Philosophica 72 (2003) pp. 35-59

THOUGHT EXPERIMENTS IN SCIENCE STUDIES

Petri Ylikoski

In this paper I will examine the role of thought experiments in the social studies of science. More specifically, I will concentrate on two strands of social studies of science: the so-called sociology of scientific knowledge and the naturalistically oriented philosophy of science with interest in social dimensions of science. I will begin by discussing David Hull's views on thought experiments in the study of science. His account will work as a foil that helps me to make some points about thought experiments. As an illustration I will discuss two example of thought experimenting in the social studies of science. The first example is the use of thought experiments by the sociologists of scientific knowledge, and the second case will be the recent work by Philip Kitcher on division of cognitive labour. With the help of these cases, I argue that Hull's negative attitude towards the use of thought experiments requires some tempering.

The notion of thought experiment will be understood broadly in this paper. One could also talk about imaginary or hypothetical examples. In social sciences the contrast for thought experiments is not an experiment, but an empirical case study. Accordingly, thought experiment in social science in an imaginary case study or a made up historical scenario. The focus of my discussion will be on the various functions and argumentative roles thought experiments have or could have in social studies of science. An empirical experiment (or a case study) is not an argument by itself, nor is a thought experiment. However, they both are used in making arguments and my aim is to look *how* they are used.

1. Naturalistic antipathy towards imaginary cases

Thought experiments are not a big issue in the recent social studies of science. They are rarely used and the attitude towards them is mostly negative. As a consequence, there is not much discussion about them. David Hull (2001) is an exception to this pattern. His two papers "A Function for Actual Examples in Philosophy of Science" (originally published in 1989) and "That Just Don't Sound Right: A Plea for Real Examples" (originally published in 1997) provide an account of thought experiments that represents the attitudes of the naturalistically oriented philosophers of science in general. (See Cummins, 1998 for a very similar attitude.)

According to Hull imaginary examples have done a massive damage to philosophy, particularly to the philosophy of science. (Hull, 2001: 219) He argues against the use of thought experiments as evidence or as a means of persuasion. He does not have objections against thought experiments if they are used merely as expository devices. However, he points out that they are usually used for more than simple illustration. Intentionally, or not, they often end up persuading people.

Hull has three principal complaints against the use of though experiments in philosophy. The first problem with thought experiments is that they inhibit innovation. This can happen in two ways. First, thought experiments have a conservative influence since they tend to rely on well-entrenched intuitions. As a consequence, the new ideas just do not sound right. For Hull, the intuitions do not have much evidential value. He points out that progress in science sometimes requires challenging common sense or received philosophical intuitions, so there is no reason to keep them sacrosanct. (Hull, 2001: 197)

A thought experiment can also inhibit philosophical progress by diverting the attention from the original problem. Instead of just being convenient illustrative tools, some thought experiments become issues themselves. Hull's primary example is Nelson Goodman's case of grue emeralds. According Hull this particular thought experiment has diverted massive attention to itself at the expense of the original philosophical issue. Philosophers have spent forty years discussing what a world would be like in which predicates like 'grue' and 'bleen' would be somehow 'natural'. All this effort spent on 'trivial illustration' seems like wasting philosophical resources. For Hull, the lesson of grue emeralds is general:

although thought experiments might seem simple and easy to explain, that is rarely the case. When the background details are added, the thought experiment becomes easily very complex and untractable. (Hull, 2001: 200) This is a fatal failure, since the principal reason for using an imaginary case instead of real one is the supposed simplicity and clarity of the former.

The second problem with thought experiments is that they are often incoherent or incomprehensible. According to Hull, this is especially true with 'cute' ones that philosophers like to entertain. By incoherence of some thought experiments he refers to the impossibility of conceiving the states of affairs described in them. He points out that this impossibility does not seem to be universal, since some people claim to able to conceive them. There are two possible explanations for this phenomenon. Either those who cannot conceive these imaginary situations are somehow psychologically deficient or people just happen to have widely different standards of conceivability. Hull seems to favour the latter explanation. (Hull, 2001: 197)

The third, and the most important, problem concerns the absence of detailed standards for the evaluation of thought experiments. According to Hull, typical thought experiments in analytical philosophy have three major deficiencies. First, they often lack sufficient detail. The thought experiments are either described so briefly that it is very difficult or impossible to understand them or to make unambiguous judgments about them. The second problem is that they lack a theoretical context that would allow the audience to fill out the lacking details by themselves. This is in a sharp contrast with the real examples where it is always possible to investigate background assumptions and conditions. Hull also points out that there is no general consensus on methodology for adding empirical details to the imaginary cases. Everyone just seems to fill in the details the way she finds convenient for her purposes. The third problem is that thought experiments trade on a notion of conceivability that is unexplicated. There is no generally accepted theory of conceivability or of the relationship between conceivability and other modal notions. (Hull, 2001: 198, 201)

Hull claims that correcting these deficiencies is bound to be difficult. Of course, one could try to go around the lack of detail by avoiding too outlandish examples and keeping them close to the actual states of affairs. For example, one could make the requirement that the examples conform to the known laws of nature. The trouble is that certain areas of science are tightly organised. This has the consequence that the modifications ramify easily. If our imaginary examples have to be consistent with everything else we know, coming up with these examples is bound to be difficult. It is not easy to keep 'everything else equal'.

As an alternative to these 'harder' imaginary cases, Hull proposes that philosophers use real examples whenever they can. His suggestion to his fellow philosophers is the following:

> *... proceed more slowly from the real world to possible worlds. Fully exploit the world as we know it before conjuring up exotic possible worlds.*" (Hull, 2001: 200 – italics in the original)

The idea is to use hypothetical examples only when we run out of real ones. Furthermore, once we have educated our intuitions with real examples, we are in a better position to come up with hypothetical scenarios, if needed.

Hull points out a number of advantages with real examples. First, one does not need any theory of conceivability to deal with real cases. All modal theories agree that what is actual is possible. The second virtue of real examples is that they always have a context that can provide us with a wealth of background information. Furthermore, there are general standards for filling in details, the rules of empirical enquiry, so there is no need for stipulation. This is in stark contrast with the lack of standards in the case of thought experiments. Finally, real examples can be used as evidence, a status Hull denies for made-up thought experiments.

Is (analytic) philosophy without thought experiments possible? According to Hull it is:

> In thirty years in publishing, I have never attempted to clarify a concept or support a position by reference to fictitious example, except when I have been forced to respond reluctantly to the examples made up by others to criticize my views. (Hull, 2001: 205)

Some comments on Hull's discussion of thought experiments are in place. Let us start with the claim that thought experiments are a hindrance to the innovation because they often rely on well-entrenched intuitions. I do not think that Hull's diagnosis is very accurate. Thought experiments do not always have a conservative effect. They can also make us face some novel situations that force us to abandon our earlier ways of thinking. In this way a thought experiment can have a destructive effect on our old conceptions. Furthermore, as I will argue later in this paper, thought experimenting can be a valuable tool in theory development. Thought experiment can also be used to develop new ideas and to show the limitations of the older. These observations show that the consequences of the use of thought experiments depend on how they are used, not on their essential nature.

I agree with Hull that there is a problem with a heavy use of thought experiments in philosophical work. However, the problem is not the relying on well-entrenched intuitions. I would rather say that often the problem is that they *create* well-entrenched intuitions. Furthermore, the problem is not so much with thought experiments themselves, but with any badly chosen set of examples. Consider for example the long debates about knowledge and justification in analytical epistemology. The problem is not really that the proposed analyses of knowledge are too closely tied to the pre-analytical intuitions. The problem is rather that most of the intuitions that drive the analyses are artefacts created by the earlier discussion. It is not that the intuitions drive the discussion, to the contrary, the discussion drives the intuitions. In fact, it would be helpful if we had some well-entrenched intuitions. One could make similar observations about many discussions in philosophy of science.

What about the charge that many thought experiments are often incoherent or incomprehensible? I do not see any reason to defend misconceived thought experiments, but again, I want to point out that this is not an intrinsic problem with thought experiments. There are also invalid arguments, unsuccessful experiments and badly conceived empirical case studies. If the thought experiment is in incoherent, incomprehensible or if it does not suggest any intuitions, it is failed. Similarly, if the thought experiment brings about widely differing intuitions, it does not confer much support for any thesis (apart from the one that says that we do not have clear intuitions about this particular case). Its argumentative force is lost. Consequently, it can be dropped from the discussion as irrelevant.

This observation suggests a general thesis about the evidential value of the intuitions triggered by thought experiment: they only have force if both parties of the dispute share them. The whole idea of the argument is to appeal to the intuitions a person has independently of his explicit theory or principles. If the opponent does not have those intuitions, the argument loses all of its original force. If one is forced to justify one's intuitive judgments (and to correct the opponents intuitions), the evidential burden shifts from the intuitions to the principles that are used to justify them. Consequently, the thought experiment loses its independent evidential value.

As Hull's terminology shows, the concept of thought experiment applies quite badly to the theory of science or to the social sciences in general. The key question is: what is the intended contrast? In physical sciences, thought experiments can be contrasted with real physically realised experiments. This is not so in philosophy or in social sciences. We have to talk about imaginary states of affairs or scenarios in contrast to empirical observation and real case studies. I do not see any problem here. We can use the term 'thought experiment' to point to the similarities between these things. If someone comes up with a better terminology, we should adopt it, but we should not let the choice of a word to be a hindrance to making some observations.

However, I suggest that we think further about the things we contrast with the imaginary in the social studies of science. Hull says that the contrast is a real example, and I mentioned previously real case studies. In principle, the contrast seems to be very clear: thought experiment talks about things that have not really happened, whereas real examples are about things that have actually taken place. In practice, things are not this simple. The fact that an account of the episode in the history of science is *about* a real event is not enough. There are other requirements. And these requirements bring some uses of historical episodes in the theory of science closer to the imaginary scenarios. We should require that the case studies used are based on serious historical research. Think about anecdotes (Newton's apple), textbook versions of history or 'rational reconstructions' used by philosophers of science (or by psychologists studying scientific discovery). The fact that these accounts are allegedly about real events should not automatically raise their evidential value above that of imaginary examples. (A similar point could be made about the highly stylized accounts of real experiments found in the textbooks of physics or chemistry.)

Despite this, Hull is right in pointing out that there is a crucial difference between real and imaginary examples. The difference is that while there is a general methodology for checking and correcting an account of a real event, no such thing exists for thought experiments. But this does not show that there cannot be a methodology for thought experimenting. For example, consider theoretical model building that has a central role in game theory. The situations that game theoretical models describe are as imaginary as wildest speculations by philosophers. Game theorist simplifies, idealizes and stipulates. But it would be strange to claim that game theorists do not have any methodology in these exercises. (I will return to model building in the end of this paper.) Another example of quite a disciplined form of thought experimenting is computer simulation. As Lennox (1991: 243) points out a computer simulation is in principle a mechanically aided form of thought experiment. The simulation does not take place in one's mind, but it is not a real experiment either. (For computer simulation in science studies, see Ahrweiler & Gilbert, 1998) These two examples show that thought experiments do not have to be as bad and suspect as Hull suggests. There is a distinction between good and bad thought experiments as there is one between good and bad experiments. With this simple observation in mind, let us have a look at various roles thought experiments have in the social studies of science.

2. How thought experiments are used in the sociology of scientific knowledge

Sociologists of scientific knowledge seem to share Hull's naturalistic antipathy towards thought experiments. This is not surprising, Like David Hull, they have for over 25 years advocated a more empirical attitude toward science and criticised traditional philosophy of science. They also share the same enthusiasm for real empirical case studies and a similar naturalistic stance towards knowledge.

For example, one looks in vain for thought experiments in Barnes, Bloor & Henry's (1996) recent statement of the basic ideas behind the sociology of scientific knowledge (SSK). However these are some thought experiments to be found in the earlier writings. In accordance with the naturalistic attitude, none of them is given an evidential role. They all illustrate philosophical ideas behind the sociological approach to the study of scientific knowledge. They deal with abstract philosophical theses that are difficult to illustrate with empirical examples (or relevant empirical studies are still to be made). Neither are they especially bizarre or wild. The imaginary part can easily dispensed with, or it has only a non-essential role in the argument.

2.1 An Azande anthropologist

As a first example, let us consider a thought experiment by David Bloor in his *Knowledge and Social Imagery* (1991 – originally published 1976). Bloor asks us to imagine an alien anthropologist (maybe a Zande anthropologist as suggested by Bruno Latour (1987: 185-195)) who after some observations about the Western society reasons as follows: "...in this culture a murderer is someone who deliberately kills someone. Bomber pilots deliberately kill people. Therefore they are murderers. " (Bloor, 1991: 142). Now, according to Bloor, we (the natives) would resist such a conclusion by arguing that the alien observer did not really understand our concept of a murderer. We would say that he or she does not see the difference between deliberate killing by an individual and a deliberate killing as an act of duty sanctioned by the government. Despite these objections the observer would make the following diagnosis:

> ... [they] see the point of his argument[s] but attempt to evade their logical force by an '*ad hoc*' and shifting tangle of metaphysical distinctions. [...] they have no practical interest in logical conclusions. They prefer their metaphysical jungle because otherwise their whole institution [...] would be threatened. (Bloor, 1991: 143)

In order to see the point of this imaginary scenario, we have to contrast it with a real case study by E. E. Evans-Prichard about the Azande tribe in Africa. According to Evans-Prichard, witchcraft plays a central role in the Azande life. For them, every calamity in human life seems to be due to ill will and malevolent powers of the witches. As a consequence, a Zande does not do anything important without consulting an oracle able to detect witches. Now, being a witch is an inherited physical trait. A male witch will transmit this trait to all his sons and a female witch to all her daughters. This principle seems simple enough and it would seem that once an oracle has identified one witch, we would be able to infer that all males in the witch's clan are witches too. Similarly, showing that a man is not a witch would clear all his male relatives from the accusation of witchcraft. However, the Azande do not act in accordance with these inferences. In theory the whole of clan of the witch should be witches, but in practice only close paternal kinsmen of a known witch are taken to be witches. Were this inconsistency to be pointed out to a Zande, he would deny it. He would refer to the difference between actual witches and potential, or 'cool', witches. The latter are not practicing witchcraft and so they are not real witches. In them, the witchcraft substance is 'cool' and inactive. (Bloor, 1991: 138-141)

For Evans-Prichard this shows that the Azande maintain their logical error on a pain of social upheaval and a need for a radical change in their ways. According to Bloor, Evans-Prichard's claim involves two components. First, there is a real logical contradiction in the Azande views whether or not they see it or not. Second, if they were able to see the error then a major social institution of their society would be untenable. So, the Azande stick to their logical error in order to maintain their social structure. (Bloor, 1991: 139).

Let us now compare these two cases. Bloor has constructed the first case in such a manner that the conclusions of the two cases match exactly. They both point to a seeming logical fallacy or an inconsistency and claim that there would be some devastating consequences to the social structure of the society if the natives were to acknowledge their mistake. This setting allows Bloor to construct his argument for symmetrical treatment of all belief systems. His starting point is the following methodological imperative:

> [...] no institutionalised body of belief depends on its adherents having defective brains or lacking natural rationality. (Bloor, 1991: 176)

This assumption rules out explanations of institutionalized beliefs in terms of widely spread psychological defects. For example, the Azande have the same psychology as we have. Consequently, the possible differences between them and us have to spring from the radically different institutions and ideas they have. This suggests that both cultures should be treated similarly. If we accept the suggested analysis in one case, then we should also accept it in another. And if we are suspicious of one case, we should be suspicious also of the other one.

However, the illustration of this methodological imperative is not the main function of the thought experiment. Bloor also wants to make a point about logic, a point that shows that Evans-Prichard and the alien anthropologist are wrong about both cultures. Bloor appeals to John Stuart Mill's view of logic. For Mill the formal structures of syllogism are connected to actual inferences by an interpretive process. Formal logic provides a mode of display in which our reasoning can always be represented. This display is itself a product of a special intellectual effort and a process of interpretative reasoning. Bloor calls this interpretative process negotiation. What is relevant here is that every application of a formal logical principle is negotiated – there is no fallacy or validity without this process. (Bloor, 1991: 133-134)

An implication of this view is that the claims made by Evans-Prichard (and the alien anthropologist) about the logical inconsistency of systems of belief do not directly follow from logic and the observations they have made. The 'logical' inference they suggest does not threaten the society or call into question entrenched beliefs.

> If the inference ever became an issue the threat would be deftly negotiated away, and this would not in itself be difficult. All that is needed is that a few cunning distinctions be drawn. (Bloor, 1991: 141)

The Azande concept of 'a cool witch' might sound *ad hoc*, but it is an allowed move in the negation process in which we reconstruct our informal thinking into a logical scheme. This shows that Evans-Prichard has misunderstood the nature of logic. The second consequence is that the rationality of the Azande does not require that they have 'an alternative logic' (as suggested by Peter Winch). It just requires that we have the right understanding of the nature of logical reconstruction. (Bloor, 1991: 139-141)

What should be said about this thought experiment and its role in Bloor's argument? It seems to me that the imaginary part of the argument can easily be dispensed. The society the alien anthropologist observes is our own (or it can be replaced with our own). Similarly, we can easily take the place of the alien anthropologist just by taking an external stance towards our own belief system. As a consequence, nothing in Bloor's symmetry argument hinges on the thought experimental part.

However, the thought experiment has a significant mediative role in the argumentation. It is not easy to get people to take an external stance towards their own thinking. For this purpose a narrative involving an alien anthropologist (or an alien from outer space (see for example Collins, 1985: 29-46)) is a helpful device. This ploy is not absolutely necessary from the point of view of the argument, but it makes the argument much more entertaining, comprehensible and compact. Notice also that this thought experiment is used to make a philosophical point, not a sociological point. This is typical for the SSK literature. Thought experiments are used to establish philosophical starting points of the SSK, but they are not used in the sociological analysis itself.

2.2 A tale of two tribes

Our second thought experiment comes from the paper 'Relativism, Rationalism and the Sociology of Scientific Knowledge' by Barry Barnes and David Bloor. Barnes and Bloor suggest the following thought experiment:

> Consider the members of two tribes, T1 and T2, whose cultures are both primitive but otherwise very different from one another. Within each tribe some beliefs will be preferred to others and some reasons accepted as more cogent than others. Each tribe will have a vocabulary for expressing these preferences. Faced with a choice between the beliefs of his own tribe and those of the other, each individual would typically prefer those of his own culture. He would have available to him a number of locally acceptable standards to use in order to assess beliefs and justify his preferences. (Barnes & Bloor, 1982: 26-27)

Barnes and Bloor use this example to illustrate the internal consistency (and plausibility) of their 'relativist' position. According to them, a relativist says about himself just the same that he would say about our imaginary tribesman. Like everybody, he has to sort out beliefs, accepting some and rejecting others. In this sorting he would naturally have preferences, and as with everybody else, these preferences will typically coincide with those of others in his locality. The words like 'true' and 'false' or 'rational' and 'irrational' provide him with an idiom in which those evaluations are expressed. And like everyone else, the relativist will probably prefer his own familiar and accepted beliefs when confronted with an alien culture. And if needed, his local culture will furnish him with norms and standards that can be used to justify such preferences. The crucial point in their argument is that a relativist accepts

> that his preferences and evaluations are as context-bound as those of the tribes T1 and T2. Furthermore, the relativist accepts that the justifications of his preferences cannot be formulated in absolute or context-independent terms. The relativist must acknowledge that, in order to escape circularity, his justifications will have to stop at some point that only has local credibility. (Barnes & Bloor, 1982: 26-27)

Here again we see an anthropological thought experiment used to illustrate philosophical position that is intended to be a foundation for the sociological study of scientific knowledge. This thought experiment is also similar to the previous one in its dispensability: the imaginary part is not required for the making the argument, but the use of the imaginary device helps its comprehension. T1 and T2 could be replaced with two cultures from a textbook in anthropology. This change would, however, only make the passage much longer and possibly distract the reader from the real issue.

2.3 A sceptical experiment in a laboratory

In *A Social History of Truth* (1994: 17-22) Steven Shapin suggest an experiment in distrust. He asks his reader to take any item of present-day factual knowledge that is considered to be a good example of true and reliable knowledge. For his illustration he takes the proposition 'DNA contains cytosine'. Now, one either holds this belief on the basis of one's own firsthand experience or not. In the first case one is an expert and in the latter case one is not. Shapin observes about the members of the latter group:

> Given modern individualistic epistemological, that group ought to be satisfied that they do not possess genuine knowledge at all, although they may be disposed to say that there are *other* individuals who are properly entitled to that knowledge on grounds of direct experience (Shapin, 1994: 17)

Therefore, non-experts do not properly know that DNA contains cytosine. How about the experts? Shapin next describes in detail how he got firsthand knowledge of this fact while working in a genetics laboratory. However, he quickly points out that there are reasons to doubt whether he really acquired that knowledge firsthand. How did he know that a certain outcome of a chemical test stood for the presence of cytosine, or that his dried precipitate was really DNA? He and all other people working in the laboratory trusted their equipment and material providers, their teachers and colleagues in other laboratories. So, in fact, their firsthand knowledge was not really firsthand knowledge. (Shapin, 1994: 18)

Shapin could have doubted the assumptions of his experiment, and this is the starting point of the sceptical thought experiment. What would have happened if he had started to doubt all claims that were not based on his own direct firsthand experience? This would have required going through the bases for his trust in the identity of the animal tissue he used, the speed of the centrifuge, the reliability of thermometric readings, the qualitative and quantitative make up of various solvents, the rules of arithmetic and so on. It would have taken enormous amounts of time and resources. Furthermore, his sceptical attitude would break the moral fabric of the laboratory. The colleagues might first be only annoyed of the sceptic's behaviour, but his distrustful attitude would finally lead to his expulsion from the community. (Shapin, 1994: 19-20)

Shapin concludes that it is safe to assume that no practicing scientist has ever carried scepticism so far. And indeed, this is Shapin's point. For both practical and moral reasons, distrust is only possible in particular instances, not as a general attitude in science. Distrust takes always place on the margins of trusting system. A sceptical attitude is an important component of science, but it is applicable only locally. (Shapin, 1994: 19) Shapin's sceptical experiment is done mentally, so it seems to be a good candidate for a thought experiment. Nobody is prepared to really distrust one's scientific colleagues, instrument makers, teachers and other authorities in a total manner suggested by the experiment. For this reason, the experiment is a kind of mental simulation, an *as if*-exercise. The sceptical experiment functions as a *reductio ad absurdum* of the naïve individualist epistemology. If the individualism were followed consistently, we would end up having no knowledge at all, which would be an absurd conclusion. The experiment also illustrates the ineradicable role of trust in knowledge production, even when we are trying to find an individual and independent grounding for our knowledge. (Shapin, 1994: 21)

3. Implicit thought experimenting in the construction of explanatory claims

The above examples show that the thought experiments are not prevalent in the sociology of scientific knowledge. However, there is one prevalent and important form of hypothetical reasoning that could be regarded to involve thought experimenting. The activity I have in mind is explanation. One of the aims of the SKK is to create *explanatory* knowledge about science. In practice, this means that sociologists of knowledge try to explain historical episodes in science.

How does thought experimenting relate to explanation of historical events? The answer is that, when making a claim about explanatory relevance (or more generally about historical significance) one is committed to a counterfactual claim. If the *explanans* had been different, the *explanandum* had not happened. In order to do this, one has to come up with an alternative scenario of how the things might have gone if the *explanans* had not happened.

In the construction of the explanatory counterfactual one only has to show that the factor in question made a relevant difference, but there is no need to establish a full alternative possible world. Therefore it could be claimed that explanation is not a full-blown thought experiment, but I think it is enough of a thought experiment. (Ylikoski, 2001, see also De Mey & Weber, 2003)

Let us take a look at Donald MacKenzie's *Statistics in Britain 1865-1930* (1981). Among other things, MacKenzie studies the influence of the eugenical movement in the development of statistical methodology. One of the statisticians in his interest is Karl Pearson. (For a fuller account, see Ylikoski, 2001: Chapter 6)

According to MacKenzie, Pearson had political and social goals related to the eugenics movement that motivated his scientific work. More specifically, his explanatory claim is that Pearson's commitment to the eugenics movement was an important influence on the formation of the cognitive goals of his statistical work. MacKenzie's explanatory claim rests on the following counterfactual:

> If Pearson had not been influenced by eugenics and committed to it, he would not have chosen the same topics of scientific research as he did and, as a consequence, he would not have developed such

statistical methods as he did.

The nature of MacKenzie's explanatory claim is often misunderstood. He is not claiming that social interests had a direct influence on Pearson's scientific choices and decisions. Rather, the eugenical motivation influenced his evaluations of the importance of certain statistical research problems, and his evaluation criteria for the results of such research. In more general terms, the basic idea of this explanatory pattern is the following. Political and social goals explain scientist's cognitive or scientific goals, which in turn explain many of the details of her scientific work.

The challenge for explanatory claims of this kind is to provide empirical evidence for the explanatory relevance of the counterfactual. To provide such evidence MacKenzie examines two scientific controversies that Pearson participated in. The first is the public controversy between biometricians and early Mendelians in the beginning of this century (MacKenzie, 1981: Chapter 6). The second is the more scholarly debate between Karl Pearson and George Yule over the statistical analysis of nominal variables (MacKenzie, 1981: Chapter 7). The basic explanatory pattern in both of these cases is similar. The controversy and its continuation are explained by the different cognitive goals of the participants, and these differences are in turn explained by their different social and political goals.

For example, in MacKenzie's analysis of the Pearson-Yule debate the *explanandum* is the difference between Pearson's and Yule's mathematical statistics and the *explanans* the difference in the goals of their statistical activities. According to MacKenzie, Pearson showed both great effort and scientific integrity in his pursuit of the research program of eugenics. This program created a specific data-processing demand for Pearson. He needed measures for the associations of nominal data that were numerically comparable to the interval-level correlation coefficient. To meet this demand, Pearson devised a series of measures for association. MacKenzie notes that Pearson's interest in measures of association diminished when his practical statistical concerns shifted. This suggests that there is an important connection between these two issues in Pearson's scientific work. In contrast, Yule did not have any involvement with any eugenics research program. His practical commitments did not give rise to a similar dominant desideratum and, as a consequence, he was prompted to develop a looser and more pluralistic approach to the measurement of association. In MacKenzie's account, this rather esoteric controversy was a result of different cognitive goals, and it was sustained not because the participants did not understand each other's positions, but because their cognitive goals were proxies for their different practical goals. (MacKenzie, 1981: Chapter 7)

I take MacKenzie's explanatory pattern to be sensible in principle. Of course, there might be some other plausible explanation candidates for this particular difference, but the evaluation of the explanation against the historical data is not my concern here. The point is that this example shows how making an explanatory claim involves counterfactual reasoning. We have to sketch an alternative scenario that shows how things might have developed if the explanatory factor did not prevail. One of the strengths of MacKenzie's study is his use of contrasts. Setting Pearson against his contemporary Yule shows in concrete terms what kind of work Pearson might have done if he had not been interested in eugenics. The more general point is that this shows how the construction and the evaluation of the explanatory counterfactuals can be based on historical evidence.

In order to see how the construction of the alternative scenario could have problems, let us take a look at another example. In his *The Social Construction of What?* (1999) Ian Hacking explores the possible effects military funding and direction of scientific research might have had in the development of physical sciences.

Hacking's example involves an explanation that is based on unintended consequences of action. This kind of explanation pattern could be called an unintentional filtering explanation. (Ylikoski, 2001: 146-149) Hacking's *explanandum* is what he calls 'the form of scientific knowledge'. By this concept Hacking refers to the idea that existing knowledge somehow determines which issues are *possible* candidates for topics of scientific research. His claim in relation to the military involvement in physical science research is that it might have influenced the form of physical science knowledge in a manner that precludes some questions about physical reality outside of what is considered to be meaningful and feasible topics of research. Of course, this influence might not have been only negative: if there had not been huge military investment, certain aspects of physical reality might have stayed outside of what are currently considered as meaningful research problems in physics. Hacking's thesis is not very specific or detailed, but the general point he wants to make is clear: military funding made a difference in the form of current physical knowledge. (Hacking, 1999: Chapter 6.)

Now let us take a closer look at what is going on in this explanation. The aim of the military and the defence industry is to promote research that is useful for the development of weapon systems and other military applications. This goal explains their funding choices. On the other hand, the scientists did not necessarily share this objective, for their idea may have been to use the resources provided to advance their own professional and cognitive goals. However, because of their dependency on funding and other resources, their research activities served the purposes of the military and can be partly explained by these interests. The interesting part in his scenario is that neither party had the goal of shaping the form of scientific knowledge to what it is. Actually, given that Ian Hacking coined this concept in the 1980's, neither party had the faintest idea of the issue.
Hacking's explanatory counterfactual is the following:

> If there had not been massive military interest in the physical sciences, the form of knowledge in physics today would have been different.

Now, in principle, this explanation sounds sensible. The fact that the *explanandum* is unintended by the agents in no way invalidates the explanatory pattern. The only problem concerns the ambiguity of the contrast. Hacking's sketchy discussion does not give a very concrete idea as to what could have been different. This is partly due his concept of form of knowledge, and partly due to the scale of issues he discusses. I guess we might try to imagine what physics would have looked like if it had not become Big Science, but thinking about alternative forms of knowledge might prove too difficult.

The failure here seems to be the same as with some philosophical thought experiments criticised by David Hull. The imaginary scenario lacks sufficient detail to allow the audience to evaluate it. Consequently, any judgments about the validity of Hacking's claim have to be suspended.

4. How thought experiments could be used

In his paper, "Darwinian Thought Experiments: A Function for Just-So Stories" James Lennox (1991) argues that thought experiments have a central role in Darwin's argument in *On the Origin of Species*. According to Lennox, Darwin used thought experiments as a test for the explanatory potential of his theory. Darwin was not arguing for the truth of his theory, rather, he was arguing that the theory was in principle able to explain a wide range of biological phenomena. His thought experiments displayed in a vivid and concrete way that,

> [...] *if* each of the mechanisms and processes referred to by Darwin's theory were to interact in particular ways, there *would* occur an accumulation of minute, random variations in a particular direction, culminating in distinct varieties and, eventually, new species. (Lennox, 1991: 229)

By using these thought experiments Darwin was able to answer to his critics, who claimed that his theory could not even in principle explain things that he claimed it could explain. As Lennox points out, this is a fully legitimate epistemic role for thought experiments. They cannot show whether a particular theory is true or false, but they can provide evidence for or against its explanatory potential.

Lennox also argues that there was a second important role for thought experiments. According to him, Darwin's thought experiments helped in the development and articulation of his theory. Lennox (1991: 230-235) describes in detail how one of Darwin's critics, Fleeming Jenkin, picked up some of Darwin's thought experiments and forced Darwin to articulate his theory and its assumptions. By making explicit some of the Darwin's implicit assumptions and by drawing out some undesired conclusions of Darwin's model he pointed out some dangerous ambiguities in Darwin's theory. For this purpose Jenkin used Darwin's own thought experiments. Darwin was able to resolve successfully these ambiguities in the later editions of *On the Origin of Species*, but it is clear that the debate around these thought experiments helped him to find and solve some conceptual problems in his theory.

According to Lennox, the use of thought experiments in these two roles is not unique to Darwin. To the contrary, he argues that throughout the history of evolutionary biology some of the greatest theoretical triumphs have began as thought experiments. These thought experiments showed that the explanation of a certain phenomenon was within the explanatory reach of the theory. (Lennox, 1991: 238) Furthermore, he points out that the so-called Just-So Stories, can have a legitimate and important role as thought experiments. When understood properly, these scenarios should be understood as theoretical how-possible accounts, not as confirmed explanations or as evidence for particular hypothesis. By keeping this distinction in mind, we can both see the legitimate role of Just-So Stories in theory development and see the important points made against them by the critics of the Panglossian paradigm in biology. (Lennox, 1991: 238-241)

Robert Brandon's account of how-possible explanations in the theory evolution agrees with Lennox's observations. Brandon notes that much of the theoretical work in mathematical population genetics consists of the construction and testing of how-possible explanations. He also points out that, how-possible explanations can have various functions. First, they advance our theoretical repertoire and understanding by telling us what could happen. In other words, they enlarge our view of possible mechanisms and their capabilities. Second, these extensions of theoretical repertoire can work as building blocks of how-actually explanations. Finally, like in Darwin's case, they help to answer some impossibility claims. (Brandon, 1990: 184)

Evolutionary biology is not the only science that uses thought experimenting in a systematic manner in theory development. Another example is economics, especially game theory. Most of the model building done in game theory can be described as straightforward thought experimenting.

The importance of model building in these sciences has not gone unnoticed. Kitcher (2002) takes the example of population genetics seriously. He explicitly suggests an analogy between biology and science studies. According to Kitcher, the history of ecology shows a development where mathematical population genetics and natural history based on field research developed in a mutually benefiting manner. The integration of mathematical models and field data has benefited the evolutionary studies of animal behaviour. Kitcher suggests that similar developmental pattern would also be fruitful in the theory of science. Theoretical model-builders and empirical observers should see the value of this kind of co-operation.

The key to the success of the integration of ecology was the mutual relevance of two ways of doing biology. Mathematically oriented biologists formulated hypotheses about the expected characteristics of organisms by devising models both about the optimal forms of phenotypes and about their constraints. These endeavours were based on the observations made by field naturalists and, in turn, their results provided concepts and hypotheses the field naturalists could take to the field for a test. The strategy did not always work, but at its best this joint activity was extremely fruitful providing increased understanding of the natural world. (Kitcher, 2002: 263)

Kitcher suggests that a similar dialectical process could be helpful in the development of the science studies. In science studies, historical and sociological studies represent a tradition of empirical observation analogical to the field studies by natural historians. What is missing is the analogical counterpart for the mathematical population genetics. In Kitcher's vision, his approach developed in *The Advancement of Science* (1993: Chapter 8), could play the role of mathematical model building. Let us take a closer look at this approach.

The basic idea in Kitcher's approach is to build highly simplified models and to see how various assumptions about the agents (or about their resources etc.) affect their behaviour, especially at the community level. Kitcher employs an analytic idiom inspired by Bayesian decision theory, microeconomics, and population biology. He writes that:

> The advantage of this idiom is that it enables me to formulate my problems with some precision, and that precision is important for both identifying consequences and disclosing previously hidden assumptions. Precision is bought at the cost of realism. My toy scientists do not behave like real scientists, and my toy communities are not real communities. (Kitcher, 1993: 305)

Kitcher's toy communities are imaginary, simplified, and there are no circumstances where they could be found in real life. Not only does he abstract away the whole social fabric of science and society around it, he also gives a highly unrealistic picture of the individual cognitive processes. Clearly they are similar to the hypothetical situations imagined in more traditional thought experiments.

To get more a concrete idea of these toy communities, let us have a look at one example. One of the issues discussed by Kitcher concerns the cognitive division of labour in a scientific community. His starting point is the following highly idealised situation:

> Once there was a very important molecule (VIM). Many people in the chemical community wanted to know the structure of VIM. Two methods for fathoming the structure were available. [...] Everybody agreed that the chances that an individual would discover the structure of VIM by using method I were greater than the chances that that individual would discover the structure by using method II. (Kitcher, 1993: 346)

Although one can see some similarities with history of science when one replaces VIM with DNA, this stipulated situation is clearly thought experimental. But this is not a point against the scenario. The idea is to start with simple (and unrealistic) assumptions and see what happens when various assumptions (about individual motivations, probability assignments, workforce requirements of the methods, etc.) are changed. In the next phase, one can make the scenario more complicated (and realistic) and see whether there are any changes in the results. In this manner it is possible to locate some of the effects one's idealizations have.

I will not go to the specific results of Kitcher's thought experiments. From the point of view of my argument, the possible advantages of this sort of theorizing are more relevant. Without evaluating Kitcher's specific results, we can say that there are two kinds of possible advantages.

First, as Kitcher points out, the models can useful by enabling us to confirm or refine qualitative arguments about the dynamics of the research process (Kitcher, 2000: 40, Kitcher, 2002: 265). This happens in two ways. Despite their apparently arbitrary assumptions, mathematical models can show that the original intuitive idea does not have a hidden flaw. In other words, the model works as a kind of consistency test for more qualitative arguments. Secondly, the model can show the limits of application of the original argument by making all its assumptions fully explicit. This is a very important advantage. It is easy to argue that various sorts of incentives could have epistemically positive or negative results in science, but it is more difficult to spell out the specific circumstances when this happens. Outside messy empirical testing, model building is the only way to make the claims more specific. In this manner, they help to strengthen the original argument.

The second advantage of models is that they help us to pose concrete questions to empirical investigation. This is the dialectic Kitcher observed in the history of ecology. The abstract models can raise issues that empirical researchers can check. They can also help them to make their research questions more precise. It needs to be emphasized here that this does not mean that the empirical research in history and sociology is subordinated to the mathematical theorists. It is important that the exchange symmetrical: ideas and critiques should go to both directions. The model building should be informed by the best empirical research. This is a crucial condition for the advance Kitcher is hoping to achieve.

One example of an approach to science studies that could profit from the adoption of Kitcher's idea is David Hull's evolutionary theory of science. His theory can be characterised as a combination of evolutionary theory, economics and sociology. Its basic idea is to give a kind of invisible hand account of how science works. In it scientists promote the goals of science (the production of credible, critically evaluated knowledge) in an optimal way by promoting their (selfish) interests (satisfaction of curiosity, maximization of credit). It also includes an evolutionary theory of conceptual and social development of science. (Hull, 1988 and 2001)

Without going into any details, I want to point out two challenges to Hull's theory. The first challenge is to connect the abstract theory to empirical reality. It is not quite clear how the theory should be applied to particular historical episodes. The theory can be fitted to the empirical material in a number of ways, and this creates a suspicion that it might be too abstract and general for its own good. Putting it through stringent tests is too difficult. Is Hull's theory a really empirical theory or is it a metatheory that can be used to redescribe almost anything?

The second challenge is related to the first one. It concerns the prospects for the further development of the theory. Hull's theory was presented in 1988, and not much has happened since. At least a partial reason for this non-progress is that nobody really knows how to further develop an evolutionary account of science. Should one refine the abstract theory or try to apply it to empirical material? Both are difficult. If one chooses the refining, one has to ask how and in which direction. And if one chooses the empirical direction, the crucial question is again how to do that.

My suggestion is that piecemeal modelling suggested by Kitcher could provide a promising way forward. It does not itself solve the problems, but it provides a very useful tool for tackling them. The application of Kitcher's suggestion to Hull's theory is quite straightforward. Hull's theory draws from economics and evolutionary theory which both use modelling extensively as a tool for theory development. So in principle, Hull and other advocates of the evolutionary account should just start practicing similar forms of theory development.

What about Kitcher's (2000) suggestion that the modelling is the right direction also for the sociologists of science? The issue is much less straightforward. It is characteristic for SSK that it lacks a layer of explicit sociological theory. There is explicit discussion about the philosophical underpinnings and there are plenty of detailed case studies. In order to initiate interesting modelling activities, some concrete sociological theories should be explicated first. And it could take a long time, especially if the people suggesting the model building approach are quite ignorant of the aims and history of sociology. (The attitude of Kitcher (2000) is quite hostile – an example of an initiative for interdisciplinarity that will never work.)

Although I advocate Kitcher's piecemeal modelling approach, this does not imply that I am committed to the idea that his particular use of this approach is the most fertile one. To the contrary, I think his extremely individualistic, or atomistic, assumptions about the social agents make his models rather uninteresting from the point of view of understanding a social institution like science. Nevertheless, his point about the importance of model building as an important tool in theory development should be taken seriously. Models can be built in a variety of ways and by using various different assumptions. And as Kitcher himself emphasises, the most crucial element is the dialogue between people doing modelling and people doing field research.

5. Conclusion

I started this paper with David Hull's critique of thought experiments in the philosophy of science. I hope my discussion in this paper shows that his theses against the use of thought experiments should be rejected or at least modified. This point does not concern his discussion of specific thought experiments in philosophy, but his generalizations from them. My basic claim is that his sample of thought experiments is biased in a manner that prevents him from seeing some important and fully legitimate forms of thought experimenting. The examples of evolutionary biology and game theory suggest that there can be a positive epistemic role for thought experiments in social studies of science. Thought experiments could provide a way to articulate and develop theoretical ideas in a manner that can later lead to empirical applications and tests. Their example also shows that thought experimenting can be done in a systematic and rigorous manner. And finally, they show that thought experimenting can require specific skills and talents. Of course, these points do not challenge the fundamental point Hull wanted to make about the crucial importance of real cases. If thought experimenting is practiced, it should be performed in a close connection with empirical studies.

Helsinki Collegium for Advanced Studies

REFERENCES

Ahrweiler, P & Gilbert, N. (eds.) (1998), *Computer Simulations in Science and Technology Studies*. Berlin: Springer.

Barnes, Barry & Bloor, David (1982), 'Relativism, Rationalism and the Sociology of Scientific Knowledge', pp. 21-47 in Hollis, M. & Lukes, S. (eds.): *Rationality and Relativism*. Oxford: Basil Blackwell.

Barnes, Barry, Bloor, David & Henry, John (1996), *Scientific Knowledge. A Sociological Analysis*. London: Athlone Press.

Bloor, David (1991), *Knowledge and Social Imagery (2. ed.)*. Chicago: The University of Chicago Press.

Brandon, Robert N. (1990), *Adaptation and Environment*. Princeton: Princeton University Press.

Collins, Harry (1985), *Changing Order. Replication and Induction in Scientific Practice*. London: SAGE.

Cummins, Robert (1998), 'Reflection on Reflective Equilibrium', pp. 113-128 in DePaul & Ramsey (eds.): *Rethinking Intuition*. Lanham: Rowman & Littlefield.

De Mey, Tim & Weber, Erik (2003), 'Explanation and Thought Experiments in History', *History and Theory* **42**, pp. 28-38.

Hacking, Ian (1999), *The Social Construction of What?* Harvard: Harvard University Press.

Hull, David L. (1988), *Science as a Process. An Evolutionary Account of the Social and Conceptual Development of Science*. Chicago: The University of Chicago Press.

Hull, David L. (2001), *Science and Selection. Essays on Biological Evolution and the Philosophy of Science*. Cambridge: Cambridge University Press.

Kitcher, Philip (1993), *The Advancement of Science. Science without Legend, Objectivity without Illusions*. Oxford: Oxford University Press.

Kitcher, Philip (2000), 'Reviving the Sociology of Science', in Howard, D. A. (ed.) *PSA98 Part II, Philosophy of Science* **67** (Supplement), pp. 33-44.

Kitcher, Philip (2002), Social psychology and the theory of science, pp. 263-281 in Carruthers, P, Stich, S. and Siegal, M. (eds.): *The cognitive basis of science*. Cambridge: Cambridge University Press.

Lennox, James G. (1991), 'Darwinian Thought Experiments: A Function for Just-So Stories', pp. 223-245 in Horowitz, T. & Massey, G. J.: *Thought Experiments in Science and Philosophy*. Lanham: Rowman & Littlefield.

MacKenzie, Donald (1981), *Statistics in Britain 1865-1930. The Social Construction of Scientific Knowledge*. Edinburgh: Edinburgh University Press.

Shapin, Steven (1994), *Social History of Truth. Civility and Science in Seventeenth-Century England*. Chicago: The University of Chicago Press.

Ylikoski, Petri (2001), *Understanding Interests and Causal Explanation*. Ph.D.-thesis, Department of Moral and Social Philosophy. University of Helsinki. Available at http://ethesis.helsinki.fi/julkaisut/val/kayta/vk/ylikoski/