STUDIA METODOLOGICZNE
NR 36 • 2016, 113–130
DOI: 10.14746/sm.2016.36.6

JULIAN REISS

Thought Experiments in Economics and the Role of Coherent Explanations*

ABSTRACT. Starting from the 'thought experimenter's dilemma' this paper develops a novel account of thought experiments, according to which these help to establish coherent explanations of phenomena of interest. Though in principle applicable to cases across science, the focus of this paper is thought experiments in economics, five different types of which are distinguished here.

KEY WORDS: thought experiments, economics, methodology, scientific explanation

1. Introduction

That thought experiments (TEs) are widely used across the sciences and social sciences poses a riddle for those who think that we can learn new facts only through observation and experimentation. The view that TEs provide genuine knowledge of new facts conflicts with deeply held beliefs about the sources of our knowledge and make it mysterious how we gain this knowledge. On the other hand, if one maintains that TEs do not provide such knowledge, we have to explain the appearance to the contrary as well as their popularity in science.

* I wish to thank John Davis and an audience at the 2016 Allied Social Science Associations conference in San Francisco for a very valuable discussion of this paper. Special thanks go to Ulrich Kühne, who didn't only provide the framework for thinking about thought experiments the paper presents but also detailed comments on a previous version.

None of the major approaches to thought experimentation in the literature offer a satisfactory answer. Brown's Platonism, which accepts TEs as genuine sources of new knowledge, builds on a highly controversial understanding of the nature and epistemology of scientific laws. Norton's empiricism, which does not, regards TEs as picturesque arguments and involves nothing more mysterious than reasoning from premises. This interpretation is, however, difficult to square with scientific practice.

In this paper I will take up Ulrich Kühne's suggestion that TEs establish *coherent explanations* of phenomena of interest [Kühne, 1995]. Specifically, I will examine different kinds of TEs in economics, defend Kühne's interpretation for these TEs, and argue that this interpretation avoids the above mentioned dilemma. I will argue that establishing coherent explanations is a highly important function especially in economics because of the ease with which researchers can come up with possible explanations, the difficulty and costliness of experiments in this field, and the ambiguity of observational research.

2. The thought experimenter's dilemma

Thought experiments are a widely used tool of scientific investigation in both the natural and the social sciences. Einstein's general theory of relativity, of which we just celebrated the centennial, is famously said to have originated in several TEs [Norton, 1991]. Einstein is also the coinventor of a TE aimed to point to what he regarded as significant shortcomings in the conceptual framework of quantum theory [(Einstein, Podolski et al., 1935]. Quantum mechanics also gave rise to numerous other TEs, including Schrödinger's cat and Heisenberg's gamma-ray microscope. Other famous physical TEs include Maxwell's demon, Newton's bucket, Stevin's derivation of the law of the inclined plane, and Galileo's falling bodies.

The method of thought experimentation is not confined to the natural sciences, however. TEs are ubiquitous in historical social sciences where they are used to establish the causes of singular events [Tetlock, Belkin,

1996; Tetlock, Lebow et al., 2006; Reiss, 2009], economics [Schabas, 2008, Reiss, 2012, Thoma, 2015] and elsewhere in the social sciences [e.g., Roberts, 1993]. The Austrian School of Economics goes so far as to call thought experimentation 'the specific method of economics' [von Mises, 1996 (1949)].[1]

The popularity and widespread use of TEs in the sciences creates a philosophical conundrum for those who subscribe to the empiricist tenet that 'all knowledge comes through the senses' – that is, most of us. This is because TEs, like real experiments, do, or at least appear to, advance our knowledge of reality. But unlike real experiments, they don't involve any new observations or data. Whatever else they are – and there is a good amount of controversy about the proper definition of the term 'thought experiment' – TEs are exercises of thinking, conducted in the armchair and not in the laboratory or observatory.

This philosophical conundrum can be stated succinctly as the following dilemma. The first horn takes the appearances to be correct and TEs to provide genuine knowledge of new facts about the world. This, however, would conflict with deeply held views about the sources of our knowledge and make it mysterious how we gain this knowledge. After all, why and how we can learn from real experiments has been understood well enough since around the time of the scientific revolution, but no analogous epistemology of the TE seems to be available. It is also completely uncontroversial that observations and experiments are sources of new knowledge of the world. The other horn of the dilemma starts from the critical voices that, throughout the history of commentary on TEs, have denied them this status. These critical voices, then, argue that TEs do not provide genuine new knowledge about the world. The problem with this suggestion is that then we have to explain the appearance to the contrary as well as their popularity in science.

Many of the accounts of TEs in the sciences can be understood as accepting either horn of the dilemma. The most prominent attempt to go

[1] A view for which the Austrian School has certainly received much criticism. See for instance Lanny Ebenstein's interview with Milton Friedman in Ebenstein, 2015.

down the first horn is James Robert Brown's [e.g., Brown, 2010]. Brown is a Platonist about the laws of nature. That is, he follows the realists about these laws, such as David Armstrong, Fred Dretske and Michael Tooley, who regard them as relations between properties or universals. Brown also accepts, with Tooley, that uninstantiated universals can exist [Dretske, 1977; Tooley, 1977; Armstrong, 1983]. A particular kind of TEs – 'Platonist TEs', which refute one theory and establish a successor at the same time – are then interpreted as means to access these relations. That is, these TEs provide *a priori* knowledge of the laws of nature.

Both the Platonist view of the laws and the view that TEs should provide access to them are deeply problematic. The Platonist view does in principle provide a criterion to distinguish between genuine laws of nature and accidental regularities. If it is a law that 'All Fs are G', then, on the Platonist view, F-ness and G-ness are real properties or universals, and they stand in a particular relation to each other: the relation of necessitation: N(F, G). This is not the case with accidental regularities. Accidental regularities such as 'All coins in my pocket are made of copper' either fails to relate two genuine properties (presumably, 'being in my pocket' does not pick out a natural kind), or there is no relation of necessitation, or both.

One problem with this account is that it does not come along with an epistemology that enables us to distinguish between genuine laws and accidental regularities empirically. On the Platonist view, laws entail regularities: N(F, G) entails 'All Fs are G'. Thus, we can confirm laws hypothetico-deductively by collecting evidence about regularities. There is an alternative hypothesis, however, that entails the very same evidence: 'Accidentally, all Fs are G'. There is no possible evidence that could distinguish between the two alternatives.

There is another problem that is highly relevant to the social sciences such as economics: there are very few, if any, strict regularities – statements of the kind 'All Fs are G' are subject to exceptions and qualifications. If exceptionlessness is a marker of proper lawhood, then there would be no laws in the social sciences, possibly no laws at all. While it is controversial how damaging the absence of laws would be for the status of the social sciences as genuinely scientific [see Roberts, 2004 versus Kincaid, 2004], it is certainly undesirable to exclude the possibility of finding social

scientific laws in this way. Moreover, the issue of finding an adequate account of TEs in economics would, in this reading, become moot.

These problems concerning the Platonist view of laws are closely related to Brown's Platonism about TEs. How can we know that a TE is successful at revealing a law of nature? In other words, how do we distinguish between good and bad TEs? There is a long list of guidelines aimed at improving the reliability of real experiments, such as 'rule out artefacts', 'empirically investigate the equipment', 'analyse experimental data using appropriate statistical methods' and so on. An experiment is a good one to the extent that these guidelines have been followed. No analogous instructions exist for TEs.

John Norton, one of the most influential contributors to the literature on TEs in natural science, accepts the second horn of the dilemma. He defends a view according to which TEs are nothing but picturesque arguments. Any TE can, in principle, be constructed as a set of propositions in which the thought experimental result appears as a conclusion. The reliability of the TE can, consequently, be determined by assessing the truth of the premises and the quality of the inference from premises to conclusions.

To give an example, Norton reconstructs Galileo's famous falling bodies TE (which Brown takes to confirm his Platonism) as follows [Norton, 1996, pp. 341–342]:

1) Assumption for *reductio* proof: The speed of fall of bodies in a given medium is proportionate to their weights.

2) From 1: If a large stone falls with 8 degrees of speed, a smaller stone half its weight will fall with 4 degrees of speed.

3) Assumption: If a slower falling stone is connected to a faster falling stone, the slower will retard the faster and the faster speed the slower.

4) From 3: If the two stones of 2 are connected, their composite will fall slower than 8 degrees of speed.

5) Assumption: the composite of the two weights has greater weight than the larger.[2]

[2] Ulrich Kühne remarked that a genuine Aristotelian would immediately reject this premise as false because he would understand 'weight' as '*specific* weight' – mass divided by volume (personal communication).

6) From 1 and 5: The composite will fall faster than 8 degrees.
7) Conclusions 4 and 6 contradict.
8) Therefore, we must reject Assumption 1.
9) Therefore, all stones fall alike.

The inference from 1 to 7 is deductive and thus hardly controversial. 8 does not follow deductively from 7, however, nor does 9 follow from 8. If two or more premises are jointly inconsistent, logic does not tell us which premise to reject. More than one resolution is possible. For 9 to be inferred deductively from 8, at minimum an additional premise is needed: 'The speed of the fall of bodies depends only on their weight' [Norton, 1996, p. 343]. This claim is, however, not part of the TE and, at any rate, would not have been accepted by Galileo's opponents.

The main issue with Norton's view is that there is no one way to reconstruct a TE as an argument, and whether or not its result can reliably be inferred from the reconstructed premises depends crucially on how the reconstruction is carried out. Even though it may well be true that *there exists* an argument that has the thought experimental result as a conclusion, this does not mean that the inferential power of the TE stems from its premises. This is in part due to the fact that not all premises that the argument needs in order to make the inference reliable will be acceptable to everyone prior to the TE. To the contrary: TEs often result in the acceptance of claims that appear dubious in the absence of the TE.

There are other accounts of thought experimentation that do not directly accept either of the two horns of the dilemma. Ernst Mach, for instance, was an arch-empiricist but accepted the usefulness of TEs because of their alleged ability to provide access to previously stored empirical information that is stored in our minds implicitly, in the form of intuitions [Mach, 1905]. In response, we have to ask why there are no other methods to make that knowledge explicit (for example, by thinking about the implications of explicitly held propositions). Thinking about economics TEs, we also have to challenge Mach's view that there is a great store of empirical information in our minds. While humans have lived in the same physical world since the beginning of time and it is therefore plausible to assume

that we inherit reliable intuitions about physical behaviours from our ancestors, it is absurd to think that we should have similar intuitions about the laws of modern capitalism.

Another view that does not accept the thought experimenter's dilemma is Thomas Kuhn's [Kuhn, 1981(1963); see also Gendler, 1998; van Dyck, 2003, Camilleri, 2012]. In Kuhn's view a TE can bring on a crisis or at least create an anomaly in an accepted theory and so contribute to a paradigm change. Thought experiments can teach us something new about the world, even though we have no new empirical data, by helping us to conceptualise the world in a new way. Whatever the virtues of this constructivist approach in general, it will not work for TEs in economics because their primary role is not to teach us something about our conceptual structures, as we will see now.

3. Thought experiments in economics: five types

There are, in my view, at least five types of TEs in economics. The first four aim to establish a causal claim. They can be illustrated in the following matrix.

Type of Causal Claim **Level of Economic Relation**	**(A) Singular**	**(B) Generic**
(1) Micro	Type 1A: The Lancashire Cotton Industry	Type 1B: Akerlof's Market for Lemons
(2) Macro	Type 2A: Fogel's Railroads and 19th Century U.S. Growth	Type 2B: Hume's Monetary Thought Experiments

Causal claims come in two different basic varieties: singular and generic. Singular causal claims concern the causes of individual outcomes such as 'Low interest rates in the early 2000's caused the 2007 financial crisis'. Generic causal claims concern relations between variables, for instance 'Increases in the quantity of money cause increases in nominal in-

come'. We find either type of claim at both the micro and the macro level, thus forming four types.

Types 1A and 2A are given by TEs that address 'What if?' counterfactuals about historical events. Their use goes back to a tradition that originates with Max Weber [Weber, 1949/1905; see also Tetlock and Belkin, 1996, and Tetlock, Lebow et al., 2006]. It is closely related to the counterfactual theory of causation according to which the truth of the counterfactual 'If C had not been, E would not have been' is sufficient for the truth of the causal statement 'C causes E'. To examine the causes of an outcome event E, the history of the world is imagined to have happened just as it did but with factors in E's past removed. Factors that make a difference to whether or not E obtains, that is, factors that are such that in the imaginary scenario in which these factors are removed the outcome does not obtain, are judged to be causes of E. Such counterfactual speculations address both microeconomic questions – such as the causes of the collapse of the Lancashire textile industry [Toms and Beck, 2007] – and macroeconomic questions – such as the causes of U.S. growth in the 19th century [Fogel, 1964].

A famous example for a TE that aims to establish a generic causal claim at the micro level (type 1B) is Akerlof's 'market for lemons'. Akerlof asks us to contemplate the following scenario [Akerlof, 1970, p. 489]:

Suppose (for the sake of clarity rather than reality) that there are just four kinds of cars. There are new cars and used cars. There are good cars and bad cars (which in America are known as "lemons"). A new car may be a good car or a lemon, and of course the same is true of used cars.

The individuals in this market buy a new automobile without knowing whether the car they buy will be good or a lemon. But they do know that with probability q it is a good car and with probability $(1 - q)$ it is a lemon; by assumption, q is the proportion of good cars produced and $(1 - q)$ is the proportion of lemons.

After owning a specific car, however, for a length of time, the car owner can form a good idea of the quality of this machine; i.e., the owner assigns a new probability to the event that his car is a lemon. This estimate is more accurate than the original estimate. An asymmetry in available information has developed: for the sellers now have more knowledge about

the quality of a car than the buyers. But good and bad cars must still sell at the same price – since it is impossible for a buyer to tell the difference between a good car and a bad car. It is apparent that a used car cannot have the same valuation as a new car – if it did have the same valuation, it would clearly be advantageous to trade a lemon at the price of new car [*sic*], and buy another new car, at a higher probability q of being good and a lower probability of being bad. Thus the owner of a good machine must be locked in. Not only is it true that he cannot receive the true value of his car, but he cannot even obtain the expected value of a new car.

Asymmetric information is thus judged to be a generic cause of low prices and traded volumes in markets where quality matters.

Margaret Schabas has given an extensive analysis of TEs of type 2B, which aim to establish generic causal claims at the macro level [Schabas, 2008]. In one example, David Hume has us contemplate the effects of an overnight doubling of the money stock in real economic quantities (and finds none). The methodology is not unlike that of the 'What if?' scenarios, but it lacks historical specificity and thereby establishes, if successful, a causal relation between variables rather than individual events.

The final type of TE (type 3) does not concern causal claims but claims about the nature and origin of economic institutions [Reiss, 2012]. They provide a fictional quasi-historical scenario that both explains the emergence of a new economic institution and (partially) justifies its existence. A famous example of this type is Menger's account of the origin of money [Menger, 1892].

Schabas argues that genuine TEs are uncommon in economics [Schabas forthcoming). She maintains that genuine TEs meet two criteria: they begin with a jarring counterfactual and they include an experiment; that is, an intervention. These criteria help, among other things, to distinguish TEs from models. Whereas models are idealised through and through, a TE 'is launched by a jarring, often bizarre counterfactual, but then restores some mental equanimity by introducing familiar objects to assist the mind of the experimenter as she reaches her destination' [Schabas, 2008, p. 2]. She recognises that there is continuity between models and TEs but sees clear cases of both categories at the ends of the spectrum.

I'd be happy in principle to accept the two necessary condition for a piece of reasoning to be a TE but I resist the conclusion that this shows that types 1A, 2A, 1B and 3 aren't genuine TEs and that (therefore) genuine TEs are uncommon in economics. All kinds of experiments that aim to establish causal conclusions – whether laboratory, controlled, field, randomised, natural or thought – proceed in essentially the same way: by systematic variation of a factor and the observation of the difference it makes (if any) on an outcome of interest. It does not matter whether the variation (or 'intervention', if the term is understood causally rather than in terms of human agency) is introduced deliberately by the experimenter or obtains naturally or is created hypothetically in the mind or *in silico*. Experiments of this kind can be analysed using Mill's method of difference or a probabilistic variant thereof.

TEs type 1A, 1B, 2A and 2B are all experiments in this sense. In 1A and 2A, the 'What if?' counterfactuals, the variation obtains between the actual course of events and a hypothetical course of events that is identical to the actual one except that one factor (or a small number of factors), has been removed and, possibly, the outcome chances in consequence. This is not unlike randomised experimentation where two groups are created by randomly dividing a sample population, treating one group and observing the difference, if any, in outcome. In a randomised experiment, variation obtains between two real experimental groups, and the outcome is probabilistic because many causal factors affect the outcome, not all of which can be controlled, but the logic of the two kinds of experiment is essentially the same. Hume's TE can be understood in this way too.

In 1B, both compared scenarios are hypothetical, but again, one factor is varied and the effect, if any, on the outcome is observed. In Akerlof's market for lemons, the variation is between a (hypothetical) situation with symmetric information and an otherwise identical (and equally hypothetical) situation with asymmetric information, and the difference this variation makes on the price and quantity of the traded cars is observed.

All these experiments start with a counterfactual – imagine key players in the Lancashire textile industry had decided differently, the railroad had been absent, an asymmetry in the information distribution were to develop,

the quantity of money were to double over night – and trace the consequences of the counterfactual on the outcome of interest. How 'jarring' the counterfactual is, is difficult to tell in the absence of a clear metric. (And as Schabas correctly observes, the literature following Lewis' seminal work on counterfactuals has made clear that no such metric is available.) What matters in my view is not how far away from actual reality the imagined scenario is, but rather whether it constitutes the right kind of counterfactual for the purpose at hand, which is causal analysis. It is important for example that the counterfactual is not implemented by varying a factor that has independent effects on the outcome [see Reiss, 2012]. If, say, Hume had us imagine a doubling of the money stock by an increase in foreign trade, the independent effect of foreign trade on the economy would confound the result of the TE. It is therefore that he has us imagine a doubling of the quantity of money 'by miracle'. So the counterfactual is jarring indeed, but this is not an end in itself. It is just one way to avoid the confounding of the result. Let me repeat: what matters is that the counterfactual is implemented by varying a factor that does not have an independent effect on the outcome of interest. One way of doing so is implementing it 'by miracle'. An alternative is to find a variable in the causal history of the factor of interest that is known by background information not to influence the outcome independently and change that [Reiss, 2012]. (As an aside, Jon Elster criticised Fogel's work on exactly this point. Fogel removes the railroad by miracle. Elster asks: How could this possibly have happened? He then argues that there is no scenario in which there exists no railroad but other things are essentially as they are. He concludes that Fogel's counterfactual is inadmissible because anything that could have implemented it would have had massive independent effects on the outcome, U.S. growth. See Elster, 1978).

Schabas calls Akerlof's Market for Lemons a *model* rather than a TE. I certainly agree that the hypothetical scenario involves more idealisations than a single jarring counterfactual. There are only four types of car (good/bad; old/new) and there is a definite probability that a given car is a good or a bad one. It is important to notice, however, that Akerlof arrives at his conclusion in two alternative ways. He first presents the informal

reasoning, the gist of which I have reproduced above. He then proceeds to reason in a much more formal, mathematical way. In so doing, Akerlof makes a large number of additional idealising assumptions. For instance, all agents are rational von Neumann-Morgenstern maximisers of utility, the quality of a good is all that matters to them, quality is a scalar that is distributed equally between 0 and 2, and so on. This piece of reasoning clearly involves a model. The inference is deductive from explicit premises. Whatever the earlier kind of reasoning is, it works in a different way. Much is left implicit. The driver of the inference is intuition. In other words, it works just like a TE (for the differences between economic models and TEs, see also Thoma, 2015).[3]

This leaves type-3 TEs, which I have called 'genealogical' [Reiss, 2012]. They are indeed different, in part because they do not aim to establish a causal effect. Schabas calls them mere 'narratives' and 'conjectured histories' that lack the crucial ingredient of experiment. I agree that they involve narratives, but all TEs do. More precisely, they involve a particular kind, namely, one that is about a hypothetical history. There is no intention to describe the actual course of history, nor a possible course of history. Menger has us imagine away all intentional creations of economic coordination in order to argue that money would have emerged unintentionally, as a by-product of people's propensity to barter.

Whether it involves an experiment depends on what precisely one means by the term. Not all experiments aim to establish a causal effect. Experiments in economics are, for instance, said to be used for 'speaking to theorists', 'searching for facts' and 'whispering in the ears of princes' [Roth, 1995]. None of these uses necessarily involves causal claims. Testing decision theories does not, for instance. 'Facts' can refer to phenomena other than causal effects. That people tend to 'overcontribute' in public goods games and the overcontribution 'decays' over the played rounds is

[3] The terminology is not used uniformly in the literature. Harro Maas refers to what I call 'economic models' as thought experiments [see Maas, 2007, Ch. 7] and Margaret Schabas classifies some of my thought experiments as models (see below). The main reason for cutting up the practices the way I do is that I am primarily interested in inference and models, and TEs, as I have argued, differ in their mode of inference very much.

an experimental phenomenon but not a causal effect [see Guala, 2005]. The 'whispering in the ears of princes' is often aimed at the creation of new institutions and again causal effects do not have to be involved. Interventions of the kind Schabas contemplates are essential to experiments but only to experiments that aim to establish causal effects.

Let me address the question whether Menger's narrative involves an experiment by a TE of my own. Suppose we implement, in the lab or in the field, the situation Menger contemplates. So we endow experimental subjects with goods of different kinds and incentivise exchange in specific ways. In particular, we make sure that the different goods have different degrees of what Menger calls 'saleability'. We also make sure that individuals cannot communicate about matters other than the relative prices and quantities of goods offered to be exchanged. Allowing individuals to play over sufficiently many rounds, we observe whether one good emerges as a standard of exchange – money in Menger's sense. Wouldn't we call this an experiment?

A final question I need to address in this section is whether or not TEs are common in economics. Schabas argues 'no', because she only regards TEs of type 2B as genuine. It is obvious that a view that also accepts the other types as TEs proper makes it much more likely that TEs are common. In my view, TEs are in fact extremely common in economics (especially, in theoretical economics), because theoretical work that involves reasoning with mathematical models is almost always preceded by more informal 'narratives' or 'stories' that establish the same conclusion by way of a type-1B TE such as Akerlof's. Of course, I have no knock-down argument to the effect that these narratives are indeed genuine TEs. What is clear is that if they are proper TEs, then TEs are very common indeed.

4. Thought experiments, coherent explanations and the thought experimenter's dilemma

The explanation of economic phenomena is certainly among the goals of economics.[4] Economists seek to explain specific economic outcomes

[4] While I sympathise with Friedman's view that prediction is the most important goal of economics[Friedman, 1953], it's clearly the case that economists also seek to explain out-

and events, recurrent patterns such as business cycles, and institutions. What all five types of TEs do, or so I want to argue, is to show that an explanation that has so far been ignored or thought to be flimsy is, in fact, perfectly coherent with commonly held background beliefs. This plays an important role in economic methodology because it *renders an explanation acceptable*. An explanation that was previously regarded (if at all) as uncompelling is now considered not only possible but *plausible*.

Let us pause for a moment to consider the notion of 'commonly held background beliefs'. The first thing to note is that the group that holds the relevant beliefs in common depends on the TE's intended audience. Depending on the context, this may be other economists or other economists within a certain tradition such as neoclassical economists, or economists of a certain political persuasion such as 'liberal' or 'conservative' economists, or policy makers or the population at large. What members of these groups find plausible or implausible differs of course a great deal, and a TE's success will crucially depend on aligning its assumptions with the actual beliefs of the intended audience. All TEs considered in this paper are primarily targeted at other economists.

Another observation is that background assumptions, beliefs, information and knowledge are not always clearly distinguished in the literature, but of course these terms mean different things. In the present context, I prefer using 'background beliefs' to distinguish the term from the idealising assumptions one typically finds in mathematical models on the one hand, and knowledge or information which is typically held to entail truth or at least reliability, on the other. The idea is that the level of commitment to the content of the beliefs is intermediate between 'known to be false' and 'held to be true'. It refers to what the relevant group takes to be credible and not in need of challenge or probing at the moment.

In each case of the TEs of Akerlof, Menger and Hume, on which I shall focus here, there is a dominant explanation of a phenomenon of

---

comes, so it would be a mistake to regard prediction as its only goal. Economics is in fact characterized by a plurality of goals that, apart from prediction and explanation, also includes description, policy analysis and making contributions to debates about normative matters such as rationality, justice, wellbeing and many others. See Reiss, 2013.

interest that is challenged by the TE, replaced by an alternative hypothesis that appears counterintuitive initially but is then shown to be consistent with background beliefs in the TE. Akerlof is, in fact, very explicit about his target: '"The usual lunch table justification for this phenomenon [of a large price differential between a new and an almost new car]"', he writes, '"is the pure joy of owning a "'new'" car"' [Akerlof, 1970, p. 489]. This is an intuitive explanation: new cars are more expensive than old cars because people prefer to be the first owners of cars. The alternative Akerlof offers is initially highly unintuitive: how can something as ubiquitous as asymmetrically distributed information can have dramatic effects including the complete collapse of market exchange and large price differences? The TE shows, however, that the causal effect of asymmetric information coheres with widely held background beliefs – other than a specific distribution of information and a small number of innocuous situational features, we do not have to make any strange assumptions. An unusual candidate is thus rendered the prime suspect.

As a result, we do not only have a new (physically or economically) *possible* explanation, we have a coherent and thus plausible explanation. The explanation seems compelling because one understands, without the need to make highly idealising assumptions and complex calculations, how prices (and quantities) must drop when an informational asymmetry is introduced. The asymmetric information account appears – post TE – entirely natural [cf. Kühne, 1995].

The same is true of Menger and Hume. The salient alternative to Menger's explanation of the emergence of money is that money was introduced deliberately, by an intentional act of economic co-ordination. It is interesting to note that Menger refutes this alternative by means of empirical data. Menger's favoured hypothesis then could be established very quickly by disjunctive syllogism. Either money was created deliberately, or it emerged spontaneously. It wasn't created deliberately (if it was, there would be evidence of such a creation, but there isn't), so it must have emerged spontaneously. QED. But Menger does *not* stop here. He adds the TE in order to show that money as an unintended consequence is a perfectly natural (in the sense of 'not contrived') phenomenon. Hume's TE,

similarly, makes the neutrality of money appear not just an abstract possibility but a compelling truth.

It is this function of a TE to 'render plausible' an explanation that shows that Norton's argument view is mistaken. Of course it is possible, after the fact, to write down an argument that has a description of the phenomenon of interest as the conclusion. But the problem is that prior to the TE, not all premissesthe premises of the argument would have been accepted by the participants of the debate. 'Asymmetric information causes low prices and quantities', 'Money does not cause real quantities', and 'Money is an unintended consequence of exchange' are hypotheses and possible explanations of economic phenomena of interest with or without the TE. But with the TE, they are more than mere possibilia – they are acceptable explanations.

Certainly, acceptable doesn't mean true. At the end of the day, an explanation rendered plausible by a TE requires empirical confirmation. Perhaps it's the joy of owning a new car that drives second-hand prices down, money isn't neutral and was a deliberately created after all. But given the ease with which possible explanations of social phenomena can be thought up [Steel, 2004], the difficulty and costliness of experimentation and the ambiguity of observational research, a purely empirical approach that put all possible explanations on an equal footing and allowed selection on empirical grounds alone would make progress in social science very difficult.

How does the proposed account address the thought experimenter's dilemma? TEs oin this view do not provide new empirical data. But it does teach a new fact: viz. that a previously unthinkable – or unthought – hypothesis is coherent with background knowledge and, in fact, given one's background knowledge, highly natural. This is a result that cannot be reached by ordinary reasoning from premissespremises, but it does not require mysterious peeks into Plato's heaven.

References

Akerlof G., (1970), "The market for 'lemons': quality uncertainty and the market mechanism", *Quarterly Journal of Economics*, 84(3), pp. 488–500.

Armstrong D., (1983). *What Is a Law of Nature?*, Cambridge, Cambridge University Press.
Brown J.R., (2010), *The Laboratory of the Mind: Thought Experiments in the Natural Sciences*, London, Routledge.
Camilleri K., (2012), "Toward a constructivist epistemology of thought experiments in science", *Synthese*, 191, pp. 1697–1716.
Dretske F., (1977), "The nature of laws", *Philosophy of Science*, 44, pp. 248–268.
Ebenstein L., (2015, *Chicagonomics: The Evolution of Chicago Free Market Economics*, New York (NY), St Martin's Press.
Einstein A. et al., (1935), "Can quantum-mechanical description of physical reality be considered complete?", *Physical Review*, 47(10), pp. 777–780.
Elster J., (1978), *Logic and Society: Contradictions and Possible Worlds*, Chichester, John Wiley.
Fogel R., (1964), *Railroads and American Economic Growth*, Baltimore (MA), Johns Hopkins Press.
Friedman M., (1953), "The methodology of positive economics", *Essays in Positive Economic,* Chicago, University of Chicago Press.
Gendler T., (1998), "Galileo and the indispensability of scientific thought experiment", *British Journal for Philosophy of Science*, 49, pp. 397–424.
Guala F., (2005), *The Methodology of Experimental Economics*, Cambridge, Cambridge University Press.
Kincaid H., (2004), "There are laws in the social sciences", [in:] C. Hitchcock, *Contemporary Debates in Philosophy of Science*, Oxford, Blackwell, pp. 168–185.
Kuhn T., (1981 [1963]), "A function for thought experiments", [in:] I. Hacking, *Scientific Revolutions*, Oxford, Oxford University Press, pp. 6–27.
Kühne U., (1995), "Thought experiments and the inference to a coherent explanation", *10th International Congress of Logic, Methodology and Philosophy of Science*, Florence, Italy.
Maas H., (2007), *Economic Methodology: A Historical Introduction*, Abingdon, Routledge.
Mach E., (1905), "Über Gedankenexperimente", [in:] E. Mach, *Erkenntnis und Irrtum*, Leipzig, Verlag von Johann Ambrosius Barth, pp. 181–197.
Menger C., (1892), "On the origin of money", *Economic Journal*, 2, pp. 239–255.
Mises L. von, (1996 [1949]), *Human Action*, San Francisco (CA), Fox and Wilkes.
Norton J., (1991), "Thought experiments in Einstein's work", [in:] T. Horowitz, G. Massey, *Thought Experiments in Science and Philosophy*, Savage, MD, Rowman and Littlefield, pp. 129–148.
Norton J., (1996), "Are thought experiments just what you thought?", *Canadian Journal of Philosophy*, 26(3), pp. 333–366.
Reiss J., (2009), "Counterfactuals, thought experiments and singular causal analysis in history", *Philosophy of Science*, 76, pp. 712–723.
Reiss J., (2012), "Counterfactuals", [in:] H. Kincaid, *Oxford Handbook of the Philosophy of Social Science*, Oxford, Oxford University Press, pp. 154–183.

Reiss J., (2012), "Genealogical thought experiments in economics", [in:] J.R. Brown, M. Frappier, L. Maynell, *Thought Experiments in Science, Philosophy, and the Arts*, New York (NY), Routledge.
Reiss J., (2013), *Philosophy of Economics: A Contemporary Introduction*, New York (NY), Routledge.
Roberts F., (1993), "Thought experiments and social transformation", *Journal for the Theory of Social Behaviour*, 23(4), pp. 399–421.
Roberts J.T., (2004), "There are no laws of the social sciences", [in:] C. Hitchcock, *Contemporary Debates in Philosophy of Science*, Oxford, Blackwell, pp. 151–167.
Roth A., (1995), "Introduction to experimental economics", [in:] J. Kagel, A. Roth, *Handbook of Experimental Economics*, Princeton, Princeton University Pres, pp. 3–110.
Schabas M., (2008), "Hume's monetary thought experiments", *Studies in History and Philosophy of Science* Part A, 29, pp. 161–169.
Schabas M., (forthcoming), "Thought experiments in economics", [in:] J.R. Brown, *Routledge Handbook of Thought Experiments*, New York (NY), Routledge.
Steel D., (2004), "Social Mechanisms and Causal Inference", *Philosophy of the Social Sciences*, 34(1), pp. 55–78.
Tetlock P., Belkin A. (eds.), (1996), *Counterfactual Thought Experiments in World Politics: Logical, Methodological and Psychological Perspectives*, Princeton (NJ), Princeton University Press.
Tetlock, P. et al, (2006), *Unmaking the West: "What-If" Scenarios that Rewrite World History*, Ann Arbor, MI, University of Michigan Press.
Thoma J., (2015), "On the hidden thought experiments of economic theory", *Philosophy of the Social Sciences* Online first.
Toms S., Beck M., (2007), "The limitations of economic counterfactuals: The case of the Lancashire textile industry", *Management & Organizational History*, 2(4), pp. 315–330.
Tooley M., (1977), "The nature of laws", *Canadian Journal of Philosophy*, 7, pp. 667–698.
van Dyck M., (2003), "The roles of one thought experiment in interpreting quantum mechanics: Werner Heisenberg meets Thomas Kuhn", *Philosophica*, 72, pp. 1–21.
Weber M., (1949/1905), "Objective possibility and adequate causation in historical explanation" [in:] M. Weber, E. Shils, H. Finch, *The Methodology of the Social Sciences*, Glencoe (IL), Free Press, pp. 164–188.

Julian Reiss
Professor of Philosophy,
Durham University and Co-Director of the Centre of Humanities Engaging Science and Society (CHESS),
Department of Philosophy, Durham University,
50 Old Elvet,
Durham DH1 3HN,
phone +44 191 334 6543; fax +44 191 334 6551;
e-mail: julian.reiss@durham.ac.uk