

*This book review was published in Theoria volv. LXIV part 1 (1998): 108-118*

Review of

Soren Haggqvist's Thought Experiments in Philosophy

(Stockholm: Almqvist & Wiksell International, 1996)

I routinely claim that round squares are impossible. Someone non-routinely replied by drawing a picture:

---

ROUND SQUARE

(Side-view)

This is a depiction of a round square by virtue of the artist's intention that it represent a round square. Intention is crucial even in the geometer's depiction of normal square. After all, the sides of the diagram are not perfectly straight. Nor is it physically possible for them to be perfectly straight. An inscribed square is a picture of a square by virtue of the geometer's fiat.

Mental images of squares also owe their representational properties to intentions. So why not cut out the middle man and just stipulate whatever we wish to be possible as possible?

In philosophy, the middle man is usually a thought experiment. Thus a friend of thought experiments must show how they go beyond mere stipulation. A variety of answers are now on the market: mental models (Nercessian 1992), Godelian style perception (Brown 1991), evolutionary nativism (Sorensen 1992), historicist groundings in entrenched scientific practice (MaCallister, 1996), and assimilation to the epistemology of argument (Norton 1991).

Soren Haggqvist's well written Thought Experiments in Philosophy comprehensively examines proposals advanced prior to 1996. Anyone seeking the latest views on thought experiments could do no better than to start here. The wide range of positions are presented in an uncluttered, fluent fashion. Haggqvist nimbly defends several positions against a number of criticisms in currency. However, he also amplifies other criticisms and raises the level of debate further with novel objections of his own. Even those familiar with the literature on thought experiment will profit from the

expository and critical portions of Thought Experiments in Philosophy.

As a New Yorker, I have just exceeded my annual allotment of praise. Mindful of my municipal obligations, I turn to Haggqvist's positive account of thought experiment.

#### I. HAGGQVIST'S MINIMALIST ACCOUNT OF THOUGHT EXPERIMENTS

Soren Haggqvist's draws inspiration from Saul Kripke's approach to the problem of transworld identity. How do we know that <>(Nixon loses the election in 1968)? We cannot look at other possible worlds through a telescope. So how do we know that the actual Nixon who won in 1968 is identical to a man in another possible world who lost the 1968 election? Kripke replies that we know by fiat. There is no need for a shadowy epistemic link between the speaker and the constituents of other possible worlds. The speaker's partial descriptions plus his intention to be talking about Nixon make it the case that the speaker does pick out Nixon. Haggqvist spots a bargain:

Similarly, then, it may seem that the thought experimenter painting a scenario C is entitled to knowledge of C's possibility simply because he or she has stipulated that C. And in a sense, this is true. Stipulation is free: if you wish to stipulate something ever so grotesque, who's to prevent you? (146-7)

At this juncture, it may seem that Haggqvist will be overly permissive -- that he will make thought experiment all too easy. Nothing could be further from the truth.

For Haggqvist immediately draws attention to a sobering catch: the thought experimenter is obliged to make the stipulated proposition cohere with a complicated corpus of beliefs. This further requirement of consistency makes knowledge of possibilities far from trivial.

We systematically under-estimate how much revision this involves. For instance, when philosophers suppose that the universe doubled in size last night, they tend to think that the assumption is accommodated by merely multiplying all magnitudes by two. However, the world is governed by geometric relations such as the inverse square law for gravity. Doubling the size of objects would

not preserve their surface to mass ratios. So the only way that the hypothesis could be really unverifiable is that the doubling would immediately cause all observers to black-out and die.

Thought experiments are not as easy as they look! Haggqvist reviews five central cases taken from the recent literature: the brain in a vat, Twin Earth, Oscar's arthritis, the Chinese room, and Newcomb's problem. His in-depth scrutiny leads him to conclude that NONE of these thought experiments provide a "decisive, clear-cut counterexample" to the thesis each was intended to refute. The thought experimenter either falls into incoherence, takes refuge in vagueness, or adopts some question-begging assumption.

How much of a surprise is this? Haggqvist's central cases are intended to be representative (187). But one could have predicted that these thought experiments would fail to involve decisive, clear-cut counterexamples just from the fact that they are so heavily debated. A theory of physics experiments should not focus on the most recently debated experiments. That would skew the investigation toward the unsettled, question-begging, innovative experiments. Most experiments are not newsworthy specimens. They are drab work horses that do a modest job in routine ways.

Soren Haggqvist is open to the suggestion that a more thorough analysis might resuscitate one of the central cases. However, he embraces a quantitative principle that generalizes his skepticism about the central cases: the amount of belief revision needed to accommodate a supposition increases with that supposition's distance from our current beliefs (156). Since the most interesting thought experiments tend to have bizarre antecedents, they are apt to impose the heaviest demand on our cognitive capacities. Hence, they are the most apt to overwhelm our ability to test the counterfactual premises of a thought experiment. So although the argument supported by the thought experiment can be regimented as a valid argument, we cannot know that its counterfactual premises are true.

Haggqvist is optimistic about our knowledge of ordinary counterfactuals such as 'If The Titanic had struck the iceberg head on, it would have remained afloat'. Hence, he is (in principle) optimistic about the success of thought experiments that only use ordinary counterfactual premises. Haggqvist does not give a single example of a successful philosophical thought experiment. Nevertheless, I think Soren Haggqvist would accept many of the old

appeals to ordinary language involving mild hypotheticals (recall how J. L. Austin distinguished between mistakes and accidents with a tale about shooting donkeys). I gather he would also endorse most Gettier counterexamples and the thought experiments Derek Parfit (1984) uses to illustrate mistakes in "moral mathematics".

## II. SELECTIVE SKEPTICISM AGAINST BIZARRE THOUGHT

### EXPERIMENTS

Soren Haggqvist recognizes that hypothetical belief revision is apt to seem infeasible even for mild suppositions. In a deterministic world, small changes ramify. Indeed, even without an assumption of strict determinism, meteorologists contend with the "butterfly effect" in weather prediction. The butterfly's movement in Tokyo can make a dramatic difference in London's weather one week later. However, Haggqvist tries to smooth the way for mild conditionals by denying that his belief revision model is intended to be psychologically realistic.

However, qualifications introduced to spare mild hypotheticals from intractibility also spare bizarre hypotheticals. None of us is aware of making titanic revisions to our beliefs systems when

accommodating the strange antecedents of thought experiments. We make only as many changes as strike us as relevant. If it does not itch, don't scratch. Haggqvist shows a need to scratch for the five thought experiments he chooses to study. But he does not show a general need to scratch. For example, he does not present the five central cases as members of a random sample. They are illustrative rather than evidential. Compare them with case studies of people who broke their legs because they did not check whether their shoes were tied. This is an insufficient basis for the conclusion that we ought to check more frequently (or that we ought to walk only in shoes that do not require laces).

So the woes of the five central cases do not threaten the default assumption that differences in detail do not matter. In any case, Haggqvist appears to accept a corresponding default principle for mild counterfactuals such as 'If I had put my key in my left pocket, it would not have fallen through the hole in my right pocket'. If my key had been in my left pocket instead of my right pocket, then the key would have gotten there via an earlier event. That counterfactual event would have occurred only if multiple earlier counterfactual possibilities had been realized. This

branching of possibilities proceeds backwards indefinitely. Yet Haggqvist does not believe that this combinatorial explosion wreaks epistemological damage. The skeptic is not entitled to demand that we exhaustively examine all the possibilities.

The number of possibilities can exceed any threatened by a combinatorial explosion. David Lewis counterexampled Robert Stalnaker's limit assumption (that there is always a closest possible world) with counterfactuals involving uncountably many possible worlds such as 'If David were over two meters tall, then he could join the basketball team'. Somehow human beings cope with big numbers. Therefore, we should not be awed into quietude by sweeping vistas of possibilia. Skeptical arguments that appeal to the scale of belief revision resemble the statistical sophistries that creationists assemble against the evolution of complex organs such as the eye.

Haggqvist has affection for thought experiments. The author's warmth for his subject matter rises off the pages of Thought Experiments in Philosophy. But Haggqvist does not love them protectively, like his children. Hence, he does not search for ways

of rescuing thought experiments from the perils of long distance accommodation.

Suppose a reviewer of Haggqvist's book did love thought experiments like his children. Especially the bizarre ones. What might he do on behalf of his peculiar brood?

First our hypothetical reviewer would note that many bizarre counterfactuals are knowable by virtue of instantiating logical and semantic relationships between the antecedent and consequent:

- (1) If everything doubled in size last night, then my pencil would not be longer than my desk.
- (2) If solipsism were true, then there would not be more than one person who understands English.
- (3) If there were a proof that had infinitely many steps, then a finite being would not be able to survey it.

These relationships have been incorporated into published thought experiments. For instance, John Passmore (1965) used counterfactuals such as (1) to demonstrate the falsifiability of 'The universe doubled in size last night'. He imagines waking up to find

that soup cans are larger than trash cans, watermelons are smaller than grapes, pencils are longer than yard sticks. Passmore confesses that he would not know what to conclude overall, but one thing he could infer: It is not the case that everything exactly doubled in size last night.

The logical relations that underwrite (1)-(3) are trivial. There are others which are just as logical but more controversial:

(4) If there were only abstract objects (sets, numbers, Cartesian egos, God), then there could have been some material objects.

(5) If there were only two spheres and each was qualitatively identical to the other except for color, then distinct things could be made qualitatively identical by a change of color.

(6) If there were time travel, then there would be causal loops.

Logic rather than imagination prompts Alvin Plantinga (1974, 59) to deploy (4) against the Barcan formula,  $(x)[\Box Fx > \Box(x)Fx$ . The connection between the bizarre antecedent and its consequent ranges from loose to tight.

Second, our gedankenexperimente-ophile would note that a thought experiment can be bizarre in one respect while conservative in another. We know all too well what it is like to be a brain in a vat. The skeptical appeal to counter-possibilities is designed to preserve the familiar course of experience. Worlds in which dream away our lives, worlds in which left and right are reversed, worlds in which everything pops into existence five minutes ago, are intended to be observationally equivalent with the actual world.

The "distance" between possible worlds is sensitive to one's standard of similarity. The Ramsey test for conditionals and all of its descendants inherit this relativity. Recall Kit Fine's criticism of David Lewis' comparative similarity analysis of counterfactuals:

The counterfactual 'If Nixon had pressed the button there would have been a nuclear holocaust' is true or can be imagined to be so. Now suppose that there never will be a nuclear holocaust. Then that counterfactual is, on Lewis' analysis, very likely false. For given any world in which antecedent and consequent are both true it will be easy to imagine a closer world in which the antecedent is true but the

consequent false. For we need only imagine a change that prevents the holocaust but that does not require such a great divergence from reality. (Fine 1975, 452)

Simply supposing that the button is defective spares the world.

Thus Lewis appears saddled with 'If Nixon had pressed the button it would have been broken'. The conditional is repugnant because we generally weight similarity with respect to laws of nature more heavily than similarity with respect to particulars. However, sometimes the weights are reversed in a way that makes the conditional acceptable. Consider a time travel scenario in which a maniacal Nixon goes back in time. Since the past is fixed, time travel stories can only be coherently told by pre-embedding the actions of time travellers into a history that closely resembles actual history. Hence, we prefer to conclude that button would have been broken.

Third, the bizarreness of many strange counterfactuals is eliminable. We wisely trade away realism in favor of radically simplified scenarios to ease calculation and to underscore relevant

variables. Counterexamples to average utilitarianism use absurdly low populations for the sake calculative convenience.

Bizarreness is also no problem when we deploy the thought experiment critically, against another supposition that is itself positioned at some distant point in modal space. For instance Nietzsche's eternal recurrence thought experiments were refuted by Georg Simmel's use of a universe that contained nothing but three wheels rotating infinitely at different rates. Our thinking about bizarre settings can also be made reliable by regimenting our description so that only a few variables are relevant. We can then take advantage of our ability to recurse.

Science and mathematics have plenty of bizarre thought experiments: Shrodinger's cat, Einstein's accelerated twin, Poincare's possibility proofs for non-Euclidean geometry, topological scenarios in which my right shoe is converted into a left shoe by being flipped through a fourth dimension, etc. Popularizers of science eagerly disseminate these to the public. Typically, they are not presented merely as expository aids or "just so" stories. They are intended as evidence just like experiments are presented as evidence.

Admittedly, the experiments and thought experiments that appear

in textbooks are partly selected for their pedagogical value. But this is compatible with their status as evidence. Parallel study of scientific thought experiments would have moderated Haggqvist's skepticism.

### III. NOTHING SUBSTANTIVE CAN BE KNOWN BY STIPULATION

Previous commentators on thought experiments have struggled to explain how thought experiments could yield substantive modal knowledge. One of the allures of Haggqvist's deflationism is that it bypasses this substantive modal epistemology in favor of a stipulative account. All the hard work gets shifted to the thinker's belief revision faculty. Nevertheless, belief revision cannot do the job alone because thought experiments purport to establish alethic possibilities. Their normal point is not to establish an "epistemic possibility". Showing that C is compatible with what I know does not show that it is really possible. The role of stipulation in Haggqvist's account is to furnish (conditional) knowledge of the thought experiments objective possibility. The "Achilles Heel" of thought experiment is discharging the conditional. Stipulation is

trivially easy. Accommodation is formidable. So says Soren Haggqvist.

This division of labor makes a mystery of all the effort thought experimenters put into "painting a scenario". Thought experimenters never brutally stipulate that  $\langle \rangle C$ . Instead, they craft scenarios that are intended to compel assent. For instance, a friend reported that he once tried to establish the possibility of a round square by imagining a small square on the surface of a sphere. The square grows and grows until it covers the whole surface of the sphere -- at which point it is both round and square. Although I share my friend's later reservations about this thought experiment, it is not a long-winded stipulation.

The thought experiment also contrasts with fictions that simply assert or presuppose that round squares are possible. If a Star Trek episode introduces round squares without any explanation, then the story provides no evidence that round squares are possible. We will say that 'Round squares are possible in the Star Trek episode' but that no more implies that round squares are possible than 'Round squares are possible according to Gene Rodenbury'. The thought experimenter wants you to assent to the

possibility of C itself rather than to 'C is possible according to the thought experimenter'.

So my objection is that stipulation is epistemically impotent. I agree that an abbreviatory definition can provide verbally novel knowledge by expressing old truths with new sentences. But thought experiments aim to do more than pour old wine into new bottles.

Stipulation cannot produce novel modal knowledge by literally creating new possibilities. What is possible is necessarily possible. Stipulations can change which sentences represent which propositions. The Indiana legislature once passed a law declaring that pi would henceforth equal 3. If the law had been obeyed, speakers would have truthfully uttered true sentences that sounded just like 'Pi equals 3'. But the statute would not have made pi equal 3.

Thought experiments bear a stronger analogy to geometrical diagrams rather than paintings. Reasoning from drawn figures was long known to be susceptible to equivocation, question-begging, and other fallacies. These can be circumvented with the formal logic developed in the last century. The superiority of formal methods

has led many professional mathematicians to demote diagrammatic proofs. Diagrammatic demonstrations are now widely regarded as merely an aid to discovery, as pedagogical tools, or at best, as proof sketches rather than proper proofs. The mathematical controversy over this demotion mirrors the philosophical debate about thought experiments.

A geometer who graphically demonstrates that the diagonal of a square must be longer than any of its sides makes numerous stipulations. But the stipulations ("Call this line A", "Let B bisect C", etc.) are mere tags designed expedite conversation. That is why one cannot object to a proof by refusing to acquiesce to a stipulation.

The point is brought out by J.E. Littlewood (1953):

Schoolmaster: 'Suppose  $x$  is the number of sheep in the problem.'

Pupil: 'But, Sir, suppose  $x$  is not the number of sheep.'

(I asked Prof. Wittgenstein was this not a profound philosophical joke, and he said it was).

Stipulations can be appraised on pragmatic grounds -- they can be wasteful, misleading, etc. But they do not play the role of challengeable premises.

If stipulation really did furnish substantive knowledge of possibilities, then it would give us knowledge of co-possibility. And if I can stipulate my way to co-possibility, then I can stipulate my way to consistency. Why accommodate?

One answer is that this kind of stipulation just changes the topic. In a trivial way, I can render 'Jack is round' compatible with 'Jack is square' by changing the meanings of 'round' or 'square'. But the task is to establish the consistency of the original statements. Given this constraint of meaning preservation, stipulation cannot give us knowledge of ' $2 + 2 = 5$ '. There is no accommodation under which ' $2 + 2 = 5$ '.

But why not just stipulate this constraint away? Just declare that the topic stays the same even if the meaning changes! Who is to be the master, words ('change', 'the', 'topic') or we inventors of words? In Beyond Good and Evil, Nietzsche asserts that we are the beings who must define ourselves.

Genuine philosophers, however, are commanders and legislators: they say, "thus it shall be" . . . . Their "knowing" is creating, their creation is a legislation, their will to truth is -- will to power. [# 211]

If conventional stipulation does not permit us this liberty, then why not stipulate a new meaning for 'stipulate' that would confer the ability?

The correct answer to this heaping of fanaticism upon fanaticism is that changing the topic as to what counts as 'changing the topic' still changes the topic. But once we restrict ourselves to meaning preserving stipulations, we have in effect accepted the requirement that any new system be a conservative extension of the previous system.

If Soren Haggqvist were to accept this requirement, he would need another bridge across the gap between epistemic possibility and alethic possibility. For instance, he might have a presumption of modal accuracy: If the thought experimenter believes  $\langle \rangle C$ , then probably  $\langle \rangle C$ . But why accept this principle? Is it equally plausible for all values of C? Why should we be reliable indicators of what

matters that transcend the actual world? Answers to these questions would force Haggqvist to revisit the substantive theories of modal knowledge that he so ably criticized.

#### REFERENCES

Brown, James (1991) The Laboratory of the Mind (London: Routledge).

Fine, Kit (1975) "Critical Notice of Counterfactuals" Mind 84: 451-458.

Littlewood, J. E. (1953) Littlewood's Miscellany ed. Bela Bollobas (Cambridge University Press, 1986, orig. 1953).

McAllister, James (1996) "The Evidential Significance of Thought Experiment in Science" Studies in the History and Philosophy of Science 27/2: 233-250.

Nercessian, Nancy (1992) "How Do Scientists Think? Capturing the Dynamics of Conceptual Change in Science" in Cognitive Models of Science (Minneapolis: University of Minnesota Press): 3-44.

Norton, John (1991) "Thought Experiments in Einstein's Work" in Thought Experiments in Science and Philosophy (Savage: Rowman &

Littlefield) ed. Tamara Horowitz and Gerald Massey (1991), pp. 129-144.

Norton, John (1996) "Are Thought Experiments Just What You Thought" Canadian Journal of Philosophy 26/3: 333-366.

Parfit, Derek (1984) Reasons and Persons (Oxford: Clarendon Press).

Passmore, John (1965) "Everything has just doubled in size" Mind 74: 257.

Plantinga, Alvin (1974) The Nature of Necessity (Oxford: Oxford University Press).

Roese, Neal J. and Olson, James M. (1995) What Might Have Been: The Social Psychology of Counterfactual Thinking (Mahwah, New Jersey: Lawrence Erlbaum Associates).

Sorensen, Roy (1992) Thought Experiments (New York: Oxford University Press).

Wilkes, Kathleen, Real People: Personal Identity without Thought Experiments, (Oxford: Clarendon Press, 1988).