Idealization and exemplification as tools of philosophy

Tommi Lehtonen
Abstract

Imaginary stories and thought experiments are often used in philosophy to clarify, exemplify, and provide evidence or counterevidence for abstract ideas and principles. Stories and thought experiments can illustrate abstract ideas and can test their credibility, or, at least, so it is claimed. As a by-product, stories and thought experiments bring literary, and even entertaining, elements into philosophy.

Even a short survey of philosophical literature quickly shows that idealization and exemplification are the main approaches in the use of thought experiments. The aim of idealizations is to conceptualize and condense the central or relevant aspects of complex realities and to make those conceptualizations testable in the “laboratory of the mind.” What aspects are considered relevant depends on the point of view and the aim of the discussion. The aim of exemplifications is to make tangible and justify abstract ideas. Problems with idealization and exemplification include the risk of the loss of information on the one hand and the risk of too much irrelevant information on the other. This paper examines and evaluates these and other risks related to the use of imaginary stories and thought experiments in philosophy.

Keywords: thought experiments, imaginary stories, idealization, exemplification, philosophy
Introduction

Imaginary stories and thought experiments are often used in philosophy to clarify and exemplify abstract ideas and principles. However, an even more important task for thought experiments is that they can provide evidence or counterevidence to support or discredit claims and counterclaims. Thought experiments are also used for conceptual analysis and the forming of concepts. For these last-mentioned tasks (i.e., evidence providing and conceptual analysis) thought experiments are used as “devices of framing and persuasion” (Gendler 2010, 128). According to Tamar Szabó Gendler, a successful and persuasive thought experiment makes people to “represent relevant non-thought experimental content in light of the thought experimental conclusion” (Gendler 2010, 6). Thus, thought experiments can help to see the reality in a certain way and from a certain perspective.

Thought experiments are an ancient method in investigation and argumentation, although it was not until the end of the nineteenth century that the term “thought experiment” was coined for philosophical discussion by Ernst Mach. Some well-known examples of philosophical thought experiments include Zeno’s paradoxes (such as “Achilles and the Tortoise,” and the “Arrow”), Hilary Putnam’s Twin Earth argument, John Rawl’s veil of ignorance, and the so-called ticking time bomb scenario. Thought experiments are thus used in different fields of philosophy, such as metaphysics, epistemology, and ethics.

One often-mentioned advantage of thought experiments is that they can illustrate (or should we say visualize or make tangible) abstract ideas. They can also test the credibility of philosophical ideas and hypotheses, or, at least, so it is claimed. As a by-product, stories and thought experiments bring literary, and even entertaining, aspects and elements into philosophy.

It should be noted that thought experiments are also used in other disciplines than philosophy. However, this article focuses solely on the philosophical use of thought experiments, while the use of thought experiments in natural sciences is excluded from consideration, with the exception of a short mention of Galileo Galilei. A central difference between thought experiments in philosophy and natural sciences is that in the latter the scope of possibilities is limited to the actual world (and possibly its physically nearest counterparts), whereas philosophical thought experiments have a much wider scope of possibilities: thought experiments in philosophy can in principle appeal to any possible or even impossible scenario. Thus, philosophical thought experiments are free from the demand and limits of practical realizability and empirical testability. This (rather than the alleged fact that philosophical reflections cannot answer empirical questions) can be considered the feature that differentiates philosophy (and formal sciences such as mathematics and logic) from other disciplines. Thus, on the one hand, thought experiments can be heuristic and
can help to solve not only conceptual but also empirical problems (as was the case when Galileo showed with a thought experiment that all bodies fall at the same speed). On the other hand, the picturesque features of thought experiments can be redundant and can hinder understanding and effective problem solving (Norton 1996).

In the following I will address and illuminate (a) the aims and advantages of the use of thought experiments in philosophy and (b) the possible risks related to the use of thought experiments. I try to alleviate the concern presented by some scholars, such as Kathleen Wilkes and Daniel Dennett, that thought experiments provide only (or to a great extent) disinformation.

**The concept of thought experiment**

Thought experiments are used as methodological instruments for philosophical inquiries. In order for a thought experiment to be a useful and serviceable instrument of inquiry, it must firstly have an intelligible and argumentative content. As regards the argumentative function of thought experiments, their job is to support or refute claims. It has even been claimed that all thought experiments are arguments, that is, ‘wholes’ consisting of premises, background assumptions, and conclusions (Norton 1996, 354). Accordingly, the goodness or badness of thought experiments depends on their logical validity, credibility, and persuasiveness.

Michael Bishop neatly defines a thought experiment as a mental representation of an experiment in which a result is derived by a process of reasoning. The reasoning used in thought experiments is not merely analytic, but also synthetic since it employs substantive assumptions about how the world works (Bishop 1998, 20). The assumptions can include, for example, laws of nature, causality, and psychological laws.

Bishop and Elke Brendel, among others, emphasize that thought experiments are indeed a certain kind of experiment, even though they are based on no new empirical data. In a manner similar to real (i.e., empirical or practical) experiments, thought experiments also study the dependency of variables on other variables and the dependency of actions on other actions. In addition, thought experiments also depend on some background assumptions or background theory (Brendel 2004, 91). However, in contrast to real experiments or simulations, the supposed outcome of thought experiments is not open (or unsettled without empirical evidence) but can be apprehended solely by the intellect.

Thought experiments usually have a practical justification. They are resorted to if real experiments are difficult or impossible to be carried out; for example, for physical, technological, or financial reasons. A thought experiment can also be the only morally permissible means of testing the boundaries of our ethics (think, for
example, of the ticking time bomb scenario and the acceptability of torture in extreme circumstances).

Some thought experiments could, at least in principle, be realized, but not all because many thought experiments depend on counterfactual, fictional, or idealized assumptions. Thus, thought experiments are in some sense “expeditions to possible worlds.” Roy Sorenson sees that such expeditions basically have a negative function: thought experiments aim to refute claims that may have appeared credible (or at least possibly true) at first (Sorensen 1992, 135). Accordingly, Sorensen calls thought experiments “alethic refuters.” However, this view is one-sided and dismissive. It is easy to see that thought experiments can also have a positive and constructive function.

Elke Brendel distinguishes the following functions and purposes of thought experiments: some thought experiments prove that certain theories or concepts involve contradictions; thought experiments can also provide supporting evidence for a theory or a concept; some thought experiments can illustrate a complex or abstract position, and some thought experiments detect the vagueness, or borderline cases, of a concept (Brendel 2004, 92). Especially in philosophy, thought experiments are used for conceptual analysis. Such thought experiments introduce us to situations in which we have to decide whether the given situation can be treated as a case of correct application of the concept in question (think, for example, of the paradox of the heap: is it true for all heaps of sand that removing one grain from it still makes a heap?) (Brendel 2004, 102). In Brendel’s view, thought experiments can thus be used positively to establish a view or theory, or they can be used negatively to refute a claim or to undermine a theory. In short, a thought experiment can be constructive or destructive (or even both).

In the light of the aforesaid, thought experiments are devices of the imagination by means of which the nature of things or concepts can be investigated. The main point in such investigations is that we seem able to get a grip on both physical and conceptual reality just by thinking (i.e., by reasoning, remembering, and imagining). Obviously, the truthfulness and helpfulness of thought experiments depends, among other things, on how correct and relevant the assumptions made in them are.

According to the Platonic view, philosophy is a maieutic art, the “midwifery of ideas.” The idea here is that the philosopher does not instill ideas but rather helps others to give birth to ideas with which they are already pregnant. An advocate of this view might claim that thought experiments are based on and make use of our store of inborn knowledge. Contrary to this view, Tamar Szabó Gendler argues that thought experiments can teach us something new about the world by helping us to reconceptualize the world in a better way. Thus, Gendler gives thought experiments an important role especially in the development of new conceptual tools. According
to her, the powers and limits of using thought experiments can be traced to the fact that when the contemplation of an imaginary scenario gives us new knowledge, it does so by forcing us to make sense of exceptional cases (Gendler 2000). Thus, in Gendler’s view, a good and informative thought experiment does not necessarily need to be as realistic as possible, but can also be fanciful. However, we will see that this view is controversial and debatable.

**Idealization and exemplification**

Above we have paid attention to the fact that thought experiments can be used constructively or destructively, to support or to refute claims. Even a short survey of philosophical literature quickly shows that idealization and exemplification are the main approaches in which the constructive and destructive functions of thought experiments are realized. In the following I will outline the central features, presuppositions, and problems of idealization and exemplification. In the subsequent sections, I will discuss these problems in more detail.

The aim of idealizations is to conceptualize and condense the central or relevant aspects of complex realities and to make those conceptualizations testable in the “laboratory of the mind”. Thus, idealizations aim to present something as better, more perfect, or simpler than in reality. Idealizations are associated with the presupposition that we can adequately, in a relatively unproblematic way, discern aspects that are relevant for the purpose of an analysis from those that are not. This discernment is dependent on the point of view or the aim of the analysis. Accordingly, the differentiation between relevant and irrelevant aspects is not unproblematic, but interest- or horizon-dependent. The obvious danger is that an idealization is circular or self-fulfilling (Dancy 1985). Thus, one might expect that philosophers get the results they desire by using suitable idealizations purposefully. Consequently, doubts arise whether thought experiments are really useful and illuminating, or somewhat loaded and manipulative. In order to alleviate this concern, the critical and self-reconstructive nature of philosophical work must be emphasized. Accordingly, philosophers and other scholars, as members of the scientific community, can also criticize and problematize the idealizations presented by themselves. Criticism can be directed, for example, to the hidden background assumptions and conceptual implications related to idealizations.

The aim of exemplifications is to make tangible (‘to expand on,’ ‘to spread out’) abstract ideas, that is, to untangle abstract ideas and construct particular stories. This story-creation is an optional and random activity, and its randomness can be a problem for philosophical inquiry, which aims to find necessary and essential features of phenomena and events. Thus, the question arises which stories among the
many alternatives are the best or the most useful exemplifications, and on what grounds.

To sum up, problems with idealization and exemplification include the risk of the loss (or concealment) of important information on the one hand and the risk of too much irrelevant information on the other. One might guess that too much information is eventually less harmful than too little information because too much information also includes, by definition, relevant pieces of information even if it may be difficult to discern the relevant information. However, one might reasonably say that idealization (i.e., abstraction) is probably a less problematic philosophical activity than exemplification or instantiation. The reason, it seems to me, is that, unlike exemplification, idealization is often based on factual data or observations received, for example, from texts or other research objects. I will defend this view in the subsequent sections.

Avoiding too high a level of generality

The discussion of the requirements and preconditions of a good and informative idealization and exemplification relates to the question of good communication in general. While discussing the requirements of good communication, philosophers and scholars of communication studies have paid a lot of attention to the importance of valid argumentation. In argumentation that aims at good communication, arguments are concrete in the sense that the central grounds used in justifying arguments are not on too high a level of generality or abstraction. Too many very general arguments are not sufficiently informative. Thus, they can be defective arguments. In order to avoid philosophical argumentation at too high a level of generality, it is helpful to use example stories and other illustrations. On the whole, the evaluation of the goodness of arguments in terms of the level of generality must be done by virtue of the context of argumentation.

In the following list, the main aims and tasks of idealization and exemplification are named:

<table>
<thead>
<tr>
<th>Idealization</th>
<th>Exemplification</th>
</tr>
</thead>
<tbody>
<tr>
<td>condensation</td>
<td>instantiation</td>
</tr>
<tr>
<td>simplification/abstraction</td>
<td>concretization/illustration</td>
</tr>
<tr>
<td>summary/“minimization”</td>
<td>expansion/“maximization”</td>
</tr>
<tr>
<td></td>
<td>specification/clarification</td>
</tr>
</tbody>
</table>
Both idealization and exemplification try to specify or clarify abstract concepts or complex realities. If successful, idealization and exemplification can help to enrich and define our concepts. In view of this, the ability to clarify our ideas and improve our understanding is the fundamental test of the success of idealizations and exemplifications. Note also that idealization and exemplification are interrelated: an idealization can be based on an example story and a story can exemplify a more general idea.

A mathematical illustration of the problems and risks of exemplifications

Let us take a mathematical illustration of the difference between idealization and exemplification. If we use the left-hand sides of the following equations as an analogy to an example story, and the right-hand side as an analogy to an abstraction or condensation of the story’s essence (i.e., its essential and philosophically interesting features), we can see something of the difference between idealization (i.e., condensation and simplification) and exemplification.

\[ \rightarrow \text{idealization (from left to right)} \]
\[
7 + 5 = 12 \\
4 \times 3 = 12 \\
1,000,012 - 1,000,000 = 12 \\
\sqrt{144} = 12 \\
(12 \times 2) \div 2 = 12 \\
(3 \times 2^2 \times \sqrt{2^2} + 63) - 75 = 12
\]

\[ \leftarrow \text{exemplification (from right to left)} \]

The above equations show that different exemplifications (in fact an infinite number of them) can yield (from left to right) the same result (12) and that can be called condensation or abstraction. However, all exemplifications are not equally enlightening or accessible; many of them include lots of confusing details and complex structures (like \((3 \times 2^2 \times \sqrt{2^2} + 63) - 75\)). In view of this, the risk of exemplification is to introduce obfuscation or confusing information.

The obvious temptation would be to conclude that we should use the abstract ideas themselves and renounce exemplifications, that is, detailed stories and thought experiments. Accordingly, we should be satisfied with what analogically could be illustrated by the following equation:
However, this approach entails tautology and is therefore not particularly attractive. So, at least for special purposes, telling a story of how we can get the same result (e.g., 12) in a particular way or in many alternative ways may be interesting and informative. In philosophy, thought experiments can thus be compared, although with obvious reservations, to equations and calculus.

The obvious reservations relate first to the fact that there is no mechanical calculus method to extrapolate the essence from a story. As we have already seen, what is relevant in a story depends on the purpose or point of view of the consideration. The results of mathematical equations are most often not, in this way, socio-constructive, as one might put it. Second, our background knowledge of the types of mathematical operations directs what equations we can derive from a number. Correspondingly, in the case of philosophical ideas, our background knowledge of language and of the context of discussion, among other things, directs what we can conceive that an idea implies. However, contextual features have even more effect on our philosophical understanding than mathematical reasoning, or so it is believed by those who consider mathematics as purely formal and content-free. George Boole (among others) seems to have thought so, when he wrote that “[o]n the point of the principle of true classification, we ought no longer to associate Logic and Metaphysics, but Logic and Mathematics” (Boole 2010, 13). However, a challenge to this formalist conception of mathematics occurs considering the fact that the consistency of a mathematical representation requires finding, “a ‘model’ (or ‘interpretation’) for the abstract postulates of a system, so that each postulate is converted into a true statement about the model” (Nagel & Newman 2001, 11). In a similar vein, W. W. Sawyer says, “Mathematics, then, is concerned with reasoning about clearly specified things or ideas. There is no reason why mathematical symbols should stand only for numbers (as in arithmetic, algebra, trigonometry) or for points (as in geometry). They can stand for anything.” (Sawyer 1982, 90.) This being so, then mathematical representation is not “content-free.” It is as much a representation of a content—a “model”—as is natural language. Now arises “one persistent source of difficulty”, namely that “the axioms are interpreted by models composed of an infinite number of elements. This makes it impossible to encompass the models in a finite number of observations; hence the truth of the axioms themselves is subject to doubt.” (Nagel & Newman 2001, 15.) This “persistent source of difficulty” is also fundamental to the status of thought experiments, since they, in a sense, represent, or correspond with, “a finite number of observations.” Therefore, one might argue, following Karl Popper, that falsificationism should be considered a basic strategy of philosophical
(as well as scientific) inquiry, because even one counterexample (or one thought experiment) can disprove a hypothesis (whereas proving a hypothesis may be more difficult or even impossible). This makes it appealing to argue that thought experiments should be primarily used to falsify claims, not to verify them.

**Twin Earth as an example of a thought experiment**

An appropriate way to test the aforesaid is to consider a known example. There are many alternatives, but let us use Putnam’s Twin Earth thought experiment as an example. It is an apt example because it has been discussed by many authors who have criticized the use of thought experiments in philosophy. As a very brief recap, the story is as follows:

Elsewhere in the universe there is a planet exactly like earth in virtually all respects. We can refer to that planet as Twin Earth. On Twin Earth there is an equivalent of every person and thing on Earth. The one difference between the two planets is that there is no water on Twin Earth. In its place there is a liquid that is superficially identical, but is chemically different, being composed not of H₂O, but of a different formula which we abbreviate (for the sake of argument) as XYZ. However, the Twin Earthlings call XYZ water. Finally, we set the date of our thought experiment to be several centuries ago, when the residents of Earth and Twin Earth would have no means of knowing that the liquids they called water were H₂O and XYZ respectively. The experience of people on Earth with water, and that of those on Twin Earth with XYZ, would be identical.

Now the question arises: when an earthling and his twin on Twin Earth say water do they mean the same thing? *Ex hypothesi*, their brains are molecule-for-molecule identical (with the exception of H₂O and XYZ molecules). Yet, at least according to Putnam, when Oscar says water, the term refers to H₂O, whereas when Twin Oscar says water it refers to XYZ. The result of this is that the contents of a person’s mind are not sufficient to determine the reference of terms they use. Instead, one must also examine the causal history that led to this individual acquiring the term. Consequently, “meanings just ain’t in the head,” to use Putnam’s phrase (Putnam 1975, 227).

I do not know exactly how and through what kind of process Putnam invented the Twin Earth story. It is said that Edmund Husserl developed a similar thought experiment nearly 70 years earlier. On the one hand, Putnam may have come to the opinion that semantic externalism and the causal-historical account of reference are true (whereas semantic internalism and the description theory of reference for theoretical terms is wrong) first and only then fabricated the Twin Earth story to support the idea. On the other hand, it is quite conceivable that Putnam had an intuitive idea of the story (or of a similar story) first and was only able to formulate
exactly what the story supposedly demonstrates afterwards. Be that as it may, both directions are possible: from the Twin Earth thought experiment to the abstract idea that “meanings just ain’t in the head,” or vice versa.

Above it was stated that, following Karl Popper, one might consider “falsificationist” use of thought experiments as primary. The Twin Earth case is an example of such a use. Another common “falsificationist” way of using exemplification is called reductio ad absurdum which I shortly discuss below.

**Reductio ad absurdum**

The history of philosophy offers various examples of the use of exemplification. The so-called reductio ad absurdum arguments are one of the special types of exemplification (Brendel 2004, 95). The core idea of the method of reductio ad absurdum is that if the conclusion of an argument cannot be accepted, the starting points from which the conclusion has been deduced must also be rejected.

The method of reductio ad absurdum is often used in Plato’s dialogues. For example, when the protagonist of the dialogue Theaetetus defines knowledge as observation, Socrates shows that if the definition presented is accepted, unwanted, absurd conclusions should also be accepted. The absurd conclusion would follow from the definition knowledge = df observation, as knowledge cannot be achieved in mathematics because the objects of mathematics, such as numbers, cannot be empirically observed. (In fact, this example is not from Plato.) Accordingly, some example stories function as a kind of reductio ad absurdum proofs. Thus, such stories show that from certain starting points unwanted conclusions follow.

If a reductio ad absurdum argument includes simplification (or exaggeration) of the view of an opponent and the refutation of this changed opinion, it is called a strawman argument. It is a distortion of the statement and thus faulty and morally questionable argumentation. In it, the statement or view of an opponent is manipulated or reformulated so that it seems absurd or ridiculous.

**Intuition pumps**

The counterfactual, fictional, or idealized assumptions made in many thought experiments can also be a source of other fallacies. Therefore, some philosophers believe that science-fiction thought experiments, such as Putnam’s Twin Earth, should be viewed with suspicion. It has been argued that when a thought experiment describes a state of affairs that is radically different from the actual one (or what we think it to be), our intuitions (i.e., instinctive insights) become unreliable. Thus, it has been thought that significant philosophical conclusions cannot be drawn from far-fetched science-fiction cases. Critics say that in the case of bizarre thought
experiments we simply do not know what to say when thinking about the situations depicted. However, thought experiments can be dismissed not only because of uncertainty due to ignorance. They can also be dismissed because they are based on implausible, incoherent, or inconsistent premises or because they involve other kinds of argumentative shortcomings like inconclusive judgments, illogical inferences, or circular inferences.

Kathleen Wilkes, among others, argues that thought experimentation engenders inconclusive and unconvincing results (Wilkes 1994). However, one may remark that counterfactual what-if questions have often been a catalyst for new discoveries. When what-if questions are posed, the researcher conceptually isolates and abstracts facts from their actual contexts and asks whether their absence or modification would alter the course or meaning of an event. Forbidding the invocation of imaginary and counterfactual scenarios in philosophical argumentation would be very problematic in light of the development and actual practice of philosophy (of which I gave a few examples in the beginning of this paper). In addition, as we have seen, thought experiments have an important negative function. For example, thought experiments can draw out a contradiction in a theory and thereby refute it. They can also show that a theory is in conflict with other very powerful beliefs that we hold. This negative function may also provide inconclusive results, but even so a thought experiment may give grounds for suspecting the accuracy of a received opinion—a suspicion that is relevant and worthy of further research. Thus, thought experiments can play an important role in helping to reveal problems in our notions and theories, and in correcting our philosophical errors.

However, as was said above, there is also some cause for doubting the reliability of our intuitions and cause for being critical of them. In this vein, Daniel Dennett calls Twin Earth and other experiments like it intuition pumps (Dennett 1984, 17–18; see also Dennett 1984 and Brendel 2004, 98–99). By intuition pumps, Dennett means conceptual tools, ideas, and beliefs that direct or channel our views in the desired direction. Thus, he warns that thought experiments can lead us to quickly and uncritically jump to a conclusion that is unwarranted. According to Dennett, thought experiments play on a strong but ultimately illusory intuition. Unfortunately, he does not specify criteria to distinguish between incorrect and correct intuitions.

One may say that the persuasive power of the Twin Earth thought experiment relies on our turning a blind eye to certain aspects of the experiment in order for it to establish what Putnam claims it establishes. Thus, the Twin Earth thought experiment is set up in such a way that one’s intuitions will be purposefully directed in the desired direction. This account gives some cause for believing that Putnam first came to the opinion that “meanings just ain’t in the head” and only then invented the Twin Earth story to support the idea.
Basically, thought experiments turn into problematic intuition pumps if the implicit assumptions of the thought experiment are disguised and the scenario is outlined in a way that leads intuitively to a conclusion which is not (or not completed) supported by the premises. In view of this, a requirement for a proper thought experiment is that all relevant background assumptions are identified and explicated. (Note that what is relevant depends on the point of view of consideration.) Brendel criticizes Putnam’s Twin Earth thought experiment for omitting this requirement. According to Putnam, the mental states of the human beings and their Twin Earth doppelgangers are identical. However, he fails to explicate and justify how this can be so even though the chemical or molecular compositions of the two worlds are different (Brendel 2004, 99). For this reason, the Twin Earth thought experiment can be criticized for being an intuition pump.

**Conclusion**

Thought experiments can be used to support or to refute claims. Idealization and exemplification are the main approaches in which these constructive and destructive functions of thought experiments are realized. The aim of idealizations is to conceptualize and condense the central or relevant aspects of complex realities and to make those conceptualizations testable in the “laboratory of the mind”. The aim of exemplifications is to make tangible and justify abstract ideas.

Literary and fictional elements related to philosophical example stories and thought experiments are not as innocent as they may prima facie seem to be. They can accidentally or deliberately direct our attention to certain context- or horizon-dependent features and can cause us to neglect certain other important features. Thus, they can in some way manipulate our intuitions. In essence, the lesson of these remarks is that when using philosophical stories and thought experiments we should be asking: “What are the central background assumptions related to this thought experiment? Are the background assumptions and prerequisites sound and justified? And what do they entail? Do they entail something that contradicts the explicit premises?”
Bibliography


