Literature as Thought Experiment ed. Faulk Bornmüller, Johannes Frazen, Mathis Lessau. Paderborn: Wilhelm Fink Verlag (2019), pp. 1-16.

Imaginative Investigations: Thought Experiments in Science, Philosophy and Literature

It is not unusual to emerge from an encounter with a novel, play, or film thinking that you have learned something. What you have in mind is not just that you've learned something about the work itself – that Shakespeare had a large vocabulary, or Hitchcock was a master of suspense – but something about the extra-literary world. Such a contention seems right, but it is epistemologically problematic. Inasmuch as the work is a fiction, it is not and does not purport to be literally true. Nor does it provide any justification, at least not any that extends beyond the fictive realm. We can readily understand how and why a work of fiction causes the reader to change her mind about things – that's a purely psychological matter. But to have *learned* is to have somehow improved one's take on things – to have gained knowledge, understanding, insight, even wisdom. And it is hard to see how imaginatively entertaining an elaborate falsehood could engender that.

Perhaps the conviction that literature is non-accidentally a source of epistemic improvement is wrong. Maybe literature only has the power to provoke us to change our minds. Inasmuch as the protagonists in a work of fiction typically do not exist, and the events described typically do not occur, perhaps we should conclude that literature's epistemic pretensions are unwarranted. This conclusion is hard to square with the profound effect literature often has on us. But there is a clear tension between the epistemic claims of literature and the requirements that epistemology standardly sets. If conveying understanding about a topic requires representing it literally and accurately, then literature does not convey understanding. Neither, however, does science. For science uses laboratory experiments and thought experiments that distance themselves from the facts in order to illuminate them. Elsewhere, I have called effective fictions and thought experiments *felicitous falsehoods* (Elgin 2017). Arguably successful laboratory experiments fall under this heading as well.

Classing laboratory experiments with fictions may appear unwarranted. Laboratory experiments seem to engage with reality in a way that fictions do not. Bench scientists manipulate real materials to produce their results – they experiment on real mice, real amino acids, real electrons, or whatever. But many of the objects they manipulate are not to be found in nature. Biologists run tests on genetically identical mice, not ordinary field mice. The mice used in experiments are artifacts whose genome has

been intentionally modified to produce a particular profile. The scientist chooses her mice from a catalog, selecting the genome that best serves her needs. The experimental setting isolates the mice and protects them from the normal slings and arrows of rodent misfortune. No hungry cats prowl the lab; no mouse traps are set. But the mice in the lab are not pampered. They are subject to insults that do not occur in nature. They are exposed to high levels of radiation, to massive doses of chemicals, to very hot or very cold temperatures, etc. Experiments are apt to go to extremes. The chemicals used in experiments are pure – the water is H_2O, not rain water or tap water; the salt is NaCl, not sea salt or table salt. The purity of the chemicals results from their having been purified – that is, intentionally altered from their natural states. The result of the experiment thus directly reveals what happens to a special, manufactured population of experimental items in highly artificial circumstances. It is then up to the scientist to extrapolate from those findings to something that happens in the wider world. The laboratory experiment does not replicate the phenomena it illuminates. It isolates, purifies, amplifies. It eliminates or controls for confounding factors in order to bring out something we would not ordinarily see.

Thought experiments distance themselves even further. They are not real experiments; many are not even possible experiments. They are imaginative exercises designed to discover what would happen if certain (perhaps unrealizable) circumstances obtained. In performing such exercises, thinkers suspend some epistemic commitments and hold fast to others. The second law of thermodynamics states that heat cannot be transferred from the hotter to the colder unless work is performed on the system. That is, entropy always increases. It follows from the second law that an ensemble of gas with some hot molecules and some cold ones will eventually evolve to a state of thermal equilibrium with a temperature between the two. To understand the law, Maxwell imagines the following scenario: An ensemble of gas molecules is isolated in a container with a barrier splitting the container in half and a small door connecting the two chambers, *A* and *B*. Some molecules in the ensemble are hot and therefore travel quickly. Others are cool and travel slowly. (They do not have to be very different in temperature; all that matters is that there is some difference.) A tiny, dexterous demon monitors the door, opening and shutting it so that only fast molecules can move to chamber *A* and only slow ones to chamber *B*. Eventually, as a result of his actions, the gas in chamber *A* will be hotter than the gas in chamber *B*, thereby violating the second law. Since the system is isolated, and the demon did nothing to the molecules themselves, no work (in the relevant sense of ‘work’) was done. So the imagined situation does not conform with the law. Still, it is conceivable. And there is nothing

thermodynamically objectionable about the set-up. Even though there is and could be no such demon, Maxwell's thought experiment shows that the second law of thermodynamics is a statistical law (Maxwell 1892). It could in principle be violated. It thus discredits the longstanding conviction that genuine laws of nature are inviolable.

Scientific thought experiments are not idle fantasies. They are grounded in shared commitments about the phenomena being studied. These set constraints on what is to be held fast and what is variable. They are imaginative exercises, not random collections of free-floating daydreams. Although the imagination is in principle free to entertain any ideas it likes, freedom, as Kant insists, is not lawlessness (Kant, 1993). Rather, in acting freely, thought experimenters, like other agents, obey laws they set for themselves – laws they think it is reasonable that they be bound by. They can justify the constellation of liberties and constraints they settle on by considering what is needed to bring out the features of interest. Thus Maxwell can ask: if the second law were merely statistical rather than universal, what would it show? Answer: a case where the overwhelmingly improbable actually happened. What would that look like? Well, imagine a tiny demon, one small enough and acute enough to discern the motions of individual gas molecules, and dexterous enough to be able to react to their trajectories. Then . . .

Both laboratory experiments and thought experiments, have a narrative structure, with a beginning, middle, and end (see Nersessian 1993). They typically begin in medias res – certain things have already happened, which set the stage for what follows. Those things may inform the choice of constraints, the grain of description, and the orientation on the phenomena. They serve as framing devices for what follows. Moreover, both laboratory experiments and thought experiments require interpretation. What shall we make of these findings? Opinions can diverge, not only with changes in background assumptions, but even when the background assumptions remain held fixed. The Einstein-Podolsky-Rosen thought experiment imagines a case where initially paired particles fly off in different directions. Once they are separated, the measurement of one should have no effect on the state of the other. But if we measure the position of one and apply the Schrödinger equation, we can determine that the other also has a definite position. This seems to violate the uncertainty principle, according to which unexamined particles have no position. What does the thought experiment show? Even the authors disagreed. Podolsky thought it showed that quantum mechanics is incomplete; Einstein thought it showed that either quantum mechanics is incomplete or that states of spatially separated objects are not independent of each other (Bokulich 2001). Others might think it undermines the uncertainty

principle itself. This sort of case discredits the stereotype that scientific conclusions are determinate and agreed upon while artistic ones are subject to interpretation. Science strives for univocality. So scientists typically design their thought experiments in hopes of achieving a univocal result. But such hopes can be dashed, and not always because the thought experiment is poorly designed. Sometimes a result is fruitfully Socratic. It shows that we do not understand what we thought we did.

Thought experiments are not peculiar to science. They are commonplace in philosophy and other disciplines as well. Some philosophical thought experiments, such as trolley problems, are relatively austere. They can be expressed in a few sentences. Some appeal to allegedly untutored intuitions. Again trolley problems are a good example. Their validity is not tied to any particular theory. Other thought experiments are more elaborate. They require stage setting. Some, such as Rawls's initial position, are embedded in theories which set the stage. Others, like Rousseau's *Emile*, stand alone. They are, like a novel, elaborate enough to set their own stage. Indeed, there is a continuum of cases from those that are clearly thought experiments through didactic fictions like *1984* and *Uncle Tom's Cabin* to more plainly literary fictions, like *Remembrance of Things Past* and *Mrs. Dalloway.* If we are going to insist that scientific and philosophical thought experiments embody and advance understanding, it will be difficult to deny that works of literature do too. Indeed, it will be difficult to deny that literary works function as thought experiments.

Gottfried Gabriel disagrees (2019). He maintains that unlike thought experiments, literary fictions are re-presentations in which something is 'shown in such a way that a fictionally reported event, due to its fictionality, loses the character of something historical-singular, and thus in becoming something particular, paradoxically gains a more general meaning' (Gabriel 2019, p.21). I hold that many scientific thought experiments do the same. Galileo's thought experiment that refuted the Aristotelian claim that heavier objects fall more quickly than lighter objects illustrates this (Galileo 1960). Consider, Galileo says, two objects: a cannonball and a musket ball. If we drop them from a tower, the Aristotelian maintains, the cannonball will reach the ground before the musket ball because the cannonball is heavier. Now consider a third object, obtained by tying a cannonball to a musket ball. It consists of the two original objects and a bit of rope. The composite object is heavier than the original objects. If we drop it, according to the Aristotelian, it should fall more quickly than the cannonball alone. But the musket ball attached to one end of the rope is falling more slowly than the cannonball. So it retards the composite object's fall. The composite should therefore fall more slowly

than the cannonball. One object cannot fall both more quickly and more slowly than another, so the Aristotelian position is false.

Arguably the final sentence – the conclusion of the thought experiment – is a proposition that is asserted. But everything leading up to it is a re-presentation of (imaginary) falling bodies. They 'lose the character of something historical-singular' and function generally to represent falling bodies as such. The scenario described does not obtain. If the descriptions of the falling bodies were asserted, they would be false, and disclose nothing about how objects fall. This the thought experiment has the re-presentational, 'as-if-ish' quality that Gabriel ascribes to fictions.

The parallels between literary fictions, thought experiments and laboratory experiments suggest that all three can advance understanding. Effective works in all three genres are felicitous falsehoods. But the epistemological challenge remains. How can they advance understanding if they do not literally and accurately represent the phenomena they purport to illuminate? How can an experiment that could not be performed, one that requires a demon that does not and could not exist, underwrite any understanding of actual heat transfer? How could a story about people chained in a cave until one escaped from his chains, emerged from the cave and saw the world outside reveal anything about appearance and reality? How could the machinations of The Party in *1984* disclose anything about Donald Trump's propensity to appeal to 'alternative facts'? I am not asking *whether* here, I'm asking *how*.

I have argued that the answer lies in exemplification (Elgin 2017). An epistemically effective fiction, whether in literature, philosophy, or science, exemplifies certain features that obtain in reality but are, or may typically be, difficult to discern in their natural setting. They may be overshadowed by more conspicuous, confounding factors. They may be hard to disentangle from their typical concomitants. But once they are identified in an artificial setting, we gain epistemic access to them, and can recognize them and appreciate their importance when we encounter them elsewhere.

To make this out, I have to say a bit about exemplification (see Goodman 1968, Elgin 1996). Exemplification is the relation between a sample and whatever it is a sample of. A sample problem worked out in a trigonometry textbook exemplifies the role of the law of cosines. It displays the way the law of cosines figures in the solution to the problem, and the problem itself is framed to be representative of a range of problems. A fabric sample exemplifies its color, pattern, texture, and weave. It shows what the particular fabric looks and feels like. Exemplification involves instantiation and reference. An exemplar refers to some of the properties it instantiates. It highlights those properties

and makes them manifest. A swatch of herringbone tweed is capable of exemplifying herringbone, for it instantiates that pattern. It cannot exemplify paisley, since it is not an instance of paisley.

The sample must also refer to the properties it exemplifies. The swatch has innumerable properties that are not exemplified. It is six centimeters long and four centimeters wide; it has ragged edges; it is a certain distance from the Eiffel Tower; it is not a giraffe. Although all of these features are instantiated, none of them is referred to by the sample, at least not in its standard use. Exemplification is selective. It highlights some properties of a sample by marginalizing or overshadowing others. Moreover, the connection between instantiation and reference is no coincidence. That is, exemplification cannot be Gettierized[1] The sample refers to a particular feature *via* its instantiation of that feature (see Vermeulen et. al. 2009).

Focus on textbook cases and commercial samples may suggest that exemplification is a device for displaying what is already known. This is not always so. An air quality inspector takes air samples to discover what no one yet knows. Her samples exemplify the levels of carbon monoxide in different regions of the building. An oncologist takes a biopsy to discover what no one yet knows – whether a tumor is cancerous. If the sample exemplifies malignant cells, it is; otherwise, probably not. In such cases, the features that the samples exemplify extend, rather than merely conveying knowledge or understanding. This is what happens in experiments. Even if the scientist has good reason to believe that an experimental result will exemplify a particular feature, he can be wrong. It may exemplify something unexpected. The Michelson-Morley experiment is a famous case in point. Although designed to exemplify the magnitude of ether drift, it exemplified the absence of ether and the inadequacy of the then current account of light propagation (Einstein 1916).

Features, as I use the term, can be static or dynamic, monadic or polyadic, descriptive or normative, thick or thin; they can be specific or general and at any level of abstraction. So every object has indefinitely many features. In principle any object that instantiates a feature (a property, pattern, or relation) can exemplify it. That is why it is so easy to adduce examples. Often, pointing to an instance suffices. But exemplification is not always so easy to achieve. Sometimes elaborate stage setting is needed to highlight a particular feature. If the feature is hard to discern, or if it is typically intermingled with or overshadowed by other features, a good deal of effort may be required to sideline the irrelevant concomitants. Consider, teaching shapes to preschool children. As Plato notes, shape is that which

1 Gettier (1963) showed that it is possible for someone to have a justified true belief without having knowledge because what makes the belief justified is not what makes it true. Because exemplification requires that an exemplar refer to a feature via the instantiation of that feature, this sort of divergence cannot occur in exemplification.

always accompanies color (Plato 1961). So you are never going to find a triangle without a color. (If a particular triangle, perhaps a wire frame, has no solid interior or has a transparent interior, we still see the color that currently lies behind it that it frames.) How then will you exemplify the shape without confounding it with a color? How do you show the children what a triangle is? Two obvious answers suggest themselves. We might tell them to ignore the color. If they can do that, all is well. But they are young, and may as yet have no clear grasp of color either. After all, color is also that which always accompanies shape. Alternatively we might display the same shape with different colors. Then we encourage them to see what is common to the blue one, the red one, and the green one. This is not a hard case. My point in mentioning it is to highlight how complex even an easy case is.

Where we cannot easily ignore irrelevant features in their natural setting, it makes sense to manufacture a setting that omits them. This is what we do in experimentation. We contrive a situation where confounding factors are either absent or controlled for. We then can focus on the behavior of the factors that remain. Science students seeking to discover whether water conducts electricity would not run their experiment using ordinary rain water or tap water. Such liquids contain impurities. If they found that these liquids conduct electricity, they still would not know whether the water or the impurities were the conductors. Rather, they would use distilled water – water from which, as far as they could tell, all impurities had been eliminated. Then a positive result is evidence that water itself conducts electricity. Where we cannot eliminate confounds, we can often control for them. Thus in an experiment to discover whether a chemical is carcinogenic, scientists expose about half the population of experimental mice to high doses of the chemical, while leaving the others unexposed. But – and this is critical – in every other respect the exposed mice and the control group will be alike, and will be treated alike. If both sets of mice develop cancer at about the same rate, their illness cannot be ascribed to the substance being tested. If the exposed mice exhibit a higher rate of cancer, the connection between the chemical and the incidence of cancer is exemplified. Although the scientists cannot eliminate all other sources of disease, by using suitable controls, they can neutralize the danger that the factors they have controlled for will mislead.

Thought experiments take things to even further extremes. Ordinary experiments are good for prizing apart features that typically commingle, thereby enabling us to discern the contributions the different features make to the phenomenon of interest. Thought experiments can go further, prizing apart features that, in reality, inevitably go together. In reality, persons are not subject to fission or fusion. A person with a single past has but a single future; a person with a single future has but one

past. Still, Parfit can ask: what would happen if fission and fusion were possible (Parfit 1984)? Suppose that for the first fifteen years Anna was a single, unified person, with one body, one set of experiences, one set of thoughts, memories, and so on. But at age fifteen she split in two – both $Anna_1$ and $Anna_2$ had the very same past, but went on to lead different lives after the split. Which, if either, would be the original Anna? The thought experiment prescinds from what is physically possible, to ask questions about the nature of personal identity. What exactly has to endure over time for two time slices to be time slices of the same person?

The rationale for asking what would happen in an unrealizable situation is that it sheds some light on what it is to be a person – something that is of interest in more mundane circumstances. Parfit's thought experiment exemplifies that personal identity at a time does not determine personal identity across time. It raises the question: what more do we need? Is there simply no fact of the matter as to whether $Anna_1$ or $Anna_2$ is identical to the original Anna? Could both be identical to the original, although they are not identical to each other? Is personal identity over time contingent on the fact that we neither fission nor fuse? The thought experiment exemplifies a host of issues about personal identity that ordinarily escape our notice because we cavalierly, if tacitly, assume each of us has and will continue to have an indisputable claim to continue over time to be the unique person that she is.

My contention that thought experiments exemplify features they share with their targets raises a worry. Exemplification requires instantiation. But thought experiments are virtual. No actual gas segregates as Maxwell imagines; no actual person divides as Parfit imagines. But if nothing instantiates the features, nothing exemplifies them. This is so. In saying that someone undergoes fusion, or some gas segregates in a way that violates entropy, I speak loosely. Strictly, the thought experiments instantiate and exemplify abstract properties that are instantiated in representations and also instantiated in actual, material phenomena. The properties in question are not peculiar to gases or persons; they are abstract, often mathematical, properties that can be shared by virtual and real items – by gas molecules and gas-molecule-representations, by persons and person-representations.

How does this bear on the epistemic contributions of literary fiction? Let's look at a case in some detail. In the *Nicomachean Ethics* Aristotle ventures the hypothesis that we should call no man happy until he is dead (Book I, Chapter 10). This sounds implausible. Part of the reason is that there is no good English translation for the Greek word 'eudaimonia'. 'Happiness' seems too subjective, and potentially too short-lived. It is possible to be happy for a mere twenty minutes, and possible to be happy even if one's happiness is based on misinformation. 'Flourishing' has been suggested as an

alternative translation. Then the hypothesis is that we should call no man flourishing until he is dead. This has the advantage that flourishing is not a purely psychological matter; it involves things actually going well for a person. Moreover, flourishing extends over time. Only a very short-lived creature could flourish for a mere twenty minutes. Aristotle's question then is whether flourishing requires an entire life-time. He recognizes that flourishing involves both character and circumstances. He maintains that only a morally and intellectually virtuous person can flourish, and only if he is not a victim of grave misfortune. A flourishing life is a life well-lived.

If we just read the *Ethics*, the contention that a *complete* lifetime is required is less than persuasive. It is plausible that an extended period of well-being is required, but perhaps not a complete life. Aristotle acknowledges that there is a temptation to think that a person could flourish throughout much, but not all of his life. But, he says, 'many changes occur in life, and all manner of chances, and the most prosperous may fall into great misfortunes in old age, as is told of Priam in the Trojan Cycle; and one who experienced such chances and has ended wretchedly no one calls happy' (Aristotle 1941a 1100a5-9). Priam, the aged king of Troy, does not constitute a compelling case for Aristotle's position. Priam's life evidently went well until the loss of Troy. His contemporaries would have been justified in calling him flourishing before the fall. Moreover, even after the fall, we may be inclined to say that he flourished throughout most of his life, even though things ended badly. That is, we do not automatically rescind our previous assessment when we find out that his life came to a wretched end. To make a better case, we need a better example – one where late wretchedness discredits previous (apparent) good fortune.

Oedipus Rex supplies one. I will argue that it serves as a thought experiment that tests Aristotle's hypothesis. Aristotle was evidently electrified by the play. The normative requirements on a good tragedy that he set out in the *Poetics* are practically a template of *Oedipus Rex*. To make my case, I do not need to show that Aristotle took the play to be a thought experiment for his theory. My concern is with how it functions, not with how Aristotle construed it. I need to do two things: first, show that the play exemplifies important features of Aristotle's account of flourishing; second, if we are to use it to support Aristotle's hypothesis rather than just to explicate it, I need to show that they can be projected so as to enable us to better understand something about the human condition.

When the play begins, Oedipus is king of Thebes, a city currently suffering a horrific plague. He is renowned for his intelligence, having years earlier solved the riddle of the sphinx, thereby delivering the city from her predation. In the first scene, he displays compassion for the city's suffering and

confidence in his own abilities. He comforts the children and the elders, assuring them that he will find a way to alleviate the plague. He thus exemplifies a cluster of moral and intellectual virtues that, according to Aristotle, are needed to flourish; he is intelligent, compassionate, brave, resourceful. Oedipus is neither a monster nor a saint. Despite his virtues, he is impulsive, quick to anger, perhaps overconfident. He is, then, a man like us, perhaps a little better, but no worse. He exemplifies the temperament Aristotle demands of a tragic hero. All of these requirements are spelled out in the *Poetics* (Aristotle 1941b 1454a16-1454b5).

The oracle reveals that the plague will be alleviated only when the killer of Laius, the previous king of Thebes, is identified and brought to justice. Unbeknownst to Oedipus, he is Laius's killer. He knows that he killed the old man at the crossroad, but not that that man was Laius. Nor, of course, does he know that Laius was his father. According to prophecy, the son born to Jocasta and Laius would kill his father and marry his mother. To prevent this from happening, that baby was left exposed on a hillside to die. He was rescued, brought to Corinth and raised as the son of the Corinthian king and queen. He was never told that he had been adopted. The baby grew up to be Oedipus. As the play progresses, it emerges that the prophecy has come true. Oedipus had killed his father and married his mother.

Oedipus was, to be sure, a victim of fate; his doom was foretold. Nevertheless, he was not completely blameless. Granted, he did not know that he was committing patricide and incest. But his bellicosity, egotism, and pride led him to missteps. Having learned of the prophecy, he fled Corinth to avoid killing the man whom he took to be his father and marrying the woman whom he took to be his mother. But it evidently never occurred to him that a sure-fire way to avoid fulfilling the prophecy would be to refrain from killing anyone, and to refrain from marrying an older woman. He took offense at the man at the crossroad, and killed him in a rage. Fortune (or anyway, coincidence) played a role. He just happened to be at the crossroad at the time when Laius and his party were there. He just happened to go to Thebes, rather than some other city, having killed the old man. Still, his character played a major role in what happened.

Not all of the traits that led to Oedipus's downfall should be characterized as flaws. His insistence on knowing the truth was, arguably, a virtue. The self-confidence that enabled him to confront the sphinx and the intelligence that enabled him to solve the riddle seem like virtues as well. Even his fortitude in leaving his home in Corinth to prevent the prophecy from being realized is admirable. So the situation is far more complicated than a focus on obvious flaws would suggest. Much

human vulnerability derives from our inevitable ignorance. Oedipus did not know who his parents were. He did not know that the man at the crossroad was his father or that Jocasta was his mother. He did not know that he himself was the source of the plague. Many of his missteps were due to ignorance.

Oedipus could not escape his fate; it was foreordained. Nevertheless he bore some responsibility. His hubris led him to think he could outrun the prophecy. His belligerence led him to take offense at the crossroad, and to violently attack the supposed offender. Toward the end of the play, his egocentricity led him to castigate Jocasta, rather than recognizing that she was as much a victim of malign fate as he was. We might even consider his relentlessness in looking for the truth to be grounded in hubris. As it became increasingly evident that Oedipus was the perpetrator of abominable crimes, Jocasta repeatedly advised him not to look any further. He dismissed her concerns, confident that he would somehow be able to master whatever he found.

How does this bear on Aristotle's hypothesis? The discovery that he had killed his father and married his mother did not merely blight Oedipus's old age, as the fall of Troy blighted Priam's. It invalidated what came before. He had never been the upstanding, honorable man that he and everyone else thought he was. He killed his own father. His claim on the throne of Thebes was due in part to regicide. His relations to his wife/mother, siblings/children were infused with corruption. The members of his family would never again be able to look on him or on their relations to him without revulsion. The life he thought he was living was not the life he actually lived. He had never flourished, even though throughout most of his life he had every reason to think that he did. The longstanding conviction that Oedipus's life was well lived turns out to be mistaken.

Aristotle goes on to ask whether even at death we have enough evidence to decide whether a person had flourished. Can the fortunes of his friends and family after his death change his status? He answers 'maybe' (Aristotle 1941a 1100 10-30). Again *Oedipus Rex* shows why. At the end of the play, Oedipus recognized how deeply and irrevocably his actions had blighted his daughters' futures.

> I weep when I think of the bitterness there will be in your lives, how you must live before the world . . . When you're ripe for marriage, who will he be, the man who'll risk to take such infamy as shall cling to my children, to bring hurt on them and those that marry with them? (Sophocles 1187-1195).

He is not dead when he says these words, but he sees how, after his death, his daughters' futures will inevitably be tarnished by his actions. Perhaps the fact that they will be in no position to flourish further undermines his flourishing as well.

I could go into more detail, but this is enough to show why *Oedipus Rex* can be read as a thought experiment testing Aristotle's hypothesis. We might then take it to function in the way that Maxwell's demon does – as teasing out the commitments of a theory. Nothing in Maxwell's thought experiment gives us reason to accept the second law of thermodynamics. It only shows what, if we accept it, we are committing ourselves to. Should we think the same about *Oedipus Rex*? Should we think that it shows only what accepting Aristotle's hypothesis commits us to? In that case, it simply exemplifies what follows from the theory. That's not nothing. But what we really want to know is whether Aristotle is right. *Should* we call no man flourishing until he is dead, or at least consider any earlier verdict merely provisional?

Can we project the features exemplified in the play onto human experience? Contemporary thinkers would be reluctant to directly export anything about oracles, prophecy, or fate onto reality. But there are less theologically loaded counterparts that we might find congenial. We regularly make predictions about others, based on our assessments of their character. We say of one person that he will come to a bad end; of another, that she will go far; of a third, that his impulsiveness will cause him difficulties. We can't make literally oracular pronouncements, but often we have enough experience and evidence that our predictions carry considerable weight. We don't believe in fate, strictly speaking. But we do believe that circumstances beyond a person's control figure in her successes and failures. He was in the right place at the right time; she was just unlucky that the flight was canceled, etc. Arguably, then, generic features that in Greek tragedy are described as oracles, fate, and prophecy, are also realized in our more mundane ways of describing the interface between character and circumstances. If we interpret the play as exemplifying these generic features, we can project them onto actual human lives. Since practically every human tragedy for which the agent is at all responsible is due to an unfortunate commingling of character and chance, doing so is straightforward.

But what of the idea that a person's entire life can be discredited in the way that Oedipus's was? Is that something we can project? If not, then even if we learn something important about the human condition from *Oedipus Rex*, it does not support the hypothesis that we should call no man flourishing until he is dead. Flourishing for a while might still be possible. Here is an example that supports

Aristotle's position, taken from the sports pages of the *Washington Post*. It strikes me as important that it is drawn from such a mundane source.

First, the background: In 2011 it emerged that Jerry Sandusky, a former assistant football coach at Penn State University, had been sexually abusing young boys for more than thirty years. Much of the abuse took place in the Penn State athletic facilities. One of the mysteries was how Joe Paterno, the longtime head football coach, whom many considered a bastion of integrity, could have turned a blind eye to his assistant's appalling actions. Admirers of Paterno, some of whom had known him for decades, were dumbfounded. How could he have been unaware of Sandusky's behavior? If he was not unaware, how could he have let it go on? Sportswriter Thomas Boswell ventures an answer:

> Everybody has weak spots in their character, fault lines where the right earthquake at the wrong time can lead to personal catastrophe. Most of us are fortunate that our worst experience doesn't hit us with its biggest jolt in exactly the areas where our flaws or poor judgment or vanity is most dangerously in play. It's part good luck if we don't disgrace ourselves. But when it does happen, as appears to be the case with Joe Paterno, that's when we witness personal disasters that seem so painful and, in the context of a well-lived life, so unfair that we feel a deep sadness even as we simultaneously realize that the person at the center of the storm can never avoid full accountability. […] Forces collide, conspire, confuse, and an icon of integrity fails to act, fails to see […] (Boswell 2011).

A great man, beset by hubris, does terrible things and is brought down by his tragic flaw. To readers or viewers of *Oedipus Rex,* this sounds familiar. Once it became evident that Paterno did nothing to stop Sandusky – that he ignored the situation as long as he could, then merely had Sandusky dismissed from his job rather than arrested; that he privileged the reputation of the team, the university, and himself over the safety of children – it was clear that he was not the man of integrity he had been thought to be, and probably thought himself to be. Indeed, Oedipus was less blameworthy than Paterno, and not only because of the inescapability of fate. Ignoring rampant pedophilia is vastly more abhorrent than inadvertent patricide and incest among consenting adults. Still, *Oedipus Rex* provides a template for understanding Paterno. Having seen the pattern in fiction, we are in a position to recognize it when we encounter it in reality. One might think (and hope) that this case is exceptional. Such late-breaking reversals of apparent fortune do not happen every day. But Aristotle does not claim that provisional verdicts about flourishing are regularly overturned. His point is that the possibility is always there. As Boswell says, everyone has character traits that, combined with unlucky circumstances, can lead to

disaster. This suggests that we might take a broader view of Aristotle's hypothesis than just reading it as a warning that mid-life verdicts are provisional. To appreciate *why* we should call no man flourishing until he is dead, we need to recognize the sources of our vulnerability, among them the inevitable limitations on what we can know, both about ourselves and about our circumstances. We need to recognize that a trait that in some situations is a virtue may in other situations be a flaw. Among the virtues we should cultivate then is an appreciation of our vulnerability to factors beyond our control, including our inevitable ignorance (and we should recognize that even that virtue may in some contexts be a flaw).

We need not consider *Oedipus Rex* in the context of Aristotle's theory to construe it as a thought experiment. Many thought experiments are free-floating; they are not allied to any particular theory. Suppose we distance ourselves from Aristotle's views on tragedy and ask what we might discover then. Let's treat it as a free-floating thought experiment like Parfit's. We can still take it to test the hypothesis that we should call no one flourishing until he is dead. The last line of the play is: 'Count no mortal happy till he has passed the final limit of his life secure from pain' (Sophocles 1530). Sophocles himself seems to venture the Aristotelian hypothesis. Much of what I sketched above would still hold. Aristotle insists that a tragedy should focus on a single protagonist. So while giving an Aristotelian reading to the play, we focused on Oedipus exclusively, treating the other characters as peripheral. But we might interpret the play differently, letting Jocasta share the spotlight. She is at least as much at fault and at least as much a victim of fate as Oedipus. She initiated the chain of events by attempting (and failing at) infanticide. She then married the murderer of her husband, who was also her king. She engaged in incest, giving birth to four further children, whose own prospects of flourishing were preempted by the conditions of their birth. She knew of the prophecy. Rather than attempting infanticide, she could have blocked the realization of the prophecy by refraining from marrying a younger man. She was driven to suicide by the realization of what she had done.

We might then reinterpret the play as a double tragedy, where the two were jointly responsible for their conjoined fate. If we extrapolate the features exemplified under this reading, we may come to understand something about seriously bad marriages. Both parties have their faults, but perhaps taken separately they are relatively minor. The disaster comes from the interanimation of those faults. We have then at least the beginning of an answer to the question: "How can two such nice people create such misery for themselves and those who care about them?" Moreover, by focusing on Jocasta, we might see something more. Oedipus is ignorant of his situation throughout most of the play. He

manages to remain in denial until the herdsman tells his story. Then in an instant he flips from ignorance to knowledge and immediately knows that he knows that the prophecy has been fulfilled. Through Jocasta, we might come not only to appreciate the epistemological point that it is possible to know that *p* without knowing that one knows that *p*, but also to recognize what that condition looks like and what it costs.

I suggested that the way fictions and other thought experiments advance understanding is that they exemplify features we can then project onto actuality. Maxwell's demon exemplifies the statistical nature of the second law of thermodynamics; *Oedipus Rex* exemplifies the interdependence of character and circumstance in human flourishing. It might seem, though, that the thought experiments *per se* do no more than provide plausible hypotheses. The exemplification of character and circumstance in *Oedipus Rex* makes certain features salient and sensitizes us to their potential significance. The play supplies a scenario that shows a hypothesis to be plausible. We then need to see whether actual human lives display the same pattern. In one sense, this is correct. A feature that is exemplified in a fiction or other thought experiment may fail to project onto the phenomena it bears on. This is a standard problem with sampling. Even a well taken sample can be unrepresentative of the phenomena, hence misleading. When we generalize from a sample we reason inductively. And in induction, there are no guarantees. But we should not therefore conclude that fictions are epistemically idle. David Lewis explains why.

> We who have lived in the world for a while have plenty of evidence, but we may not have learned as much from it as we could have done. This evidence bears on a certain proposition. If only that proposition is formulated, straightway it will be apparent that we have very good evidence for it. If not, we will continue not to know it. Here, fiction can help us. If we are given a fiction such that the proposition is obviously true in it, we are led to ask: and is it also true *simpliciter*? And sometimes, when we have plenty of unappreciated evidence, to ask the question is to know the answer. (Lewis 1983, p. 279).

References

Aristotle (1941a). *Nicomachean Ethics* in *The Basic Works of Aristotle,* Richard McKeon editor. New York: Random House, pp. 935-1026.

Aristotle (1941b). *Poetics* in *The Basic Works of Aristotle,* Richard McKeon editor. New York: Random House, pp. 1455-1487.

Bokulich, Alisa (2001). 'Rethinking Thought Experiments'. *Perspectives on Science* 9: 285-307.

Boswell, Thomas (2011). 'Penn State Coach Joe Paterno Reaches a Sad Conclusion' *Washington Post,* November 9.

Einstein, Albert (2016). *Relativity: The Special and General Theory.* New York: Taylor and Francis.

Elgin, Catherine (2017). *True Enough*. Cambridge MA: MIT Press.

Elgin, Catherine (1996). *Considered Judgment*. Princeton: Princeton University Press.

Gabriel, Gottfried (2019). 'The Cognitive Value and Ethical Relevance of Fictional Literature'. *Literature as Thought Experiment?* Falk Bronmüller, Johannes Franzen, Mathis Lessau editors. Leiden: Wilhelm Fink Verlag, pp. 17-30.

Galilei, Galileo (1960). *On Motion and On Mechanics: Comprising de Motu*. Madison: University of Wisconsin Press.

Gettier, Edmund (1963). 'Is Knowledge Justified True Belief?' *Analysis* 23: 121-123.

Goodman, Nelson (1968). *Languages of Art*. Indianapolis: Hackett.

Kant, Immanuel (1993). *Grounding of the Metaphysics of Morals*. Indianapolis: Hackett.

Lewis, David (1983). 'Truth in Fiction: Postscript' *Philosophical Papers* vol 1. Oxford: Oxford University Press, pp. 276-280.

Maxwell, James Clerk (1872). *Theory of Heat.* New York: Appleton.

Nersessian, Nancy (1983). 'In the Theoretician's Laboratory: Thought Experimenting as Mental Modeling' *PSA 1992* vol. 2. Philosophy of Science Association, pp. 291-302.

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

Plato (1961). *Meno* in *The Collected Dialogues of Plato.* New York: Pantheon, pp. 353-384.

Sophocles (1960). *Oedipus the King* in *Greek Tragedies* vol 1. Chicago: University of Chicago Press, pp. 107-176.

Vermeulen, Inga, Georg Brun, Cristoph Baumberger (2009). 'Five Ways of (not) Defining Exemplification' *Nelson Goodman: From Logic to Art.* Frankfurt a. M. Ontos: pp. 219-250.