a wooden rod and touch the eel with the tip, the sensation would be more severe than with a plain wooden rod,
- c) Touching the eel with an iron rod wrapped in dry cloth should produce no sensation, but if the cloth is made wet it should again be felt,

- d) If five persons hold hands while the first touched the eel, the fifth would be numbed less severely.

On the other hand, if the mechanism is not electric, then these predictions will not come true. As it happened, van 's Gravesande successfully carried out the experiments. This counted as strong evidence for the electric hypothesis.⁷

4.3 Orientation experiments versus inter-level experiments: structural features

It is important here to compare these orientation experiments with Craver's inter-level experiments. One similarity stands out: both types of experiments require intervention (cf. structural feature (a) in Section 3). In inter-level experiments, one intervenes on either the explanandum-phenomenon, or on one or more of the components of the mechanism responsible for the explanandum phenomenon. As we have already said, one can only measure change in, or intervene on, real entities or activities that are already successfully located, not hypothetical ones. This means that for inter-level experiments, the component intervened upon (in the case of bottom-up experiments) or measured (in the case of top-down experiments), are real and have already been located. In contrast, orientation experiments do not intervene on or measure a component of the mechanism, because the mechanism is still hypothetical, and components have not yet been localized (structural feature (b)). Instead, the intervention is on the explanandum-phenomenon, and something else on the same level of description. In the case of experiment a) for example, one intervenes on the way the eel is touched, and measure a change in the numbing sensation. Both variables of this intervention reside on the same level, and in contrast to Craver's experiments, orientation experiments are horizontal. (structural feature (c))

Of course, orientation experiments are not foolproof (e.g. in the case of experiment b, even in the absence of electricity, the numbing might still be more severe because of some hidden factor), but this is countered by the usual methods to eliminate variables.

4.4 Orientation experiments versus inter-level experiments: goals

As should be evident, the primary goal of the orientation experiments devised by Gronov, and performed by 's Gravesande, was not to sort relevant from irrelevant parts of the mechanism, as in the case of Craver's inter-level experiments. Rather, the goal is to arrive at a rough, qualitative characterization of the mechanism, to constrain future research. As we will show in our next two case studies, this is a goal that orientation experiments are always meant to achieve; other goals are present in some experiments, but not in others.

In time, scientists became increasingly convinced of the electric hypothesis, as more evidence was gathered for it, and other hypotheses were abandoned. The orientation phase gave way to the next phase, where anatomical studies and inter-level experiments are used to give a detailed mechanistic explanation of the explanandum-phenomenon. For example, John Hunter started to dissect eels and produce illustrations of the electrical organs. Here, the fact that one type of mechanism is agreed upon also helps to guide research: if the

⁷ At the time, one serious problem for the hypothesis was that unlike the Leyden jar, touching the eel did not produce a spark, which was thought by many to be a hallmark of electricity. When in 1770s, scientists finally managed to draw a spark from an eel at the moment of discharge, this helped to turn many skeptics into supporters of the electrical hypothesis (Piccolino & Bresadola 2002).

mechanism is electrical, one looks specifically for things (organs) which can generate, store and discharge electricity. Thus, there is no sharp boundary between the orientation phase and the inter-level experimentation phase.

Summarizing, the orientation phase of the research in the capacity of South American eels to produce numbing sensations involved generating numerous hypotheses about the character of the underlying mechanism. With the help of orientation experiments, evidence for and against these hypotheses can be gathered. Though very general, the information that the mechanism is electric rather than e.g. spring-like, has important implications for further research.

5. Second case study: puerperal fever (a.k.a. childbed fever)

Our second case is from the 19th century: the work of Ignaz Semmelweis at the Vienna Maternity Hospital, which was made famous by Carl Hempel in his 1966 book *The Philosophy of Natural Science*. Hempel himself used the case to illustrate the H-D method, while we use it for a different aim. Our account is mainly based on Loudon 1992. While Hempel uses the popular term “childbed fever”, the standard (medical and historical) term (used by Loudon and many others) is “puerperal fever”. We use the latter term. This case study will be used as a further illustration of the structural features of orientation experiments, but also to argue that such experiments can serve additional goals, apart from the general goal of providing a qualitative sketch of the mechanism and constraining future research can serve.

5.1 Background

What is puerperal fever? Loudon quotes descriptions Charles Meigs (professor at Jefferson Medical College in Philadelphia from 1841 to 1861):

[T]here is no occasion to be surprised or astonished when, after having left your patient at ten o'clock in the evening, comfortable and apparently with any untoward symptom or accident, you find here, at six in the morning, a prey to the most unspeakable disorders of innervation, the respiration, and the circulation. (Meigs 1848 p. 597; quoted from Loudon 1992, p. 54)

Meigs reports to have seen patients ...

... who not only suffer intolerable pain, but in whose minds that pain appeared to excite the most unspeakable terror. (1848, p. 596; quoted from Loudon 1992, p. 56)

Loudon also quotes William Leishman (professor of midwifery at the University of Glasgow from 1868 to 1894):

The belly swells further and becomes tense, with great aggravation of the suffering, so that the patient can now no longer bear even the pressure of the bedclothes [...] (Leishman 1875, pp. 779-780; quoted from Loudon 1992, p. 56)

In any case, Semmelweis's own investigation starts with the observation of a difference between the first and second clinic of the Vienna Maternity Hospital. In the first clinic, puerperal fever was much more common than in the second, so that one might propose the following causal claim:

Being assigned to the first clinic of the Vienna Maternity Hospital (as opposed to the second clinic) is a positive causal factor of dying of puerperal fever.

In other words, Semmelweis believed that, if all women would give birth in the first clinic there would be more cases of puerperal fever in the Vienna Maternity Hospital than if all women would give birth in the second clinic. His evidence for this claim is a combination of (a) methodologically relevant facts about the structure and procedures of the hospital and (b) empirical data. First, the structure of the hospital:

In 1833 the Maternity Hospital was divided into two clinics. Medical students and midwives attended both clinics. From 1840 the first clinic was reserved for the instruction of medical students and doctors, the second for midwives. Patients were admitted to the clinics on alternate days, producing, unintentionally, a system of random allocation. There was no selection of complicated cases for the first clinic and uncomplicated for the second. (Loudon 1992, p. 65)

This quote not only clarifies what is meant with the two clinics. It also clarifies that there was a randomisation procedure inherent in the hospital's operational procedures. Second, there are the empirical data. Here is a selection from the table in Loudon 1992 (p. 67) which gives the maternal mortality rates per 1000 births:

In 1841, per 1000 births 77.0 women died in the first clinic, versus 35.0/1000 in the second clinic.

In 1844 the mortality rates were: 82.0/1000 (first clinic) versus 23.0/1000 (second clinic).

In 1846 the mortality rates were 114.0/1000 (first clinic) versus 27.0/1000 (second clinic).

The average for the period 1841-1846 was: 98.0/1000 (first clinic) versus 39.9/1000 (second clinic).

Combined with the (unintentional) random design, these data provide good evidence for the causal claim above.

5.2 *The orientation phase*

Semmelweis' aim was to explain the causal relation in order to get rid of it. He considered several types of mechanisms. Before we go into that, it is useful to clarify what we know *now*. Because this knowledge was mostly developed near the end of the 19th century and the beginning of the 20th, it was unknown to Semmelweis; we use it to serve as contrast. Here are some relevant quotes from Loudon 1992:

Puerperal fever is an illness which results from infection of the uterus during or after delivery. (p. 53)
[F]or most practical purposes, it is correct to think of puerperal fever as a streptococcal disease. Epidemics of puerperal fever [...] where wholly streptococcal in origin. (pp. 77-78).
Once the infection had gained entry into the uterine cavity it could spread out into the pelvic tissue causing pelvic cellulitis, the pelvic veins leading to pelvic thrombophlebitis, the peritoneal cavity causing peritonitis. It could also spread into the bloodstream causing puerperal septicaemia. (p. 55)

Whereas a mechanistic explanation of the causal relation requires a detailed model, the

quote provides only a very rough sketch: it mentions some entities (organs, bacteria), and suggests some activities (infection, spreading out) and organisation.

We find none of this in Semmelweis. Instead, what he claims is that:

“Cadaverous particles” adhering to the hands of the doctors are transferred from the post-mortem room to the labour wards of the 1st maternity clinic.

He also uses the term “morbid matter”. The background of this claim was the following habit:

At the Vienna Maternity Hospital it was the custom for medical students to attend post-mortem examinations of women who died of puerperal fever before walking over to the labour wards where they undertook numerous vaginal examinations in labour as part of their routine training. ... [T]hey saw no reason to wash their hands, and, of course, rubber gloves were not worn. (Loudon 1992, pp. 66).

The midwives working in the second clinic did not have post-mortem examinations as part of their training, so there was a difference between the two clinics in this respect.

We will look at Semmelweis’ experiments which provided evidence for this in Section 5.3. What interests us here is that there is an important similarity between his views and the contemporary perspectives. According to current views, Semmelweis was wrong in several respects: there are no “particles” involved but living organisms (bacteria); and the bacteria can, in principle, come from various sources, not only from dead bodies (so “morbid” and “cadaverous” are inadequate labels in general, though they correctly point at where the bacteria came from in the particular environment of the Vienna Maternity Hospital). However, the mechanism he proposes is basically biochemical: decaying organic material is the crucial ingredient in Semmelweis’ view. To drive this point home, let us compare this with an idea that he considered but rejected:

Even religious practices did not escape attention. The hospital chapel was so located that when the priest was summoned to administer last rites in the second clinic, he could go directly to the room set aside for ill patients. On the other hand, when he was summoned to the first clinic he had to pass through five other rooms because the room containing ill patients was sixth in line from the chapel. According to accepted Catholic practice, when visiting the sick to administer last rites, the priest generally arrived in ornate vestments and was preceded by a sacristan who rang a bell.

...

One can imagine the impression that was created on the other patients when the priest came several times a day, each time accompanied by the clearly audible bell (Semmelweis 1983, p. 71)

The potential mechanism described in this quote is a *psychosomatic* one. We will discuss Semmelweis’ reasons for rejecting this possibility in Section 5.3. Here we focus on the consequences of believing in one type of mechanism rather than the other. If we believe that the mechanism is biochemical, this guides our attempts to unravel the mechanism in the following two ways:

(a) Combined with the type of examinations done in the labour wards, it gives us an hypothesis (which has to further investigated) about the part of body where the mechanism initiates: the uterus.

(b) Given the amount of blood in the uterus, it is plausible to conjecture that, if the cadaverous particles enter through the uterus (cf. the first hypothesis), they their way into the blood stream, leading to blood poisoning. This is a second hypothesis that comes up.

Both hypotheses can be further explored by means of inter-level experiments.

If we believe that the mechanism is psychosomatic, we would be directed towards different mechanism sketches. The place where the mechanism starts would be the brain. In this scenario, blood poisoning would not be something to explore. If we believe that there are two mechanisms, of different types, operating parallel to each other, that would also provide useful orientation: it would entail that, in order to give a good mechanistic explanation, we have to give models for *two* mechanisms instead of one.

5.3 *Orientation experiments versus inter-level experiments: structural features*

Let us have closer look at the structure of Semmelweis' experiments. About his test of the priest hypothesis, he writes:

During my first period of service, I appealed to the compassion of the servant of God and arranged for him to come by a less direct route, without bells, and without passing through the other clinic rooms. Thus no one outside the room containing the ill patients knew of the priest's presence. The two clinics were made identical in this respect as well, but the mortality rate was unaffected. (1983, pp. 71-73)

The line of reasoning is clearly hypothetico-deductive: if the H (the hypothesis that there is a psychosomatic mechanism that involves the priest) is true, then one expects that a certain operation O (introducing the new route) has a certain effect E (lower mortality rate). However, the mortality rate remains unaffected, so we reject H. More formally, the experiment starts with the establishment that $H \rightarrow (O \rightarrow E)$. Then we manipulate the world so that O is the case. We observe $\neg E$. O and not-E together imply $\neg(O \rightarrow E)$, so we can reject H by modus tollens. Thus, this experiment exhibits the three structural features mentioned in section 3: it requires intervention (choosing another route), there is no localization of any parts of the mechanism, and the intervention is intra-level, since the relata of the supposed causal relation are the explanandum phenomenon, namely the occurrence of childbed fever, and the environment, namely the presence or absence of the priest.

The experiment with respect to the cadaverous particles is well known (see e.g. Loudon 1992, p. 66) and has the same structure: Semmelweis reasons that if H (the hypothesis that cadaverous particles adhering to the hands of the doctors are transferred from the post-mortem room to the labour wards of the 1st maternity clinic) is true then washing hands in disinfectant before going to the labour wards (operation O) is expected to reduce mortality rates (effect E). This effect occurs, so the hypothesis is confirmed. Again, the experiment requires intervention, but only at the level of the explanandum phenomenon and its environment, and there is no localization of parts. Needless to say, in our views this is not the end of the investigation. Rather, it is a signal that the orientation phase has been completed, so that the hard work can now begin. This is a fundamental difference between our view (and that of the mechanists in general) and Hempel's: he considers the research process to be finished after the HD-experiments are done.

5.4 *Orientation experiments versus inter-level experiments: goals*

Thus, the orientation experiments performed by Semmelweis exhibit the same structural features as do the experiments on the electric eels. But what about the goals? The general goal, namely to provide a rough characterization of the mechanism, is again

present: from the experiments, we have learned that the mechanism is not psychosomatic, but rather involves infection. However, in this case, there is another piece of information that is provided by the experiments: they suggest a rough, hypothetical localization of the mechanism, namely the uterus rather than the brain. Note that this is not the same as identifying parts of the mechanism and localising them. What we get are hypotheses that are to be investigated further.

Let us explain briefly why further investigation is necessary. As Raphael Scholl (2013, p. 74) notes, there is another explanation for the effectiveness of hand-washing: if the chlorinated lime solution cures puerperal fever, it is also expected that hand-washing reduces mortality (the curative substance is then transferred from physician to patients during examinations). According to Scholl (2013, p. 75) Semmelweis was aware of the need for further investigation. It motivated him to do animal experiments (with rabbits) that are, according to Scholl, usually neglected by philosophers and historians of science but are considered important by Semmelweis himself. In these experiments the uteri of the rats were treated with several fluids (e.g. pus from abscesses) and then the pathological changes were observed. This illustrates two points. First, the importance of performing orientation experiments to rule out proposals about the qualitative character of the mechanism: if Semmelweis had not ruled out the psycho-somatic proposal, he would never have performed the experiments on animals. Second, the usefulness of the hypothetical, localization that results from the orientation experiments: this leads Semmelweis to treat the *uteri* of the rats with pus and other fluids, rather than say have the rats swallow them.

6. Third case study: pigeon navigation

Our third case study⁸ is from the 20th century and relates to the capacity of homing pigeons (*Columba livia*) to home. A trained pigeon can be taken from its home, transported over very long distances (hundreds of miles are not uncommon), in total darkness,⁹ and still find the way back to its home after being released. Besides a final illustration of the structural features, and the general goal, of orientation experiments, this case study will reveal a further additional goal they can serve.

6.1 Do pigeons have a sun compass?

What mechanism could be responsible for the capacity of pigeons to home? One possibility is that the mechanism involves a sun compass, i.e. the pigeons use the position of the sun as their cue. To test this hypothesis, scientists derive certain expectations from the presence of this type of mechanism and then test these expectations by means of orientation experiments. In the case of hand, if pigeons use the position of the sun as a cue, then since

⁸ This section draws upon Keeton & Gould 1986, pp. 575-580.

⁹ Transporting them in total darkness is in fact itself an orientation experiment: it rules out a possible mechanism, namely that the pigeons build up an internal map on their journey from home, so that they can retrace the route. The schema of this experiment is: if pigeons home by means of an internal map, then in the absence of visual clues, the homing capacity should break down. As in fact this capacity is not impaired when transported in total darkness, this counts against this particular candidate mechanism; at least as the only mechanism (the internal map may be used in combination with some other mechanism).

the position of the sun changes, they also need to have some internal sense of time such as to allow them to take into account the westward movement of the sun. To test the proposed mechanism, one can manipulate the internal clock, which keeps track of time by means of visual clues coming from outside. The expectation derived is that if the pigeons are kept in a room whose lights go on and off six hours early, so that the room is lit at midnight, and dark at noon, their ability to home will be impaired. If such a bird is released at noon, it will see the sun in the south but interpret its location as west, flying at a 90 degree angle to its actual route home. For example, if its home is in the south, it will fly east.

Such clock shift experiments constitute another example of orientation experiments. By interventions, they provide evidence that a mechanism including a sun compass is present. However, unlike Craver's inter-level experiments, the variables upon which the intervention is performed, and the variables that change as a result of the intervention, are located at the level of the explanandum-phenomenon and its environment: they do not 'descend into' the mechanism itself. Yet, they result in a rough qualitative characterization of a mechanism that is present.

6.2 *Do pigeons have a magnetic backup compass?*

From observation we know that they are able to exercise this capacity both in sunny weather and on clouded days. The sun compass leaves the second capacity unexplained. Moreover, clock-shift experiments that were done on clouded days revealed that in these circumstances the manipulations of the internal clock of the pigeons did not impair their capacity to find their home. So there is a second mechanism that works differently.

One proposal was that pigeons use a magnetic compass to home. Again, this is only a very rough characterization of the responsible mechanism, but it does generate test implications. In order to test the magnetic compass hypothesis, biologist William T. Keeton released two groups of pigeons, one group that had magnets attached to them, and one group that had brass rods (which function as placebos) attached to them. He performed this experiment both on sunny and on clouded days. On clouded days the birds carrying magnets became disoriented. The hypothetico-deductive structure of this orientation experiment is clear: we have H (pigeons have a magnetic compass which operates *if* the sky is overcast), O (putting magnets around the neck) and an expected effect E (disorientation on clouded days). We see here the same structure as the experiments on eels and puerperal fever, and indeed the experiments on the presence of a solar compass in pigeons: from the hypothesized presence of a mechanism, we derive expectations about the presence or absence of certain causal relations. We then test these expectations by means of orientation experiments.

On sunny days, the ability to home was unaffected in both groups. This means that the magnetic compass is a *backup* compass: it is switched off on sunny days though it could be operational in these circumstances (the magnetic field of the earth is always available). The conclusion from both performances of the experiment taken together is that pigeons have a magnetic compass which operates *if and only if* the sky is overcast.

6.3 *Results of the orientation phase*

According to Keeton, pigeons have the capacity to home on both sunny and clouded days because they have a sun compass and a magnetic backup compass. This is important information which guides our search for mechanistic explanations in the "schema

generation” phase. What we know thanks to Keeton’s experiments is that:

- there are two mechanisms involved.
- the mechanisms differ radically in the type of information they process (angle relative to the sun vs. angle relative to the magnetic field of the earth); so they are qualitatively different.
- the magnetic compass is a backup mechanism, and is switched off on sunny days.

This information guides our search in the next phase in several ways. For instance, we know that there are two mechanisms to unravel. And with respect to the second, we know that we should try to find magnetoreceptors somewhere in the body of pigeons (at this moment, it is still unclear where these are located).

6.4: One final time: orientation experiments versus inter-level experiments.

Reflecting on this case study, it is easy to see that the experiments described in this section exhibit the structural features of orientation experiments: they required intra-level intervention (attaching magnets to pigeons) but do not involve localization of parts of the mechanisms. As for the goals, the general goal is again present, as the experiments result in a rough, qualitative characterization of the mechanisms: “magnetic” and “sun based”. Besides this, an additional benefit of these experiments is that some other knowledge is obtained: in the case at hand, that there are two mechanisms, and that one of those mechanisms serves as a back-up to the other. Thus, both their structural features and their goals set these experiments apart from Craver’s inter-level experiments.

7. Conclusion

In this article, we have argued that the literature on the discovery of mechanisms, though it is very detailed in its description of the different strategies used by scientists to construct detailed models of the mechanisms responsible for all kinds of capacities, has neglected an important strategy that is prominent in the discovery of mechanisms: the orientation experiment. These experiments are crucial in the early stages of investigation, in fact constituting a distinct phase that comes between the identification of the explanandum phenomenon, and schema generation: the orientation phase. It is a phase characterized by proposing the existence of various mechanisms, and then gathering evidence for and against these proposals. This is done by inferring expectations about certain causal relations, and then testing these expectations by intra-level interventions – the orientation experiments themselves. Although they do not offer the kind of detailed information that inter-level experiments presuppose, and they do not involve localizing parts of the mechanism, they do give us a general characterization of the mechanism, which then constrains further research.

We have shown that this orientation phase, along with its particular experimental techniques, has been crucial in the research into at least three capacities: the power of eels to numb the hand, puerperal fever, and the ability of pigeons to home. It is our belief that by studying the history of science, as Darden encourages us to do (2002 p. S364), many more examples of this research strategy can be found.

References

- Allemand, J. N. S. (1756). Kort verhaal van de uitwerkzelen, welke een Amerikaanse vis veroorzaakt op de geenen die hem aanraaken. *Verhandelingen Hollandsche Maatschappye der Weetenschappen* Haarlem 2: 372–379.
- Bechtel, W. & Abrahamsen, A. (2005). Explanation: a mechanistic alternative. *Studies in History and Philosophy of the Biological and Biomedical Sciences* 36: 421-441.
- Borelli, Giovanni Alfonso. 1680–1681. *De Motu Animalium*. 2 vols. Roma: Bernabo.
- Craver, C. F. (2007). *Explaining the Brain*. Oxford: Clarendon Press.
- Darden, L. (2002). Strategies for discovering mechanisms: Schema instantiation, modular subassembly, forward/backward chaining. *Philosophy of Science* 69 (supplement): S354-S365.
- Darden, L. (2009). Discovering mechanisms in molecular biology. Finding and fixing incompleteness and incorrectness. In J. Meheus & T. Nickles, *Models of Discovery and Creativity*. Dordrecht, Springer: 34-55.
- Gronov, L. T. (1758). Brief van D. Laurens Theodorus Gronovius aan C. N. over nieuw-ontdekte zeediertjes en byzonderheden van den siddervis of beef-aal (Leide October 1758). *Uitgezogte Verhandelingen uit de Nieuwste Werken van de Societeiten der Wetenschappen in Europa en van Andere Geleerde Mannen* 3: 464–478.
- Hempel, C. F. (1966). *Philosophy of Natural Science*. Englewood Cliffs: Prentice Hall.
- Illari, P. M. & Williamson, J. (2012). What is a mechanism? Thinking about mechanisms across the sciences. *European Journal for Philosophy of Science* 2:119-135.
- Keeton, W. & Gould, J. (1986). *Biological Science*. New York and London: W. W. Norton & Co (4th edition).
- Koehler, P. J., Finger, S. & Piccolino, M. (2009). The “Eels” of South America: Mid-18th-century Dutch contributions to the theory of animal electricity. *Journal of the History of Biology* 42: 715-763.
- Leishman, W. (1875). *A System of Midwifery* (with additions by John S. Parry). Philadelphia.
- Leuridan, B. (2012). Three problems for the mutual manipulability account of constitutive relevance in mechanisms. *The British Journal for the Philosophy of Science* 63:399-427.
- Loudon, I. (1992). *Death in Childbirth*. Oxford: Clarendon Press.
- Machamer, P.K., Darden, L., and C.F. Craver. (2000). “Thinking About Mechanisms.” *Philosophy of Science* 57:1-25.
- Meigs, C. (1848). *Females and their Diseases. A Series of Letters to his Class*. Philadelphia.
- Richer, J. (1679). Observations astronomiques et physiques faites en l’Isle de Caienne. Various authors (1678–1693). *Recueil d’observations faites en plusieurs voyages par ordre de sa majeste’, pour perfectionner l’astronomie et la geographie*. Paris: Imprimerie Royale.
- Scholl, R. (2013). Causal inference, mechanisms, and the Semmelweis case. *Studies in History and Philosophy of Science* 44: 66-76.
- Semmelweis, I. (1983). *Etiology, Concept and Prophylaxis of Childbed Fever*. Madison, WI: The University of Wisconsin Press.