

Capturing the Scientific Imagination

Fiora Salis and Roman Frigg

1.1 Introduction

Maxwell, when investigating lines of force, sets himself the task of studying “the motion of an imaginary fluid,” which he conceives as “merely a collection of imaginary properties” (1965, 159–160). Einstein explains the principle of equivalence by inviting the reader to first “imagine a large portion of empty space” and then “imagine a spacious chest resembling a room with an observer inside” (2005, 86). Maynard Smith asks us to “imagine a population of replicating RNA molecules” (quoted in Odenbaugh 2015, 284). In his study of the growth of an embryo Turing notes that “the matter of the organism is imagined as continuously distributed” (quoted in Levy 2015, 782). And in his investigation into the nature of contractual relations Edgeworth proposes to “imagine a simple case—Robinson Crusoe contracting with Friday” (quoted in Morgan 2004, 756).

These are examples of leading scientists appealing to the imagination. They do so talking about either a scientific model (SM) or a thought experiment (TE). So the imagination is seen as crucial to the performance of both. Philosophers concur. Brown presents one of Newton’s TEs as asking the reader to “imagine the universe completely empty” (2004, 1127). Laymon paraphrases TEs as “imagined but truly possible experiments” (1991, 192). And Gendler describes them as “imaginary scenarios” (2004, 1154). Weisberg reports that Volterra in his model “imagined a simple biological system” (2007, 208) and accepts that “modelers often speak about their work as if they were imagining systems” (2013, 48). Godfrey-Smith suggests we “take at face value the fact that modelers often take themselves to be describing imaginary biological populations, imaginary neural networks, or imaginary economies” (2006, 735), and he sees modeling as involving an “act of imagination” (2009, 47). Harré sees models as things that are “imagined” (1988, 121). Sugden regards models as “imaginary” worlds (2009, 5).

Cartwright understands modeling as offering “descriptions of imaginary situations or systems” (2010, 22). Frigg (2010), Levy (2015), and Toon (2012) present analyses that place acts of imagination at the heart of the practice of scientific modeling, and Levy submits that “the imagination has a special cognitive role in modeling” (2015, 783).

This enthusiasm notwithstanding, philosophers of science typically do not offer explicit analyses of imagination. It is, however, common to associate imagination with mental imagery.¹ This is not surprising given that the word “imagination” derives from the Latin *imago*, which means “image,” “portrait,” “icon,” and “sculpture.” In this vein Levy observes that “imagining typically involves having a visual or other sensory-like mental state—a ‘seeing in the mind’s eye’” (2015, 785). Brown regards performing a TE as “a case of seeing with the mind’s eye” (2004, 1132), he characterizes TEs as being “visualizable” (1991, 1), and he regards being “picturable” as a “hallmark of any thought experiment” (1991, 17). Gendler emphasizes that “the presence of a mental image may play a crucial cognitive role” in a TE (2004, 1154). Likewise, Harré sees the “imagining of models” as providing scientists with a “picture of mechanisms of nature” (1970, 34–35). And Weisberg attributes to Godfrey-Smith the view that scientists form a “mental picture” of the “model system” (2013, 51).

Those who hoped that this was going to be a rare occasion of philosophers agreeing with each other have gotten their hopes up too quickly. The veneer of harmony unravels as soon as we probe the nature of imagination and the role it plays in TEs and SMs. While some authors, most notably Gendler (2004) and Nersessian (1992, 1999, 2007), affirm the imagistic character of the imagination and see it as an asset in explaining how TEs and SMs work, most scientists and philosophers draw back as soon as the imagination is linked to mental imagery. Norton thinks that TEs “are merely picturesque argumentation” (2004, 1142). And Weisberg dismisses a view of SMs based on imagination as “folk ontology” (2013, ch. 3). Talking about the necessary statistical treatment of atomic phenomena within quantum mechanics, Bohr recognized “the absolute limitation of the applicability of visualizable conceptions of atomic phenomena” ([1934] 1961, 114). And Dirac famously proclaimed that “the object of physical science is not the provision of pictures” (1958, 10).

¹ An exception is Odenbaugh (2015, 287), who explicitly recognizes a propositional variety of imagination.

We now find ourselves in a paradoxical situation. On the one hand, the imagination is widely seen as having an important role to play both in TEs and SMs. On the other hand, the imagination is dismissed because of its allegedly imagistic character. But one cannot both dismiss the imagination as ill-suited for scientific reasoning and see it as being crucial to TEs and SMs. The way out of this predicament, we submit, is an investigation into the character of the imagination.

Fortunately, such an investigation does not have to start from zero. There is a rich and intricate literature in aesthetics and philosophy of mind about the notion of imagination. But there has been little, if any, contact between that body of literature and debates in the philosophy of science. We therefore review this literature in a way that makes it relevant to TEs and SMs, and we propose a novel taxonomy of varieties of imagination that helps philosophers of science to orient themselves in this jungle of positions. One of the core messages emerging from this review is that the association of imagination with mental imagery has been too quick: there are propositional kinds of imagination that aren't in any way tied to mental images. This indicates the way for a resolution of the paradox mentioned previously: we argue that SMs and TEs involve a specific kind of propositional imagination, namely, *make-believe*.

We begin the chapter by reflecting on the relationship between TEs and SMs. So far we have mentioned TEs and SMs in one breath, thereby suggesting that they can be treated side by side. First we argue that TEs and SMs indeed involve the same kind of imagination. Then we present the main arguments for and against the involvement of the imagination in TEs and SMs: Norton's on the con side, and Gendler's and Nersessian's on the pro side. Following that, we review the positions on the imagination in aesthetics and philosophy of mind and propose a classification of these positions. We analyze the arguments previously introduced with the instruments subsequently developed. We argue that imagistic imagination is unnecessary for the performance of TEs and use of SMs, and that a propositional kind of imagination is necessary. We examine what the different kinds of propositional imagination introduced earlier offer for an analysis of SMs and TEs, and we tentatively suggest that this imaginative activity is best analyzed in terms of *make-believe*. Then we briefly summarize our results and draw some general conclusions.

Before delving into the discussion, a number of caveats are necessary. The term "imagination" has many meanings. To avoid getting started on

the wrong foot, let us set aside those meanings that are not relevant to our questions. First, “imagination” is often used as a synonym for “creativity.” Something is said to be “imaginative” if it is new, original, groundbreaking, or innovative. Needless to say, great scientific achievements are imaginative in this sense. Yet not all imaginative activities involve creativity, and not all creative activities involve imagination. A student who studies field lines, the principle of equivalence, or the nature of contracts has to engage in imaginative activities, but these aren’t creative because she is merely asked to retrace the steps outlined by Maxwell, Einstein, or Edgeworth. The *creative imagination* emerges when our imaginative abilities intersect with creativity to produce a novel output of any kind.² The imaginary acts we are interested in can be creative but need not be.

Second, “imagination” is often used to refer to false beliefs and misperceptions. This popular figure of speech is of no systematic interest because there is no specific ability to falsely believe or misperceive something. Rather, there is an ability to believe and an ability to perceive, both of which can go wrong.³ There are two corollaries to this point. First, imagination can be about real objects. We can imagine *of* Putin that he is a gambler to explore certain underlying features of his personality. In this case Putin is the focus of imaginative activities that are directed at improving our understanding of him. Second, imagination is independent of truth and belief. As Walton points out, “imagining something is entirely compatible with knowing it to be true” (1990, 13). So, for example, when reading Tolstoy’s *War and Peace*, we imagine that Napoleon was ruined by his great blunders, which is something that we also know to be true.

Finally, a terminological comment. As it is common in the literature on imagination, we take “imagination” to refer to the mental attitude of the person who imagines something; we use the noun “imagining” for an act of imagination and “imaginings” as the plural for several such acts.

1.2 Models and Thought Experiments

Is there a force needed to keep an object moving with constant velocity? In a classic TE Galileo argued that the answer to this question was no (Sorensen

² See Gaut 2003, 2010 and the contributions in Gaut and Livingston 2003 for current discussions on the relation between creativity and imagination.

³ See Currie and Ravenscroft 2002, 9, for a similar remark on imagination and false belief.

1992, 8–9). Galileo asked us to imagine a U-shaped cavity, imagine we put a ball on the edge of one side, and imagine we let the ball roll down into the cavity. What is the trajectory of the ball? Galileo argued that it would have to reach the same height on the other side irrespective of the shape of the cavity. This is Galileo's *law of equal heights*. Of course Galileo realized that the ball's track was not perfectly smooth and that the ball faced air resistance, which is why the ball in an *actual* experiment does not reach equal height on the other side. So Galileo suggested considering an idealized situation in which there is neither friction nor air resistance and argued that the law was valid in that scenario.

Galileo then asks us to continue the TE and derive the law of inertia from the law of equal heights. The law of inertia says that a body either stays at rest or moves at constant velocity if no force acts on it. Now imagine a situation in which the U-shaped cavity is bent downward on the right side so that the cavity becomes flatter on that side while the height is still the same on both sides. According to the law of equal heights, a ball starting on top of the left side still eventually reaches the top of the right side, no matter how much you bend the cavity. We can now imagine a series of variations of this thought experiment in which the right side of the cavity is bent ever more—and in each of them the ball reaches the top of the right side. If we continue this series indefinitely, we reach a scenario in which the right side is bent down all the way so that it becomes horizontal. The law of equal heights still applies, and so the ball should eventually reach the height at which it started on the left. However, since the right side of the cavity is horizontal now, the ball can't move upward, and so it keeps moving forever. From this Galileo drew the conclusion that no force is needed to keep a ball moving with constant velocity, which is the law of inertia.

Now consider a variation of this situation. Our protagonist is Malileo, a presumed mechanical philosopher of the nineteenth century. Malileo masters Lagrangean mechanics and can solve even difficult equations. He doesn't trust any result that isn't proven mathematically, and so he's suspicious of Galileo's informal reasoning. To get a mathematically rigorous justification of the law of inertia he assumes, with Galileo, that the cavity is frictionless and that there is no air resistance. He assumes that the ball is a perfect sphere with a homogenous mass distribution and with a radius that is much smaller than the width of the cavity. He further assumes that the only force acting on the ball is linearized gravity (that is, he screens off electromagnetic forces, etc.). He then conceptualizes the cavity as a

conjunction of two half-segments of a parabola that meet at the vertex. The right segment's equation contains a parameter a controlling the inclination of the half-parabola (the smaller a is, the flatter the parabola). He then uses the machinery of Lagrangean mechanics to write down the equation of motion of a ball moving under the constraint of the cavity. He solves the equation. The solution still depends on the parameter a . He then takes the limit for $a \rightarrow 0$ and finds that in the limit the trajectory tends toward constant linear motion. This is formal proof of Galileo's result.

Malileo constructed a model of the cavity and the ball's motion. In fact, when telling Malileo's story it was difficult to avoid the word "model." It would have been more natural to say that he models the ball as an ideal sphere with homogenous mass distribution, that he models the cavity as a parabola, and so on. His construct is a bona fide SM, similar to other SMs such as the logistic growth model of a population or the ideal chain model of a polymer. This observation matters because the kind of imaginings that Malileo entertains are the same as Galileo's. Both imagine cavities and the motion of balls. For sure, Malileo also adds a mathematical description and uses a background theory (Lagrangean mechanics). But this does not detract from the fact that he imagines the same sort of objects in the same way as Galileo, who doesn't have the additional formal apparatus.

The conclusion we draw from this little scientific fairy tale is that insofar as imaginings are involved when a scientist performs a TE, these imaginings are of the same kind as the ones she has when working with a SM (and vice versa). Of course, the exact mental content is typically different. Malileo's mathematical expressions are not on Galileo's mind, but when Galileo and Malileo think about a cavity that can be flattened on one side and about a ball moving in it, they engage in the same kind of imaginative activity. This observation generalizes: TEs and SMs involve the same kind of imagination. The imaginative activities involved in SMs and TEs can be analyzed together.⁴

Views gesturing in the same direction have been voiced before. Harré submits that a "model is imagined and its behavior studied in a *gedanken*-experiment" (1988, 121–122), thereby putting SMs and TEs in the same category. Cartwright urges that models "are often experiments *in thought*" (2010, 19). Del Re, commenting on Galileo, observes that in

⁴ We here set aside reconstructions of SMs in terms of set theoretical structures (for a discussion of this view, see Frigg 2010). We agree with Weisberg (2013) that even those who think that the model-world relation is ultimately purely structural will have to admit fictional objects such as perfect spheres and unbounded populations at least as "folk ontology" into their understanding of models.

Gedankenexperimente we explore objects of an ideal world, and adds that “‘physical models’ applies to the objects of which this ideal world is made” (Del Re 2000, 6).⁵

1.3 Exorcism and Veneration

As we have seen, there are diametrically opposed positions on the nature and role of the imagination in philosophy of science. In this section we review in some detail the most explicit pronouncements on either side of the divide.

Norton advances a view of TEs as devoid of imagination. He characterizes TEs as picturesque arguments that “(i) posit hypothetical or counterfactual states of affairs, and (ii) invoke particulars irrelevant to the generality of the conclusion” (1991, 129). Condition (i) gives TEs their thought-like character, otherwise they would be mere descriptions of real states of affairs. Condition (ii) gives them their experiment-like character. The claim that TEs are arguments is motivated by Norton’s empiricism, the view that knowledge of the physical world derives from experience. Because TEs do not involve any new experimental data, “they can only reorganize or generalize what we already know from the physical world. . . . The outcome is reliable only insofar as our assumptions are true and the inference valid” (1996, 335).

Norton introduces two related theses. According to the *reconstruction thesis* (ReT), “the analysis and appraisal of a thought experiment will involve reconstructing it explicitly as an argument” (1991, 131). According to the *elimination thesis* (ET), “thought experiments are arguments which contain particulars which are irrelevant to the generality of the conclusion” (1991, 131), but “these elements are always eliminable without compromising our ability to arrive at the conclusion,” and therefore “any thought experiment can be replaced by an argument without the character of a thought experiment” (1996, 336). Norton’s ET can be interpreted in two ways. According to a weak interpretation, ET is a thesis about the nature of the *conclusion* of a TE, which is a general proposition that does not involve any reference to the specific elements of a TE. According to a strong interpretation, the irrelevant particulars can also be eliminated from the argument itself.

⁵ *Gedankenexperiment* is the German word for TE; sometimes it’s also spelled *Gedanken-Experiment*.

What is the role of the imagination in this framework? Norton barely mentions the word “imagination” and never explores the notion. However, when he talks about the “picturesque” character of TEs (1996, 2004), he seems to associate imagination with mental imagery. On other occasions he also seems to condemn imagination as irrational thinking, as when he writes that “empiricist philosophers of science . . . must resist all suggestions that one of the principal foundations of science, real experiments, can be replaced by the *fantasies of the imagination*” (1996, 335, italics added). So he seems to regard imagination as irrelevant both to the derivation of the outcomes of TEs and to their analysis and assessment.

Nersessian and Gendler defend different versions of the imagistic view against the idea that TEs are mere logical arguments involving propositional reasoning. While they do not discuss Galileo’s TE, their proposals entail that when performing this TE we form a perception-like representation of a U-shaped cavity and a ball rolling down into the cavity. Gendler claims that some TEs crucially require imagistic reasoning and that “the presence of a mental image may play a crucial cognitive role in the formation of the belief in question” (2004, 1154). To lend support to these claims she presents a series of examples from problem-solving contexts where similar imagistic abilities would be crucial. For example, she asks the reader to imagine whether four elephants would fit comfortably in a certain room and suggests that “presumably . . . you called up an image of the room, made some sort of mental representation of its size . . . , called up proportionately-sized images of four elephants, mentally arrayed them in the room, and tried to ascertain whether there was space for the four elephants within the confines of the room’s four walls” (2004, 1157).

Nersessian develops this approach to TEs by appealing to the literature on mental modeling and mental simulation.⁶ On her view, the performance of a TE involves the manipulation of a mental model within the constraints of a specific domain of scientific inquiry. A mental model (which is distinct from a SM) is a mental analogue of a real-world phenomenon. Accordingly, much of the work in Nersessian’s account goes into articulating the nature of mental analogues. She appeals to the distinction between two different kinds of mental representations enabling two different kinds of cognitive processes. On the one hand, there are linguistic and formulaic representations that enable logical and mathematical operations, which are rule-based and truth-preserving. These representations “are interpreted as referring to physical

⁶ See especially Johnson-Laird 1980, 1982, 1983, 1989.

objects, structures, processes, or events descriptively” (2007, 132). Their relationship to what they refer to “is truth, and thus the representation is evaluated as being true or false” (2007, 132). On the other hand, there are iconic representations, which include analogue models, diagrams and imagistic representations. They “involve transformations of the representations that change their properties and relations in ways consistent with the constraints of the domain” (2007, 132). For example, Nersessian asks the reader to think about how to move a sofa through a doorway and writes that “the usual approach to solving the problem is to imagine moving a mental token approximating the shape of the sofa through various rotations constrained by the boundaries of a doorway-like token” (2007, 128). Iconic representations enable the latter sort of processing operations, or *simulative model-based reasoning*. They “are interpreted as representing demonstratively” (2007, 132). And their relationship to what they represent “is similarity or goodness of fit. Iconic representations are similar in degrees and aspects to what they represent, and are thus evaluated as accurate or inaccurate” (2007, 132). Mental models are mental analogues of real-world phenomena. And mental analogues are iconic representations that cannot be reduced to a set of propositions.

In the next section we discuss positions on the imagination found in aesthetics and philosophy of mind, and based on the insights gained in this discussion we evaluate the positions introduced in this section. We argue that Gendler and Nersessian overstate the importance of the imagistic imagination, which we find to be unnecessary for the performance of TEs and the use of SMs. Norton’s account, by contrast, underplays the importance of the imagination. We argue that construing TEs as arguments presupposes a propositional kind of imagination, which we argue is necessary for the performance of TEs and SMs.

1.4 Varieties of Imagination

This section provides tools for a reevaluation of the role of the imagination in TEs and SMs by presenting positions from the rich and intricate literature on imagination in aesthetics, philosophy of mind, and cognitive science in a way that makes them applicable to problems in the philosophy of science. In doing so we also offer a novel taxonomy of imaginative abilities.

Central to accounts of imagination is the distinction between the *content* of a mental state and the *attitude* an agent takes toward this content. Different mental states can have the same content. One can believe that

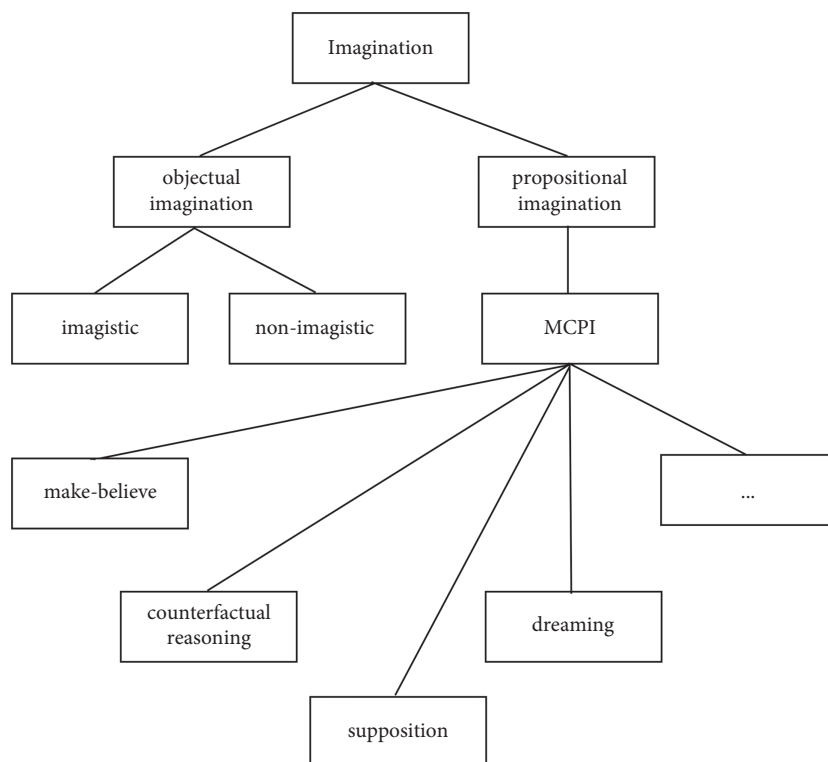


Figure 1.1 Varieties of imaginative abilities

there is a tree in the garden and one can imagine that there is a tree in the garden. Imagination and other states must therefore differ at the level of attitude. This said, a crucial distinction pertains to the kind of content toward which an imaginative attitude is taken. We can imagine *that there is a tree in the garden*, and we can imagine *a tree* in the garden. Whether we imagine a proposition⁷ or an object leads to the distinction between the two main varieties of imagination: *propositional imagination* and *objectual imagination*. Figure 1.1 shows the different accounts that we will discuss in this section along with their logical relations to each other to aid orientation.

⁷ Philosophers of language disagree about the nature of propositions. For the purpose of this chapter it suffices to say that propositions are the intersubjective objects of propositional attitudes, that they are the bearers of truth-values, and that they are expressed by using syntactically well-formed sentences.

1.4.1 Objectual Imagination

The objectual imagination is a mental relation to a representation of a real or nonexistent entity. One can imagine London or the fictional city Macondo, Napoleon or Raskolnikov, a tiger or a unicorn. Yablo characterizes objectual imagination as having referential content of the kind “that purports to depict an object” (1993, 27). Yet he emphasizes that depicting an object does not require forming a mental image of it, which is why we can imagine objects that are hard (or even impossible) to visualize. We can imagine a chiliagon (a thousand-sided polygon) even if we cannot form a mental image of it (1993, 27 n. 55). However, if we cannot form a mental image of a chiliagon, how can we imagine it without imagining *that* it is so-and-so? Yablo does not consider this issue, but Gaut offers a natural solution: “Imagining some object x is a matter of entertaining the concept of x , where entertaining the concept of x is a matter of thinking of x without commitment to the existence (or nonexistence) of x ” (2003, 153). Imagining a chiliagon simply amounts to entertaining the concept of a chiliagon.

In contrast with this somewhat minimalist view, a long philosophical tradition characterized objectual imagination as a kind of *imagery*: a relation between a subject and an image-like representation of an object (real or nonexistent). Different varieties of imagery experiences correspond to different sensory modalities. The most common is visual imagination, often referred as “seeing in the mind’s eye,” “visualizing,” or “imagining seeing.” Other modalities give rise to “imagining hearing,” “imagining feeling,” and so on. Colloquially, the term “mental image” is used to denote the *phenomenal* character of the imagery experience—that is, what it feels like to form a mental image. Scientists use the term in this pre-theoretical way when they report certain imagery experiences as the source of scientific discoveries. Kekulé’s famous introspective report of a reverie involving a snake-like figure closing in a loop as if seizing its own tail involves a mental image of this kind.⁸

The contemporary debate on mental imagery is vast, and there is disagreement on many foundational issues.⁹ Most of these issues can be set aside safely in the context of a discussion of SMs and TEs. Two issues are pertinent for our discussion: the nature of the representational format of mental images and the role of imagery in cognition.

⁸ See Shepard 1978 for more paradigmatic examples.

⁹ See Nigel 2014 for an excellent review.

Within the debate on the representational format, Kosslyn's (1980, 1983, 1994, 2005) *quasi-pictorial theory of visual imagery*, or *analogical theory*, has been influential in recent debates about TEs, and we therefore concentrate on it here. According to the quasi-pictorial theory, visual mental images have intrinsic spatial representational properties: they represent in a way that is analogous to the way in which pictures represent. But what is meant by a mental image having spatial representational properties? To pump intuitions, consider an example taken from an important experiment (Shepard and Metzler 1971). Subjects were presented with pairs of images showing three-dimensional objects from different angles, and they had to say whether the two objects were in fact identical. The experiment showed that the reaction time was a linearly increasing function of the angular difference in the orientations of the objects. Subjects reported that they had to form mental representations with spatial properties that allowed them to rotate the object in their mind and check whether some rotation would yield a view that was congruent with the second picture.

Kosslyn takes this to show that mental images have much in common with perceptual images.¹⁰ He offers the following analogy: perception is like filming a scene with a camera while at the same time watching the scene on-screen; mental imagery is like playing back on-screen what has been recorded previously. This view is backed by the fact that visually imagining something with our eyes closed activates 92% of the regions of the brain that are also activated when we visually perceive something similar. However, Kosslyn is quick to add that the analogy is not perfect in one crucial respect: imagistic imagination is not just a passive playback process. In fact, images are put together actively. This allows us to vary the setup we have perceived. For instance, we can move around, in our mind, the pieces of furniture in a room and imagine the room arranged differently. So imagistic imagination is informed but not constrained by what we perceive.

A time-honored philosophical tradition attributed a central role to mental imagery in *all* cognitive processes. This idea is usually traced back to Aristotle's claim that "the soul never thinks without an image" (1995, iii 7, 431a15–17), and it lived on in classical British empiricism. It was largely abandoned in the wake of influential objections by Frege ([1884] 1953), Ryle (1949), and Wittgenstein (1953). The dominant view nowadays is that

¹⁰ See "PhotoWings Interview: Stanford Cognitive Scientist Stephen Kosslyn—Mental Imagery and Perception," Vimeo, uploaded December 7, 2012, by PhotoWings, <https://vimeo.com/55140759>.

most thinking is sentential—or propositional—and non-imagistic. Fodor (1975, 174–194) recognizes that mental images play some role in cognition, but submits that their meaning—what they are images of or what they represent—must be determined by a description in a language of thought, or *mentalese*. Even modern proponents of Kosslyn's view do not attribute a central cognitive role to imagery, which is seen as deriving most or all of its semantic content from *mentalese*.

A dissenting voice is Barsalou's (1999), which has been influential in recent discussions about TEs. He proposes an alternative theory of perceptual symbols according to which cognition uses the same representational systems as perception. He distinguishes between what he calls "amodal" and "modal" symbols.¹¹ Amodal symbols are the not imagistic language-like symbols of *mentalese*. They are akin to words in that they are "linked arbitrarily to the perceptual states that produced them . . . Just as the word 'chair' has no systematic similarity to physical chairs, the amodal symbol for *chair* has no systematic similarity to perceived chairs" (1999, 578–579). Modal symbols, by contrast, are subsets of perceptual states stored in long-term memory. They are analogical because "the structure of a perceptual symbol corresponds, at least somewhat, to the perceptual state that produced it" (1999, 578). Barsalou emphasizes that modal symbols should not be identified with mental images,¹² but he conceives of modal symbols as closely related to traditional conceptions of imagery and as involved in our conscious imagery experiences. Unlike proponents of the quasi-pictorial view, Barsalou attributes a crucial role to modal symbols and claims that they are involved both in perception *and* in cognition.

Returning to our earlier distinction between attitude and content, it should be emphasized that objectual imagination cannot be *defined* in terms of the presence of mental images because mental images can accompany episodes of memory, belief, desire, hallucination, and more. What makes the deployment of a mental image an instance of imagination is the attitude we take toward the mental image. We may, for instance, suspend belief and not react to images (imagining a fighter jet flying at us does not make us run to the bomb shelter). What exactly the relevant attitudes are is an interesting

¹¹ The use of the term "modal" in this context has nothing to do with the use of the same term in modal logic. A modal symbol is one that pertains to the relevant sensory modality (e.g., visual modality, haptic modality, olfactory modality).

¹² His reason for this is that mental states may sometimes be active even when the agent is not conscious of them. Paivio (1986), however, suggests that mental images can be active even when we are not consciously aware of them.

question. However, an answer to this question does not matter for the discussion of TEs and SMs to come, and so we set it aside here (yet we do pay attention to attitudes in the context of the propositional imagination, and some of the insights gained there could be carried over, *mutatis mutandis*, to the context of objectual imagination).

1.4.2 Propositional Imagination

The propositional imagination is a relation to some particular proposition (or propositions). We analyze propositional imagination by first individuating a *minimal core of propositional imagination* (MCPI), which provides necessary and sufficient conditions for something to be an instance of propositional imagination. Different varieties of propositional imagination can then be distinguished by the further conditions they satisfy. Hence, each kind X of propositional imagination can be characterized by filling in the blank in the scheme

$$X = \text{MCPI} \ \& \ \underline{\hspace{1cm}}.$$

Three main features of the propositional imagination emerge from the literature. Taken together, these form MCPI.

First, we are not free to believe whatever we want, but typically we are free to imagine whatever we want.¹³ To believe that *p* is to hold *p* as true at the actual world, and whether the actual world makes *p* true or false is not up to us. To imagine that *p* does not commit us to the truth of *p*. We can decide freely what to imagine, and we can engage in spontaneous imaginative activities such as daydreaming where our imagination is not guided consciously. We refer to this feature as *freedom*.¹⁴

Second, propositional imagination carries inferential commitments that are similar to those carried by belief, hence manifesting *mirroring*.¹⁵ If we believe that Anna is human and that humans have blood in their veins, we infer that Anna has blood in her veins irrespective of whether Anna is real or fictional. The inferences we make may depend on background assumptions

¹³ See, e.g., Currie and Ravenscroft 2002; Nichols and Stich 2000, 2003; and Velleman 2000.

¹⁴ We here set aside the issue of *imaginative resistance* (Walton 1994), which is fraught with controversy.

¹⁵ See, e.g., Gendler 2003; Leslie 1987; Nichols 2004, 2006; Nichols and Stich 2000; and Perner 1991.

and on the specific aims and interests that direct our reasoning, but this is true in both cases.

Third, imagining that p does not entail believing that p . Typically, imagined episodes are taken to have effects only within the relevant imaginative context, hence manifesting *quarantining*.¹⁶ More generally, mental states of propositional imagination do not guide action in the real world. When watching a stage performance of *Othello* we may not want Desdemona to die, but only a hopeless country bumpkin would jump onstage to save the heroine. Quarantining does not imply that nothing of “real-world relevance” can be learned from an act of pretense. Dickens’s *Oliver Twist* mandates us to imagine that many orphans in London in the mid-nineteenth century were cruelly treated. We may well also believe that this was true. Such “exports” are, however, one step removed from the imagination.

In sum, MCPI consists of freedom, mirroring, and quarantining. We are now in position to discuss specific varieties of propositional imagination. We consider supposition, counterfactual reasoning, dreaming, daydreaming, and make-believe. There is no claim that this list is exhaustive, but we submit that it contains the main varieties needed to discuss SMs and TEs.

Supposition

Scientists often introduce SMs and TEs via the use of expressions such as “suppose,” “assume,” and “consider.” These are typically used interchangeably and so we regard them as synonyms, at least in the context of SMs and TEs. If a description of a model starts with “Suppose that three point masses move quantum-mechanically in an infinite potential well . . .,” then we are invited to engage in a particular imaginative activity. So when scientists introduce TEs and SMs by inviting us to suppose something, they typically invite us to imagine something without any commitment to its truth. The same use of the term can also be found in formal logic, where we sometimes assume a proposition in a process of inferential reasoning without any commitment to its truth—for example, when we suppose that p in a proof by reductio.

Supposition satisfies the three features of MCPI. We can suppose that most sentient life in the universe will soon be destroyed by an asteroid hitting the earth (freedom). The inferences we draw from this are similar, in relevant ways, to the ones we would make if we were to assume an attitude of belief

¹⁶ See, e.g., Gendler 2003; Leslie 1987; Nichols and Stich 2000; and Perner 1991.

(mirroring). Yet we do not take action to protect the well-being of our family and friends (quarantining).

There are two standard features of supposition that typically distinguish it from other varieties of propositional imagination: *epistemic purpose* (EP) and *rational thinking* (RT). These features fill the blank in our schema:

Supposition = MCPI & EP & RT.

Supposition is typically associated with ratiocinative activities aimed at specific epistemic purposes. By “ratiocinative activities” we mean the sort of activities wherein a consequence is derived from certain premises via deductive or inductive reasoning. By “epistemic purpose” we mean that supposition is usually aimed at gaining knowledge.

Some might doubt that supposition is a species of propositional imagination. In this vein Peacocke claimed that imagination is a “phenomenologically distinctive state whose presence is not guaranteed by any supposition alone” (1985, 20) because “to imagine something is always at least to imagine, from the inside, being in some conscious state” (1985, 21). This distinction is artificial since many of our imaginings do not involve any imagining from the inside, as when we imagine that Anna Karenina is in love with Vronsky without having any sort of love-like experience ourselves. And some paradigmatic cases of supposition may involve a phenomenologically distinctive experience, as when we are invited to engage in hypothetical reasoning about being in such-and-such state or having this or that experience.¹⁷ Hence, supposition is a variety of propositional imagination, and one that is typically associated with ratiocinative activities aimed at specific epistemic purposes.

Counterfactual Reasoning

Counterfactual reasoning involves thinking about alternative scenarios and possible states of affairs via the use of counterfactual conditional statements of the form “If *A* were the case, then *C* would be the case,” or “ $A \Box \rightarrow C$ ” in the standard formal notation. Counterfactual reasoning satisfies MCPI and therefore qualifies as a variety of propositional imagination. This ties in with

¹⁷ Another argument against regarding supposition as a kind of imagination is Gendler’s (1994) argument from imaginative resistance. Arguments pulling in the same direction have also been offered by Moran (1994) and Goldman (2006). We agree with Nichols (2006) that these arguments remain inconclusive.

the fact that Williamson recently advanced an account of counterfactual reasoning in terms of propositional imagination. He writes: "When we work out what would have happened if such-and-such had been the case, we frequently cannot do it without imagining such-and-such to be the case and letting things run" (2005, 19). On this view, if King Lear thinks, "If only I had not divided my kingdom between Goneril and Regan, Cordelia would still be alive," he imagines a relevant situation in which he does not divide the kingdom between his two older daughters and from this he further imagines that Cordelia would still be alive. In order to do this, imagination must be constrained in specific ways.

Stalnaker (1968) and Lewis (1973) advanced semantic analyses of counterfactuals that offer implicit criteria for how imagination should be constrained in counterfactual reasoning. The leading idea of both analyses is that a counterfactual $A \Box \rightarrow C$ is true if and only if in the closest possible world where A is true C is also true (we discuss differences between Stalnaker's and Lewis's development of this idea in section 1.6). It is important that the notion of closeness in the phrase "closest possible world" means closeness to the *actual world*, or to *reality*. Let us call a possible world in which A is the case an A -world. The counterfactual conditional $A \Box \rightarrow C$ is then true if and only if C is true in the A -world that is closest to the actual world.

The truth conditions for counterfactuals provide the essential clues for the analysis of counterfactual imagination. The first essential feature is *selectivity* (S). When King Lear imagines what would have happened if he had not divided his kingdom between his two older daughters, he selects an antecedent that is contrary to a relevant fact in a *very specific way*. When thinking counterfactually one does not merely ponder that things could have been different. One selects a particular manner in which things could have been different (specified in A) and then reasons about a world in which this difference is the case (the A -world). The second feature is *reality orientation* (RO). There could be many possible worlds in which A is true, and one could check for the truth of C in any of them. But those conditions don't treat all A -worlds on par. They single out an A -world (or, as we shall see, a class of A -worlds) that is *closest to reality* as the one that determines the truth of the counterfactual conditional. When King Lear pondered what would have happened if he had divided his kingdom differently, he wondered how things would be in a world that is *just like the real world* apart from the distribution of property in his family. Minimal departure from the actual world is an essential constraint on counterfactual reasoning.

We can then fill the blank in in our schema as follows:

Counterfactual reasoning = MCPI & S & RO.

Contemporary work on counterfactual reasoning in empirical psychology backs the idea that when people evaluate counterfactual conditionals their imaginings are constrained in a reality-oriented way. Byrne (2005) presents a series of experiments suggesting that people tend to imagine worlds with the same natural laws, with alternatives to more recent events rather than earlier events, and with alternatives to events that they can control rather than events that they cannot control.¹⁸ This is consonant with the reality orientation that emerges from Stalnaker's and Lewis's analyses. We note, however, that a more fine-grained analysis of RO faces important issues. Stalnaker appeals to the "intuitive idea that the nearest, or least different, world in which [the] antecedent is true is the one that should be selected" (1981, 88) but provides no guidance as to what counts as "least different." Lewis assumes a notion of similarity of worlds that is taken as a primitive, which, as Arló-Costa and Egré notice, "leaves the notion of similarity unconstrained and mysterious" (2016, sec. 6.1).

Dreams

Scientists sometimes refer to their dreams as a source of inspiration for their discoveries, as in Kekulé's introspective report mentioned earlier. Dreams satisfy MCPI to the extent that they are free, they usually mirror standard inferential mechanisms of reasoning, and they are quarantined since their content does not export to real-world contexts. The individuating features of dreams are that they are *solitary* imaginative activities (SIA) that are performed while asleep (SI). These will fill the blank in the scheme:

Dream = MCPI & SIA & SI.

Walton describes dreams as also being "spontaneous, undeliberate imaginings that the imager not only does not but cannot direct (consciously)" (1990, 16), and so one might be tempted to add these features to the list of

¹⁸ Johnson-Laird 1983 and Roese and Olson 1995 offer further empirical evidence that counterfactual reasoning is constrained in a reality-oriented way. See also Weisberg 2016 for a discussion of philosophical and psychological treatments of how much of the real world is imported in counterfactual scenarios.

conditions. However, Ichikawa (2009) points out that those of us who can engage in *lucid dreaming* (which involves the subject's awareness that he or she is dreaming) are able to consciously guide and explore their dreams. Dreams are often thought to involve some variety of imagery, but forming a mental image while dreaming is not necessary: we can dream conversations, jokes, philosophical arguments, and so on.¹⁹

Make-believe

Make-believe emerges as a specific theoretical notion within Walton's (1990) theory of fiction. Walton characterizes make-believe as "the use of (external) props in imaginative activities" (1990, 67). Anything capable of affecting our senses can become a prop in virtue of there being a *prescription to imagine* something—that is, a social convention either explicitly stipulated or implicitly understood as being in force within a certain game. Props are generators of fictional truths. Fictional truth is a property of those propositions that are among the prescriptions to imagine of a certain game. Walton's notion of fictional truth is intrinsically *normative* and *objective* to the extent that the statement "it is fictional that *p*" is to be understood as "it is to be imagined that *p*." Walton thinks that works of fiction are props in games of make-believe. When reading the Sherlock Holmes stories we imagine that Holmes lives at 221B Baker Street in virtue of Conan Doyle's prescription to imagine that this is the case. We can imagine that Holmes lives in Paris, but this does not conform to the story.

Fictional truths divide into primary truths and implied truths, where the former are generated directly from the text while the latter are generated indirectly from the primary truths via general principles and standard rules of inference. These are called principles of generation. Sometimes implicit fictional truths are generated according to the so-called *reality principle*, which keeps the world of the fiction as close as possible to the real world. For example, from the primary fictional truth that Sherlock Holmes lives in Baker Street and our knowledge of London's geography we can infer the implied fictional truth that Holmes lives nearer to Paddington Station than to Waterloo Station. Depending on the context of interpretation, however, implied truths can also be generated according to the *mutual belief principle*, which is directed toward the mutual beliefs of the members of the community in which the story originated. Many of the implied truths of Dante's *Divine Comedy*

¹⁹ Closely related to dreaming is daydreaming. For a discussion, see Walton 1990, 13.

are generated from the primary truths of the story and the medieval belief in the main tenets of the Ptolemaic geocentric system.

Two main features of make-believe emerge from Walton's characterization: make-believe is a social activity (SA) and it involves props that convey a normative aspect (NA) to its content. It obviously satisfies the MCPI conditions, and so we obtain:

$$\text{Make-believe} = \text{MCPI} \ \& \ \text{SA} \ \& \ \text{NA}.$$

Some might question the characterization of make-believe as a variety of propositional imagination. Walton himself distinguishes between "imagining a *proposition*, imagining a *thing*, and imagining *doing* something—between, for instance, imagining that there is a bear, imagining a bear, and imagining seeing a bear" (1990, 13). In particular, he develops the latter notion as *imagining de se*, as opposed to mere propositional imagination, and further claims that games of make-believe involve a sort of participation that crucially requires *de se* imagining. The motivation for Walton's claim is that on his view literary fictions have a specific cognitive purpose in granting us insight into ourselves, which requires imagining things from a participatory perspective.²⁰

However, Currie (1990) argues, rightly in our view, that make-believe, just like belief and desire, is a propositional attitude. He does not think of make-believe as a phenomenologically distinctive attitude, although he does accept that make-believe, like belief and desire, "is a kind of state that can be accompanied by or give rise to introspectible feelings and images" (1990, 21). This, however, is not necessary and hence not a defining feature of make-believe.

According to this characterization of make-believe, episodes of supposition and counterfactual reasoning are also episodes of make-believe if they involve props and are therefore constrained by the prescriptions to imagine in a game of make-believe. In this way, they also satisfy NA and SA. Dreams, by contrast, cannot be interpreted in a similar way. Dreaming is a solitary activity that does not satisfy SA and NA because it does not involve props.

²⁰ Cf. Currie 1990, sec. 1.4, 7.5.

1.5 Reconsidering the Scientific Imagination

We now return to the views we introduced in section 1.3. As we have seen, Norton puts forward ET, suggesting that the picturesque character of a TE can be eliminated. However, at the same time condition (i) claims that TEs posit hypothetical or counterfactual states of affairs. As we have seen, counterfactual reasoning constitutes a variety of propositional imagination, which would suggest that conducting a TE involves propositional imagination. This suspicion firms up when we look at Norton's reconstructions of TEs. Consider Galileo's falling bodies, which Norton (1996, 341–342) reconstructs as a *reductio ad absurdum*:

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

This argument satisfies ReT and the weak interpretation of ET since (9) is a general claim about all falling stones. However, it does not conform to the strong interpretation of ET because it does posit imagined states of affairs involving imagined particulars. Steps (2), (4), (5), and (6) explicitly involve reference to the objects described in Galileo's original TE. None of the situations specified by these statements actually obtains in the real world. We assume them in the imagination for the purpose of drawing the relevant inferences. This does not mean that the general laws and principles reached via TEs could not be reached via some other means. But in TEs the arguments leading to the general conclusions involve imagined scenarios and particulars.

We have pointed out that the propositional imagination is characterized by MCPI, positing an ability to ponder and evaluate alternative scenarios that is deliberate, mirrors the inferential mechanisms of belief, and quarantines content. This is exactly the sort of imagination required by TEs. Galileo deliberately imagines a certain hypothetical scenario, he develops a deductive reasoning leading to a contradiction, and he quarantines its content since he explicitly invites us to imagine a non-actual situation. We conclude that TEs involve propositional imagination.²¹ The remaining question is, which kind of propositional imagination? We come back to this issue in section 1.6.

Let us now consider the view that the imagistic variety of objectual imagination is crucial to the performance of TEs. We focus on Nersessian's proposal because she offers the most detailed defense of this view. As we have seen, her account is based on the notions of mental analogues and iconic representations. She develops these concepts by appealing to Barsalou's distinction between modal and amodal symbols, which we discussed in section 1.4.1. Mental models are iconic representations that can be composed of either modal or amodal symbols. So, for example, a cat-like representation on a plane-like representation is a mental model constituted by modal symbols (modal iconic). A circle resting on a square for a cat being on a plane is a mental model constituted by amodal symbols (amodal iconic).²²

Iconic representations (be they modal iconic or amodal iconic) are *imagistic* according to the currently dominant notion of imagery, which, as we have seen, rejects the identification of mental images with pictures in the mind.²³ Figure 1.2b is not a picture. The circle and the square are arbitrarily linked to what they represent, yet they preserve the spatial relations that Figure 1.2a has. Figure 1.2b is more abstract than Figure 1.2a, but it is an image nevertheless.

The main problem with Nersessian's proposal, as well as with other accounts produced within the literature on mental models, is that there is no general consensus on many foundational issues of this framework, a point that Nersessian (2007, 129ff.) herself acknowledges. In particular, the appeal to similarity and goodness of fit as the kind of relationship that characterizes iconic representations is controversial. As we have pointed out, most

²¹ This admission is also implicit in Sorensen's (1992, 202–203) discussion of supposition.

²² Thanks to Nancy Nersessian for suggesting these two examples to us in personal communication.

²³ In fact, Nersessian rejects the old pictorial notion of imagery. See Nersessian 1992, 294; 2007, 133 and 149 n. 6. She declares, however, that iconic mental models are imagistic in the contemporary sense of the term (cf. 2007, 137).



Figure 1.2a Modal iconic

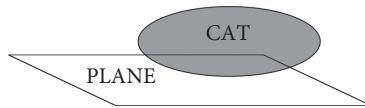


Figure 1.2b Amodal iconic

cognitive scientists nowadays recognize that mental images have a specific representational format. Yet the standard view is that the relationship between a mental image and the object it represents is determined by a description couched in mentalese. Mental images might share some properties with what they represent, but this is not what makes them representations of what they represent. As long as these basic issues remain unresolved, Nersessian's claim that TEs are iconic representations and that the execution of a TE consists merely in the manipulation of such representations remains in need of clarification.

However, even if we assume, for the sake of argument, that these issues can be resolved in a satisfactory manner, two concerns about the imagistic view remain. The first is whether imagistic reasoning is *sufficient* to the derivation of the outcome of a TE. The problem is that not all factors that matter to the successful performance of a TE seem to have sensory-like correlates. When considering Galileo's cavity we do not seem to have a perception-like representation of the cavity being frictionless or of the lack of air resistance. Likewise, we cannot form a perception-like representation of the concept of force without having a theoretical definition, which is usually given in linguistic and formulaic symbols. Similarly, Malileo's SM assumes these concepts, but he also requires theoretical knowledge of Lagrangean mechanics,

general principles and laws, mathematical abilities, and logical inferential abilities. We could not even begin to reason about the model and its domain of inquiry without the relevant theoretical, mathematical, and logical abilities. So it is not surprising that Nersessian admits that “information deriving from various representational formats, including language and mathematics, plays a role in scientific thought experimenting” (2004, 147). However, this form of reasoning is, by her own lights, fundamentally different from the reasoning with iconic representations, and so it is difficult to see how it fits into a view that places iconic representations at the heart of TEs. Imagistic reasoning therefore seems insufficient for the performance of TEs and use of SMs.

The second concern is whether imagistic reasoning is *essential* (or necessary) to the performance of TEs. Our abilities to form mental images and perform the relevant kinds of operations are highly subjective and idiosyncratic. Yet it would be implausible to argue that individuals with a poor imagistic ability could not derive the correct outcome of Galileo’s TE (or, for that matter, of any TE).²⁴ Presumably, one could perform the TE and draw the relevant conclusion by understanding the propositional content of the argument underlying it. When performing the TE we do not have to form a mental image of the U-shaped cavity and the series of transformations we described in section 1.2. We need to grasp the relevant concepts, with or without forming a mental image of the objects and transformations they stand in for. The problem becomes even more perspicuous when we consider SMs. Malileo’s SM could be illustrated with figures that facilitate a scientist’s reasoning by making it more vivid, and some of us might form a mental image of the parabola and the ball. However, this is not necessary. We can calculate the trajectory of the ball by going through the relevant mathematical calculations and by deploying the mathematical and theoretical notions that are relevant for this specific domain of inquiry.

1.6 Analyzing the Scientific Imagination

We have argued that while TEs and SMs do not require imagery, the propositional imagination is crucial to them. But what sort of propositional

²⁴ As Arnon Levy pointed out to us, this would be an interesting empirical question.

imagination is required? In section 1.4.2 we individuated supposition, counterfactual reasoning, dreaming, and make-believe as different varieties of propositional imagination. Scientists sometimes report their dreams as a source of inspiration for scientific discoveries. But these imaginative activities are typically subjective and unconstrained, and, more to the point, they are not involved in the performance of a TE or the exploration of a SM. So we can safely set dreams aside.

This leaves the other three varieties as contenders. They are genuine options and deserve to be taken seriously. We now discuss what it would take to analyze TEs and SMs in terms of each of these options and make the challenges that emerge explicit. Our tentative conclusion is that SMs and TEs are most naturally explained in terms of make-believe. The conclusion is tentative because we don't claim to present a complete account of the scientific imagination, and a final analysis may well end up incorporating elements from all three accounts.

Let us begin with supposition. Often scientists introduce TEs and SMs by explicitly inviting us to suppose that some (real or non-actual) objects are endowed with certain properties and that they behave in certain ways. To perform a TE or use an SM would then amount to supposing a number of things and deriving consequences from them with the aim of gaining knowledge. Unfortunately, this is too weak. Supposition, as we have characterized it, is not an essentially social activity (since it can be purely private), and as such, it does not account for the social character of scientific activities. Furthermore, it does not have a normative element to it, and such elements seem to be characteristic of scientific thought. One can suppose anything, and as long as no further restrictions are imposed, one can conclude almost anything from certain assumptions. The notion of supposition imposes no constraints on inferences beyond those that follow from mirroring, which is part of MCPI. This is too little. First, mirroring alone is too weak to capture the way in which the imagination is constrained in TEs and SMs. Second, mirroring only provides a thin inferential structure that consists primarily of logical operations, but it doesn't offer the kind of principles that would guide a process of investigation to the kind of inferred truths that the study of TEs and SMs aims to uncover. For these reasons supposition does not offer a satisfactory analysis of the propositional imagination in TEs and SMs.

Let us now consider counterfactual reasoning. From this point of view the performance of a TE or the use of an SM amounts to evaluating the counterfactual $M \Box \rightarrow C$, where " M " is a description of the SM or TE. A claim C is

then true in the TE or SM if the counterfactual $M \Box \rightarrow C$ is true. For instance, it is true in Newton's model of the solar system that planets move in elliptical orbits if the counterfactual "if planets were perfect spheres gravitationally interacting with each other and nothing else, then they would move in elliptical orbits" is true.

A first challenge for this analysis of TEs and SMs is the issue of completeness. Possible worlds are complete. Intuitively, a possible world is complete when the principle of the excluded middle holds and for any proposition p it is the case that either p or *not- p* holds.²⁵ But models are not complete in this sense. Claims about the date of the Battle of Waterloo, the height of the tallest building in London, and the average rainfall in China last year are neither true nor false in, say, Einstein's elevator TE or a mechanical model of the atom simply because battles, buildings, and levels of rainfall are not part of these TEs and SMs. However, the closest possible world in which M is true is one in which there are matters of fact about these things (because possible worlds are complete), and so the counterfactual $M \Box \rightarrow C$ may have a truth-value for claims that have nothing to do with the model. For instance, the counterfactual "if planets were perfect spheres gravitationally interacting with each other and nothing else, then the height of the tallest building in London would be 310 meters" could come out true. But in fact the truth-value of this counterfactual should be indeterminate (i.e., $M \Box \rightarrow C$ should be neither true nor false). So the worry is that the standard semantics for counterfactuals would make TEs and SMs complete.

Whether this worry is a real problem depends on the details of the account. The crucial question is whether the account one adopts accepts the so-called principle of conditional excluded middle (CEM).²⁶ CEM says that for all C either $M \Box \rightarrow C$ is true or $M \Box \rightarrow \sim C$ is true (where " $\sim C$ " stands for "*not- C* "). Stalnaker's semantics works with a selection function that picks a *unique* nearest world w , and hence the truth-value of $M \Box \rightarrow C$ is simply the truth-value of C in w . Since C is either true or false in w , either $M \Box \rightarrow C$ or $M \Box \rightarrow \sim C$ is true and CEM holds. Stalnaker (1981) has defended CEM, and a

²⁵ See Van Inwagen 1986 for a critical discussion of the notion of completeness and the metaphysics of possible worlds, and Priest 2008 for a discussion of the notion of completeness in modal logic. Stalnaker (1986, esp. 117–118) further discusses the notion of completeness and its role in framing the distinction between possible world semantics and situation semantics (e.g., Barwise and Perry 1983, 1985), where completeness applies only to possible worlds as total states that include everything that is the case, while situations can be construed as partial worlds or small parts of worlds and therefore cannot be complete.

²⁶ We are grateful to Timothy Williamson, Matthieu Gallais, and Sonia Roca-Royes for helpful discussions about CEM.

number of recent authors have followed suit (see, e.g., Cross 2009; Williams 2010). But CEM conflicts with the incompleteness of TEs and SMs, and defenders of CEM have to find a way around this problem.

In contrast with Stalnaker's, Lewis's semantics works with a comparative similarity relation, which defines a weak total ordering of all possible worlds with respect to each possible world. When several possible worlds tie in for similarity, the truth of $M \Box \rightarrow C$ requires the truth of C in *all* the nearest M -worlds. If C is true in some of the nearest M -worlds but not in others, then both $M \Box \rightarrow C$ and $M \Box \rightarrow \sim C$ are false in the actual world and CEM fails. The failure of CEM is a step in the right direction, but by itself this is insufficient to solve the problem of incompleteness. For a solution of this problem not only requires that for *some* C neither $M \Box \rightarrow C$ nor $M \Box \rightarrow \sim C$ is true, but it requires that this be the case for *all* C s that don't belong to the TE or SM. This implies that for all C s about which the TE or the SM remain silent, it must be the case that there are some M -worlds in which C is true and some other M -worlds in which C is false that are at the same distance from the actual world. Since the set of C s that belongs to the TE or SM is different from case to case, this approach requires that we give up on the notion of a universal similarity metric between possible worlds and postulate that each TE or SM comes with a tailor-made cross-world similarity metric that ensures that $M \Box \rightarrow C$ has no determinate truth value for all the right C s.

The next issue is how we acquire counterfactual knowledge. Roca-Royes submits that "how capable we are of counterfactual knowledge depends on how capable we are of tracking the similarity order" (2012, 154). In agreement with Kment (2006), she also holds that our capability for counterfactual knowledge "needs to be based on rules that permit us to determine which propositions are cotenable with a given antecedent" (Kment 2006, 288). Thus, any epistemology of counterfactuals needs to identify the relevant rules. This, however, is no easy feat. A rule that merely states that we shouldn't go beyond considering possible worlds that are maximally similar to the actual world needs an indication of what counts as a maximally similar world. Kment (2006) offers a metaphysical account of different types of similarity facts and of their relative weights. However, there is no general agreement on these issues. These problems are inherited by a counterfactual epistemology for TEs and SMs. As previously noted, the set of C s that belongs to a TE and a SM is different from case to case. Thus, we need a tailor-made cross-world similarity metric for each case, or perhaps we can identify a series of overarching types of metrics for different types of TEs and SMs.

A tenable account of counterfactual imagination will have to address these issues.

Let us finally turn to make-believe. Analyses of SMs in terms of make-believe have been suggested by Frigg (2010), Levy (2015), and Toon (2012), and of TEs by Meynell (2014). On this view, to perform a TE or use an SM amounts to exploring a fictional scenario that is defined by the primary truths and the principles of generation. In doing so, the scientist discovers things about the scenario and finds out what holds and what doesn't hold in it.

Make-believe is a highly constrained form of imagination. The constraints come from the use of props and the principles of generation that are constitutive of a game of make-believe. These constraints capture well how TEs and SMs work. When performing Galileo's TE we imagine that so-and-so is the case in virtue of Galileo's prescriptions. We could imagine that instead of a ball we put a toothpick on the edge of the cavity. But this is a violation of the prescriptions to imagine in force within Galileo's TE. Furthermore, we derive the law of inertia from the law of equal heights (a general principle of generation) and the appropriate variations of the TE setting as further prescribed by Galileo. Likewise, when working with Malileo's model we could imagine that the ball is oval and has an inhomogeneous mass distribution that causes it to wobble inside the cavity. But this is a violation of the rules of Malileo's game of make-believe. To use the model properly, we have to engage in the official game and derive the outcome from Malileo's prescriptions in combination with the mathematical equation and theoretical principles of Lagrangean mechanics.

Not only is make-believe constrained due to its reliance on props and socially sanctioned principles of generation, but it is also an essentially social imaginative activity. It has an objective content that is normatively characterized in terms of social conventions implicitly or explicitly understood as being in force within the relevant game. The social character and objectivity of make-believe are typical for the sort of imaginative activities involved in TEs and SMs.

The props in the game are the linguistic descriptions, graphs, and mathematical formulae used by scientists in the performance and communication of TEs and in the development and exploration of SMs as props. In this way, we can explain the notion of truth in a TE and truth in a SM in terms of fictional truth. The latter carries over to TEs and SMs simply by interpreting the propositions that are true in a TE and true in a SM as being among the prescriptions to imagine specified in their original assumptions, either explicitly or implicitly. In contrast with possible worlds, the content generated by a game of make-believe is incomplete. Propositions that do not belong to

the game of make-believe of a certain TE or SM are neither mandated to be imagined nor mandated not to be imagined, and hence they are neither fictionally true nor false.

Make-believe also accounts for the mechanisms of generation of the implicit truths of TEs and SMs. The performance of a TE and the exploration of an SM consist in finding out what is true according to a TE and what is true according to an SM, which goes beyond what is explicitly stated in the original assumptions. These implicit fictional truths can be inferred according to certain principles of generation. This also provides an epistemology for fictional truths: we investigate a TE or an SM by finding out what follows from the primary truths of the model and the principles of generation. This is in line with scientific practice, where a significant part of the work goes into studying the consequences of the basic assumptions of the TE or SM. Eventually this leads to the generation of hypotheses about the real world that can be tested for genuine truth or falsity.²⁷

What principles of generation constrain the contents of TEs and SMs? We have presented the reality principle and the mutual belief principle as those constraining the generation of implicit fictional truths in stories. While these principles can be at work in certain TEs or SMs, other options may be possible. Meynell (2014, 4162–4163) points out that different kinds of TEs make use of different principles, and which ones are chosen depends on disciplinary conventions and interpretative practices. Specifically, she points out that “which principles of generation a physicist brings most automatically to a TE will tend to reflect her beliefs about reality as well as the various theories and projects upon which she currently works” (2014, 4163). For this reason neither the reality principle nor the mutual belief principle is in any way privileged, and different principles may be needed in specific domains of scientific inquiry. It is an advantage of the framework of make-believe that it has the flexibility to accommodate such context-specific principles.

Make-believe is at once constrained (due to its reliance on props and principles of generation) and flexible (due the freedom of choosing different principles). This renders make-believe a promising analysis of the kind of imaginative activity at work in TEs and SMs.

²⁷ See Salis 2016 for a discussion of theoretical hypotheses generated in SMs in connection with make-believe and for different analyses.

1.7 Conclusion

This chapter investigated the nature of imaginative activities involved in TEs and SMs. We find ourselves in the seemingly paradoxical situation that the imagination is at once deemed crucial and dismissed because of its purportedly intrinsic imagistic character. This tension can be resolved, we submit, by recognizing that there is a propositional variety of imagination. A discussion of both imagistic and propositional kinds of imagination leads us to the conclusion that while propositional imagination is crucial to the performance of TEs and the use of SMs, imagistic imagination is neither sufficient nor necessary. We then tentatively suggest that the imaginative activities in SMs and TEs are most naturally analyzed in terms of make-believe, leaving open the possibility that a final analysis may well end up incorporating elements from other varieties of propositional imagination.

Acknowledgments

We would like to thank Nancy Nersessian for an extremely helpful email exchange, and Peter Godfrey-Smith, Arnon Levy, and Mike Stuart for comments on earlier drafts. Thanks to Alisa Bokulich, Ruth Byrne, Greg Currie, Stacie Friend, Matthieu Gallais, Manuel García-Carpintero, Sonia Roca-Royes, Alberto Voltolini, Michael Weisberg, and Tim Williamson for helpful discussions. Previous versions of this chapter were presented at workshops at the Van Leer Institute in Jerusalem, at the Institute of Philosophy in London, and at the Centre for Philosophy of Natural and Social Science at the London School of Economics. We would like to thank the audiences of these events for their comments. Frigg would like to acknowledge financial support from the Spanish Ministry of Science and Innovation (MICINN) through grant FFI2012-37354. Salis would like to acknowledge financial support from the European Union's Horizon 2020 Research and Innovation Programme under the Marie Skłodowska-Curie grant agreement no. 654034.

References

- Aristotle. (1995). "De Anima." In *The Complete Works of Aristotle: The Revised Oxford Translation*, Vol. 1. Bollingen Series LXXI. Princeton, NJ: Princeton University Press.

- Arlo-Costa, H., and Egré, P. (2016). "The Logic of Conditionals." In *The Stanford Encyclopedia of Philosophy* (Winter 2016 ed.), edited by E. N. Zalta. <http://plato.stanford.edu/archives/win2016/entries/logic-conditionals>.
- Barsalou, L. W. (1999). "Perceptual Symbol Systems." *Behavioral and Brain Sciences* 22: 577–660.
- Barwise, J., and Perry, J. (1983). *Situations and Attitudes*. Cambridge, MA: MIT Press.
- Barwise, J., and Perry, J. (1985). "Shifting Situations and Shaken Attitudes." *Linguistics and Philosophy* 8: 103–161.
- Bohr, N. ([1934] 1961). *Atomic Theory and the Description of Nature: Four Essays with an Introductory Survey*. Cambridge: Cambridge University Press.
- Brown, J. R. (1991). *The Laboratory of the Mind*. Cambridge: Cambridge University Press.
- Brown, J. R. (2004). "Peeking into Plato's Heaven." *Philosophy of Science* 71, no. 5: 1126–1138.
- Byrne, R. (2005). *Rational Imagination. How People Create Alternatives to Reality*. Cambridge, MA: MIT Press.
- Cartwright, N. (2010). "Models: Parables v. Fables." In *Beyond Mimesis and Convention: Representation in Art and Science*, edited by R. Frigg and M. C. Hunter, 19–32. Berlin: Springer.
- Cross, C. (2009). "Conditional Excluded Middle." *Erkenntnis* 70: 173–188.
- Currie, G. (1990). *The Nature of Fiction*. Cambridge: Cambridge University Press.
- Currie, G., and Ravenscroft, I. (2002). *Recreative Minds: Imagination in Philosophy and Psychology*. Oxford: Clarendon Press.
- Del Re, G. (2000). "Models and Analogies in Science." *Hyle* 6: 5–15.
- Dirac, P. A. M. (1958). *Principles of Quantum Mechanics*. 4th ed. Oxford: Clarendon Press.
- Einstein, A. (2005). *Relativity: The Special and the General Theory*. New York: Pi Press.
- Fodor, J. A. (1975). *The Language of Thought*. New York: Thomas Crowell.
- Frege, G. ([1884] 1953). *The Foundations of Arithmetic*. Translated by J. L. Austin. Oxford: Blackwell.
- Frigg, R. (2010). "Fiction and Scientific Representation." In *Beyond Mimesis and Nominalism: Representation in Art and Science*, edited by R. Frigg and M. Hunter, 97–138. Berlin: Springer.
- Gaut, B. (2003). "Imagination and Creativity." In *The Creation of Art: New Essays in Philosophical Aesthetics*, edited by B. Gaut and P. Livingston, 148–173. Cambridge: Cambridge University Press.
- Gaut, B. (2010). "The Philosophy of Creativity." *Philosophy Compass* 5, no. 12: 1034–1046.
- Gaut, B., and Livingston, P. (Eds.). (2003). *The Creation of Art: New Essays in Philosophical Aesthetics*. Cambridge: Cambridge University Press.
- Gendler, T. S. (2004). "Thought Experiments Rethought—and Reperceived." *Philosophy of Science* 71: 1154–1163.
- Godfrey-Smith, P. (2006). "The Strategy of Model-Based Science." *Biology and Philosophy* 21: 725–740.
- Godfrey-Smith, P. (2009). "Models and Fictions in Science." *Philosophical Studies* 143: 101–116.
- Goldman, A. (2006). "Imagination and Simulation in Audience Responses to Fiction." In *The Architecture of the Imagination*, edited by S. Nichols, 41–56. Oxford: Clarendon Press.
- Harré, R. (1970). *The Principles of Scientific Thinking*. London: Macmillan.
- Harré, R. (1988). "Where Models and Analogies Really Count." *International Studies in the Philosophy of Science* 2: 118–133.

- Ichikawa, J. (2009). "Dreaming and Imagination." *Mind and Language* 24, no. 1: 103–121.
- Johnson-Laird, P. N. (1980). "Mental Models in Cognitive Science." *Cognitive Science* 4: 71–115.
- Johnson-Laird, P. N. (1982). "The Mental Representation of the Meaning of Words." *Cognition* 25: 189–211.
- Johnson-Laird, P. N. (1983). *Mental Models*. Cambridge, MA: MIT Press.
- Johnson-Laird, P. N. (1989). "Mental Models." In *Foundations of Cognitive Science*, edited by M. Posner, 469–500. Cambridge, MA: MIT Press.
- Kment, B. (2006). "Counterfactuals and the Analysis of Necessity." *Philosophical Perspectives* 20: 237–302.
- Kosslyn, S. M. (1980). *Image and Mind*. Cambridge, MA: Harvard University Press.
- Kosslyn, S. M. (1983). *Ghosts in the Mind's Machine: Creating and Using Images in the Brain*. New York: Norton.
- Kosslyn, S. M. (1994). *Image and Brain: The Resolution of the Imagery Debate*. Cambridge, MA: MIT Press.
- Kosslyn, S. M. (2005). "Mental Images and the Brain." *Cognitive Neuropsychology* 22: 333–347.
- Laymon, R. (1991). "Thought Experiments by Stevin, Mach and Gouy: Thought Experiments as Ideal Limits and as Semantic Domains." In *Thought Experiments in Science and Philosophy*, edited by T. Horowitz and G. J. Massey, 167–191. Savage, MD: Rowman and Littlefield.
- Leslie, A. (1987). "Pretense and Representation: The Origins of 'Theory of Mind.'" *Psychological Review* 94, no. 4: 412–426.
- Levy, A. (2015). "Modeling Without Models." *Philosophical Studies* 172, no. 3: 781–798.
- Lewis, D. K. (1973). *Counterfactuals*. Oxford: Basil Blackwell.
- Maxwell, J. C. (1965). *The Scientific Papers of James Clerk Maxwell*. Edited by W. D. Niven. Mineola, NY: Dover Publications.
- Meynell, L. (2014). "Imagination and Insight: A New Account of the Content of Thought Experiments." *Synthese* 191: 4149–4168.
- Moran, R. (1994). "The Expression of Feeling in the Imagination." *Philosophical Review* 103: 75–106.
- Morgan, M. (2004). "Imagination and Imaging in Model Building." *Philosophy of Science* 71: 753–766.
- Nersessian, N. J. (1992). "In the Theoretician's Laboratory: Thought Experimenting as Mental Modeling." *Philosophy of Science* 2: 291–301.
- Nersessian, N. J. (1999). "Model-Based Reasoning in Conceptual Change." In *Model-Based Reasoning in Scientific Discovery*, edited by L. Magnani, N. J. Nersessian, and P. Thagard, 5–22. New York: Kluwer Academic/Plenum.
- Nersessian, N. J. (2007). "Thought Experimenting as Mental Modeling: Empiricism Without Logic." *Croatian Journal of Philosophy* 7, no. 20: 125–154.
- Nichols, S. (2004). "Imagining and Believing: The Promise of a Single Code." *Journal of Aesthetics and Art Criticism* 62: 129–139.
- Nichols, S. (2006). "Just the Imagination: Why Imagining Doesn't Behave Like Believing." *Mind and Language* 21, no. 4: 459–474.
- Nichols, S., and Stich, S. (2000). "A Cognitive Theory of Pretense." *Cognition* 74: 115–147.
- Nichols, S., and Stich, S. (2003). *Mindreading: An Integrated Account of Pretence, Self-Awareness, and Understanding Other Minds*. Oxford: Clarendon Press.

- Nigel, T. J. T. (2014). "Mental Imagery." In *The Stanford Encyclopedia of Philosophy* (Fall 2014 ed.), edited by E. N. Zalta. <http://plato.stanford.edu/archives/fall2014/entries/mental-imagery>.
- Norton, J. (1991). "Thought Experiments in Einstein's Work." In *Thought Experiments in Science and Philosophy*, edited by T. Horowitz and G. J. Massey, 129–148. Savage, MD: Rowman and Littlefield.
- Norton, J. (1996). "Are Thought Experiments Just What You Thought?" *Canadian Journal of Philosophy* 26, no. 3: 333–366.
- Norton, J. (2004). "On Thought Experiments: Is There More to the Argument?" *Philosophy of Science* 71: 1139–1151.
- Odenbaugh, J. (2015). "Semblance or Similarity? Reflections on Simulation and Similarity." *Biology and Philosophy* 30: 277–291.
- Paivio, A. (1986). *Mental Representations: A Dual Coding Approach*. New York: Oxford University Press.
- Peacocke, C. (1985). "Imagination, Experience, and Possibility." In *Essays on Berkeley*, edited by J. Foster and H. Robinson, 19–35. Oxford: Oxford University Press.
- Perner, J. (1991). *Understanding the Representational Mind*. Cambridge, MA: MIT Press.
- Priest, G. (2008). "Many-Valued Modal Logics: A Simple Approach." *Review of Symbolic Logic* 1, no. 2: 190–203.
- Roca-Royes, S. (2012). "Essentialist Blindness Would Not Preclude Counterfactual Knowledge." *Philosophia Scientiae* 16, no. 2: 149–172.
- Roese, N. J., and Olson, J. M. (1995). *What Might Have Been: The Social Psychology of Counterfactual Thinking*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Ryle, G. (1949). *The Concept of Mind*. London: Hutchinson.
- Salis, F. (2016). "The Nature of Model-World Comparisons." *Monist* 99, no. 3: 243–259.
- Shepard, R. N. (1978). "The Mental Image." *American Psychologist* 33, no. 2: 125–137.
- Shepard, R. N., and Metzler, J. (1971). "Mental Rotation of Three-Dimensional Objects." *Science* 171: 701–703.
- Sorensen, R. (1992). *Thought Experiments*. New York: Oxford University Press.
- Stalnaker, R. (1968). "A Theory of Conditionals." In *Studies in Logical Theory*, edited by N. Rescher, 2:98–112. American Philosophical Quarterly Monograph Series, vol. 2. Oxford: Blackwell.
- Stalnaker, R. (1981). "A Defense of Conditional Excluded Middle." In *Ifs: Conditionals, Belief, Decision, Chance, and Time*, edited by W. L. Harper, R. Stalnaker, and G. Pearce, 87–102. University of Western Ontario Series in Philosophy of Science vol. 15. Dordrecht: D. Reidel.
- Stalnaker, R. (1986). "Possible Worlds and Situations." *Journal of Philosophical Logic* 15, no. 1: 109–123.
- Sugden, R. (2009). "Credible Worlds, Capacities and Mechanisms." *Erkenntnis* 70: 3–27.
- Toon, A. (2012). *Models as Make-Believe: Imagination, Fiction, and Scientific Representation*. New York: Palgrave Macmillan.
- Van Inwagen, P. (1986). "Two Concepts of Possible Worlds." *Midwest Studies in Philosophy* 11: 185–213.
- Velleman, J. D. (2000). "The Aim of Belief." In *The Possibility of Practical Reason*, 244–281. Oxford: Oxford University Press.
- Walton, K. L. (1990). *Mimesis as Make-Believe: On the Foundations of the Representational Arts*. Cambridge, MA: Harvard University Press.

- Walton K. L. (1994). "Morals in Fiction and Fictional Morality." *Proceedings of the Aristotelian Society* 68: 27–50.
- Weisberg, D. S. (2016). "How Fictional Worlds Are Created." *Philosophy Compass* 11, no. 8: 462–470.
- Weisberg, M. (2007). "Who Is a Modeler?" *British Journal for the Philosophy of Science* 58: 207–233.
- Weisberg, M. (2013). *Simulation and Similarity: Using Models to Understand the World*. New York: Oxford University Press.
- Williams, R. (2010). "Defending Conditional Excluded Middle." *Notus* 44, no. 4: 650–668.
- Williamson, T. (2005). "Armchair Philosophy, Metaphysical Modality, and Counterfactual Thinking." *Proceedings of the Aristotelian Society* 105: 1–23.
- Wittgenstein, L. (1953). *Philosophical Investigations*. Edited by G. E. M. Anscombe and R. Rhees. Translated by G. E. M. Anscombe. Oxford: Blackwell.
- Yablo, S. (1993). "Is Conceivability a Guide to Possibility?" *Philosophy and Phenomenological Research* 53, no. 1: 1–42.