Coase and Models as Foils

By: Brian C. Albrecht and Brian Kogelmann

Abstract: This paper articulates a way of reasoning with economic models that, we believe, has not been sufficiently elaborated in the literature. We begin by highlighting an example of the kind of reasoning we are interested in by taking a close look at the work of Ronald Coase. Then we argue that current accounts of modeling in the philosophy of economics literature are unable to account for the kind of reasoning in question. We fill this gap by proposing what we call *models as foils.* The basic idea is that a model can give us insights concerning feature *f* of the actual world by constructing a model world in which *f* is not present; from here, economists draw inductive inferences concerning the role *f* plays in the real world by comparing the real world to the world of the *f*-less model.

"Even imaginary constructions which are inconceivable, self-contradictory, or unrealizable can render useful, even indispensable services in the comprehension of reality, provided the economist knows how to use them properly."

- Ludwig von Mises, *Human Action: A Treatise on Economics*, p. 236

1. Introduction

Economists are in the business of building models: models of markets, models of voting behavior, bargaining models, and so on. Philosophers of economics, on the other hand, are interested in (among other things) how models are used in economic inquiry. Our paper takes up this topic. Some argue that models are valuable tools of economic inquiry when they generate testable predictions that can then either be empirically verified or falsified. Others hold that models are valuable as tools of economic inquiry when they act as credible or plausible worlds that allow us to draw inductive inferences about our actual world. Of course, other answers can be found in the literature, each proposing different ways in which models are used in economic reasoning.

We – an economist and a philosopher – are pluralists about models. We believe that there are many valuable things models can do. Moreover, these functions need not compete with one another. Some models generate testable predictions, others are credible worlds, still others are used to isolate or gain insight on capacities, and so on. We are thus not interested in adjudicating between different accounts of how models are used in economic reasoning. Instead, we wish to *add* to the list. Though we do not think many will be particularly surprised at the form of reasoning we wish to highlight in this paper, we do believe that it has yet to be formally articulated within contemporary work in the philosophy of economics, though there is some similarity between this form of reasoning and John Stuart Mill's *method of difference*, something we discuss in §4 below. Of course, we do not claim that all models must be interpreted in the way we propose. Rather, we wish to highlight one more way in which models can serve as valuable tools, though we think a significant one. Indeed, as we shall see, one of the most celebrated models in economic theory gains its value precisely by fulfilling the function we articulate.

As motivation and a case study in this form of reasoning, we start by looking to Ronald Coase's model of a world without transaction costs from his seminal paper "The Problem of Social Cost" (§2).[1] The general idea here is that Coase wants to draw a conclusion concerning the function some feature *f* of the real world plays by looking to a model world in which feature *f* is conspicuously absent. Though this seems to be a plausible way of reasoning, contemporary accounts of economic modeling, we argue, cannot account for the way in which Coase draws such inferences (§3).

This opens the door for us to articulate a new account of how models are used in economic reasoning, what we call *models as foils* (§4). The basic idea is that models can give us insights concerning feature *f* of the actual world by constructing a model world in which *f* is not present; from here, economists such as Coase draw inductive inferences concerning the role *f* plays in the real world by comparing the real world to the world of the *f*-less model.[2] Though again, we do not think many in the philosophy of economics literature will be particularly surprised by this style of

[1] Coase is not the only person who reasons in the way we highlight. Indeed, we believe this is how those in the Austrian economics tradition interpret equilibrium models. Relevant works here include Mises (1949/2007: 245-260); Hayek (1946/2014: 109); Hayek (1961/2014: 420, 425); Kirzner (1973/2013: 11).

[2] Our account of models as foils is inspired by Boettke (1997)'s insightful history of general equilibrium theory in the latter half of the twentieth century, particularly Boettke's analysis concerning how different schools of thought interpreted the Arrow-Debreu result in diverging ways.

reasoning, we certainly *do* think that this form of reasoning fails to receive adequate attention in the literature. We hope our paper is a first step toward remedying this.

2. Coase on Transaction Costs

Coase begins his 1960 paper by telling us that he is interested in "those actions of business firms which have harmful effects on others" (Coase 1960/1988: 95). That is, he is interested in externalities: costs which parties not privy to voluntary exchange pay nonetheless. It is intuitively plausible that underlying property law matters in how externalities get resolved. If Althea wants to run loud machinery on her property, and Bertha wants peace and quiet on her adjacent property, then how this dispute will be resolved will depend on the allocation of rights: does Althea have a right to run machinery, or does Bertha have a right to peace and quiet? If Althea has the right then the machinery will run; if Bertha has the right then it will not.

But, Coase shows, this is not necessarily so. Coase begins his analysis by assuming that "the pricing system works smoothly (strictly this means that the operation of a pricing system is without cost)" (Coase 1960/1988: 97). That is, he assumes there are zero transaction costs. From here Coase goes through examples showing that, in such a world, (*i*) all disputes are resolved in an efficient manner, but more stunningly (*ii*) all disputes result in the *same* efficient outcome regardless what the relevant property law says: "with costless market transactions, the decision of the courts concerning liability for damage would be without effect on the allocation of resources" (Coase 1960/1988: 106). To continue our toy example from the paragraph above, if there are no transaction costs then the dispute between Althea and Bertha will be (*i*) resolved in an efficient manner and (*ii*) be resolved in the *same* manner regardless of whether Althea has a right to use noisy machinery or whether Bertha has a right to enjoy peace and quiet. In such a world, property rights are simply irrelevant.

Though it is not a mathematical model of the kind most contemporary economists build, Coase's imaginary world can be studied as a model nonetheless. But, given the highly unrealistic assumption of no transaction costs, what is the point of such a model? What purpose does it serve in economic inquiry? This is a controversial question. For many, the takeaway from this model is the so-called Coase Theorem – coined by George Stigler – which says that "under perfect

competition private and social costs will be equal" (Stigler 1966: 113). Coase's model is thus interpreted as another theorem of welfare economics: absent transaction costs – the typical assumptions of most general equilibrium models at the time – all externalities will be bargained to an efficient point.

Note that the Coase Theorem as articulated by Stigler tells us little about the actual world. This is puzzling, for Coase understood his model as telling us something deeply important about our actual world.[3] He writes:

> "The Problem of Social Cost," in which these views were presented in a systematic way, has been widely cited and discussed in the economics literature. But its influence on economic analysis has been less beneficial than I had hoped. The discussion has largely been devoted to sections III and IV of the article and even here has concentrated on the so-called "Coase Theorem," neglecting other aspects of the analysis. In sections III and IV, I examined what would happen in a world in which transaction costs were assumed to be zero. My aim in doing so was *not* to describe what life would be like in such a world but to provide a simple setting in which to develop the analysis and, what was even more important, *to make clear the fundamental role which transaction costs do, and should, play in fashioning the institutions which make up the economic system* (Coase 1988b: 13) (emphasis ours).[4]

The purpose of Coase's model, *contra* Stigler but how we and Coase see it,[5] is not to derive propositions concerning what would happen in the imaginary world of zero transaction costs and perfect competition. Rather, the purpose is to tell us something about our actual world, *even though* the model world departs radically from reality as we know it. In particular, the model is meant to tell us something about the role transaction costs play in the real world, even though transaction costs are conspicuously absent from the model world. We do not want to get into a fight concerning the uniquely correct way of interpreting Coase's model. Instead, we take for granted that there is

[3] For an overview of how the Coase theorem has been interpreted by economists see Medema (2011).
[4] For similar remarks about how Coase believed his model was misunderstood, see Coase (1988a: 174); Coase (1988b: 15).
[5] For another who interprets Coase along the same lines as we do, see what Bertrand (2010) calls the "heuristic role" of the Coase theorem.

some merit to Coase's interpretation of his own model, and then focus on the inductive strategy he employs in drawing the conclusions he does about the actual world.

What does Coase think his model teaches us about the actual world? Coase takes the main point of his model to be as follows:

> Of course, if market transactions were costless, all that matters (questions of equity apart) is that the rights of the various parties should be well defined and the results of legal action easy to forecast. But as we have seen, the situation is quite different when market transactions are so costly as to make it difficult to change the arrangement of rights established by the law. In such cases, the courts directly influence economic activity. It would therefore seem desirable that the courts should understand the economic consequences of their decisions and should, insofar as this is possible without creating too much uncertainty about the legal position itself, take these consequences into account when making their decisions. Even when it is possible to change the legal delimitation of rights through market transactions, it is obviously desirable to reduce the need for such transactions and thus reduce the employment of resources in carrying them out (Coase 1960/1988: 119).

We interpret this passage as Coase drawing two key conclusions from his model. First: it is *precisely because* of transaction costs that different allocations of property rights result in diverging allocations of goods with differing levels of efficiency. That is, when and because transaction costs are present, the dispute between Althea and Bertha will not necessarily get resolved in the same manner depending on who has the relevant right (whether Althea has a right to use noisy machinery on her property or whether Bertha has a right to uninterrupted enjoyment of her property). Indeed, there is no reason to think that these different ways of specifying the relevant rights will lead to an efficient outcome in either case. This leads to Coase's second conclusion, which is normative: because transaction costs have massive effects on how different allocations of property rights affect economic efficiency, we should structure property law to attain the most efficient outcome possible.

We will not be interested in Coase's normative conclusion for the rest of the paper. But we will be interested in his positive conclusion. More specifically, we will be interested in how he

draws this positive conclusion – a conclusion about the relationships between property law, allocations of goods, and efficiency *in the real world* – from a model that is radically unrealistic in its assumptions. How exactly is the model being used in economic reasoning?[6] And, moreover, are contemporary accounts articulating the role of models in economic reasoning able to capture the way Coase uses this model in his own reasoning?

3. Philosophical Accounts of Modeling

The last section highlighted a special kind of reasoning economists sometimes employ with their models by examining closely the work of Coase. In building a model, the economist purposefully leaves a feature *f* of the world out of the model. This is not atypical. But what is interesting here is that the economist then draws conclusions concerning the function this absent feature *f* actually plays in the real world by looking closely at the world of the *f*-less model. With Coase the missing feature is transaction costs – we learn about the role transaction costs play in the real world by examining the model world which transaction costs are absent. Before specifying more clearly exactly how this form of counterfactual reasoning proceeds, we wish to show in the current section that several contemporary accounts of reasoning with economic models found in the philosophical literature have trouble making sense of this particular mode of analysis.[7] Note, this is not to say that such accounts of economic modeling are flawed. As we noted in the introduction: we are pluralists about models, believing that models can do many different things and be valuable in many different ways. But insofar as currents philosophical accounts of modeling, as we argue in this section, have difficulty making sense of what Coase is doing, we believe that such a fact leaves open the door for us to articulate a new account of reasoning with economic models, which we offer in the paper's final section.

3.1 *Prediction.*

[6] There have been some attempts at articulating the kind of reasoning Coase employs, but we believe that the account we offer in §4 below is distinct from these. See Hsiung and Gunning (2002); Hsiung (2004).
[7] Due to space constraints our overview of the literature is incomplete. For a more complete and very recent overview see Mireles-Flores (2018).

Models are often used in economic inquiry as tools for generating hypotheses that can then be tested. Milton Friedman (1953) is often seen as an advocate of this understanding of models, but recent work suggests his account is actually subject to deep ambiguities (e.g., Mäki 2009). A clearer account of this way of reasoning with models is offered by Kevin A. Clarke and David M. Primo, who distill what they call *hypothetico-deductivism* into a three-step process: "The first step is to write down the model… The second step is to derive deductive implications, or predictions, from the model. The third step is to test the model by checking these implications against the data" (Clarke and Primo 2012: 23).[8]

The reasoning here is quite clear: a model is built, hypotheses are generated from the model, and then these hypotheses are tested.[9] Though this is certainly a valuable way of reasoning with models, we think that such an account clearly fails to describe Coase's inferential strategy. This can most clearly be seen by thinking about what kind of hypotheses Coase's model of a world without transaction costs generates. Such a model would seem to generate the following hypothesis: that different schemes of property rights should *not* result in different allocations of resources or differing levels of efficiency (for this is, after all, the outcome of Coase's model). While this is proposition is in principle testable, the lesson Coase draws from his model is precisely the opposite of the model's outcome: different schemes of property rights *do* matter and importantly so for the distribution of goods as well as how efficient these distributions are. This asymmetry between (*i*) what can plausibly be construed as the testable hypothesis of Coase's model and (*ii*) the conclusion Coase actually draws from the model suggest that it was not intended to generate a testable hypothesis in the first place.

Indeed, some economists *do* test Coase's model, proffering hypotheses that are inconsistent with those conclusions that Coase himself draws. For example, Elizabeth Hoffman and Matthew L. Spitzer (1982) test the Coase Theorem in a laboratory setting. Though the authors interpret Coase as saying that a "change in a liability rule will leave the agents' production and consumption decisions both unchanged and economically efficient" – the precise opposite of what Coase concluded – they focus their test on whether agents' production and consumption decisions are efficient, not whether the liability rule changes the allocation. The authors find that when

[8] Similar here is Alexandrova's view of models as *open formulae*. On this account, "models are used as suggestions for developing causal hypotheses that can then be tested by experiment" (Alexandrova 2008: 396).

transactions costs are absent, all two-person groups negotiated to the efficient allocation. For Hoffman and Spitzer, the testable implications of Coase's model are about what happens in a world without transactions costs or with minor transactions costs – this should not be surprising, for this *is* what happens in the model world. But for Coase, the model was intended to highlight the effects transaction costs have in the real world, not to generate hypotheses about what would happen in a fantasy world in which these costs were completely absent.

3.2 *Approximations and Credible Worlds*.

Distinct from understandings models as generating testable hypotheses, Allan Gibbard and Hal R. Varian argue that models are used in reasoning as *approximations*.[10] The general idea here is that a model explains some feature of economic reality by being approximately close to this reality: "an explanation by the approximate truth of a model takes the following form. First, if the assumptions of the applied model were *true*, the conclusions would be—here the proof is mathematical. Second, the assumptions in fact are sufficiently close to the truth to make the conclusions approximately true" (Gibbard and Varian 1978: 670).

In what we believe is a similar vein, Robert Sugden argues that models are employed in economic reasoning as *credible worlds*. According to Sugden, a model gains its value by allowing us to draw inductive inferences about our actual world from the model world. But, as Sugden notes, "if we are to make inductive inferences from the world of a model to the real world, we must recognize some significant similarity between those two worlds" (Sugden 2000: 23). More bluntly: "I want to suggest that we can have more confidence in [our inductive inferences], the greater the extent to which we can understand the relevant model as a description of how the world *could* be" (Sugden 2000 24).[11] Both the Gibbard-Varian and Sugden accounts of reasoning with models thus understand models as being instructive in virtue of the fact that they are reasonably close representations of reality.

[10] Gibbard and Varian suggest another way of reasoning with models, which we discuss below.
[11] Ylikoski and Aydinonat (2014: 19) take this view further, offering what they call the *family of models thesis*, which essentially says that simplified approximately true or credible world models are not useful on their own. Rather, such models gain value "when considered in the context of a family of related models."

Like before, we believe that such accounts of reasoning with economic models – that models gain their value by explaining the world through, in some general sense, offering reasonably accurate depictions of the world – cannot explain Coase's reasoning. For, in assuming away all transaction costs, Coase is clearly not constructing a credible world or some close approximation of reality. Indeed, common reflection, as well as rigorous empirical estimation, suggests that transaction costs are pervasive and massively so. As an example of this, the classic estimate by John Wallis and Douglass North (1986) puts transaction costs sectors of the economy at 55% of GNP in 1970. Recent empirical work reaches the same conclusion: transaction costs are quite large (e.g., Libecap and Lueck 2011). This omission of transaction costs in Coase's model is not like leaving out the color of the sky in order to focus on something else. Transaction costs are the focus. As such, imagining bargaining situations in which transaction costs are completely absent can in no way be construed as credible or approximately close to the real world. For this reason, the current account of reasoning with models cannot capture what it is Coase is doing.

3.3 *Derivational Robustness*.

On another account of reasoning with economic models, the purpose of modeling is to allow the theorist to engage in some kind of robustness analysis (Kuorikoski and Lehtinen 2009: 125-128; Kuorikoski *et al* 2010: 544-545). The idea here is to show that a result is constant over a changing set of assumptions. This is significant because "a result is more likely to be real or reliable if a number of different and mutually independent routes lead to the same conclusion" (Kuorikoski *et al* 2010: 544). As an example, suppose that we have a model consisting of features f_1 and f_2 that induces outcome *O*. How confident should we be in the accuracy of the model, though? According to the derivational robustness framework, we can increase our confidence in the model by changing the assumptions and showing that outcome *O* still obtains. So perhaps a different model with features f_1, f_2, and f_3 induces *O*; then perhaps a different model with features f_1 and f_3 induces *O*. This should increase our confidence in the reliability of the original result. The purpose of modeling is thus to generate models that yield outcomes that are then robust to changing assumptions of the model. When we do this, we can be confident in the results of our models.

Importantly, considerations of robustness help explain how one can use a model of the world with deeply unrealistic features (like zero transaction costs) to explain aspects of the actual world. For the deeply unrealistic feature is one of the variables that changes and, if all goes well, the outcome of the model is constant in the face of such alterations. When this is the case we should have increased confidence in our model. Though such a framework might be able to explain why some models assume away all transaction costs, it cannot explain why Coase specifically did so. Indeed, the outcome of the model – that all externalities are bargained to the same efficient point – is *not* robust under changing assumptions. Once transaction costs are added back to the model, everything changes. In Coase's words: "But as we have seen, the situation is quite different when market transactions are so costly as to make it difficult to change the arrangement of rights established by the law" (Coase 1960/1988: 119).

Not only is the model not robust to the addition of transaction costs, but this seems to be one of the central points Coase seeks to highlight. It is precisely the *difference* between (*i*) how externalities are bargained in a world *without* transaction costs and (*ii*) how externalities are bargained in a world *with* transactions that allows Coase to draw the conclusion he does: that transaction costs are the reason why property rights matter for the allocation of resources in the real world. Given that Coase's model is not robust to changing assumptions, and given that this failure of robustness seems to be an important part of Coase's reasoning, we believe that the derivational robustness framework fails to capture how it is Coase draws inferences from his model.

3.4 *Isolation.*

Another view – advanced primarily by Uskali Mäki – says that models are used in economic reasoning for their ability to isolate certain features of the world (in the same way that physical experiments do)[12] and, as a result, provide insight concerning the causal role these isolated features of the world play. We believe that Coase's model *does* isolate. But even so, Coase does not draw inferences from this isolation in the way that Mäki proposes. As such, though Mäki's

[12] On this direct comparison between models and experiments see Mäki (2005).

isolation view comes close to describing Coase's reasoning, it does not fully capture it. We explain in more detail below.[13]

> To get a better feel for Mäki's view, let us first get clearer on what exactly isolation entails:
>
> In an *isolation*, something, a set X of entities, is "sealed off" from the involvement of influence of everything else, a set Y entities; together X and Y comprise the universe. The isolation of X from Y typically involves a representation of the interrelationships among the elements of X. Let us call X the *isolated field* and Y the *excluded field*. It should be obvious that any representation involves isolation: isolation is ubiquitous in human cognition (Mäki 1992: 321).

After one has separated the isolated field X from the excluded field Y, one can then get a better grasp on the causal impact of the isolated field X on certain outcomes. This is made possible by "sealing off" all those features in the set Y, in the same way that experiments control for certain variables. By sealing off other variables, we can then know for certain that it is those features in the isolated set X driving the results.

The problem with using Mäki's framework to describe Coase's reasoning is that transaction costs are intentionally omitted from Coase's model, meaning that they are best understood as being part of the excluded field (the set Y). But according to Mäki, the purpose of a model is to draw inferences about the causal powers of those things included in the isolated field (the set X); we are allowed to do this, recall, precisely because they are sealed off from those things in the excluded field Y. But Coase's model gains insight on some feature of the world *f* not by including it in the set X, but rather by excluding it in the sealed-off set Y, and then seeing what sorts of outcomes the set X (importantly, *not* the set Y) produces. In other words, on Mäki's view, we gain insight on feature *f* by putting it in set X, sealing it off from set Y, and then seeing what X produces; on Coase's view, we gain insight on feature *f* by putting it in set Y, which is sealed

[13] Very similar to Mäki's account of models as isolations is Nancy Cartwright's (2009) account of models as revealing capacities. Indeed, Mäki's and Cartwright's views are often lumped together (e.g., Grüne-Yanoff 2009: 1-2; Knuuttila 2009: 60-61). As such, we believe that reasons we give here concerning why Mäki's framework does not capture Coase's reasoning can be applied, *mutatis mutandis*, to Cartwright's framework as well.

off from set X, and then seeing what X produces. For this reason, we do not think Mäki's account of models as isolations can make sense of the kind of reasoning Coase employs with his model.[14]

3.5 *Cased-Based Reasoning*.

In a relatively recent piece Itzhak Gilboa, Andrew Postlewaite, Larry Samuelson, and David Schmeidler advance a new way of understanding the value of economic modeling. To get a hold on what their view is we need to understand the distinction between *rule-based reasoning* and *cased-based reasoning*. Write the authors:

> In everyday as well as professional life, people use both rule-based reasoning and case-based reasoning for making predictions, classifications, diagnostics and for making ethical and legal judgments. Rule-based reasoning, in which the reasoner formulates general rules or theories, was formally introduced by the ancient Greeks, in the development of logic. In case-based (or, equivalently, analogical) reasoning, in contrast, the reasoner identifies similar past cases and uses those cases to guide the prediction (or classification, diagnosis, or ethical or legal judgment) in the current case (Gilboa *et al* 2015: F516-517).

As an example of rule-based reasoning, we might have a hard rule we employ every time we are in a situation relevant to the rule: for instance, always tip twenty percent. As an example of case-based reasoning, when we are in novel situations we can decide what to do by looking at similar situations confronted by us or others in the past.

The general thesis the authors advance is that (*i*) reasoning about the economic world is an instance of case-based reasoning rather than rule-based reasoning,[15] and (*ii*) the relevant cases we use in our reasoning are models. So, when we confront an interesting case in reality – what, for instance, will happen to unemployment if policy *p* is implemented? – we think about this in terms

[14] Indeed, Mäki has written on Coase (and transaction costs more generally) using his framework of models as isolations. Mäki's interpretation of Coase and his model is distinct from the one we offer which, we believe, is evidence that understanding models as isolations is inconsistent with the account of modeling we develop in the next section. See Mäki (1998); Mäki (2004).

[15] As an interesting aside, Mary Morgan's (2012: 1-2) work in the history of economic thought shows how economics shifted from a law-based science to a model-based science in the mid-twentieth century, mirroring the authors' claim that economic inquiry is essentially cased-based reasoning, *not* rule-based reasoning (though perhaps it *used* to consist of rule-based reasoning).

of cases which, really, are just models. We select the model that most closely resembles the situation at hand and then make our prediction.

On this view Coase's model is one out of many models we can appeal to in our cased-based reasoning: when we confront an economic situation that resembles closely enough Coase's model of a world absent transaction costs, then we use this model to form our predictions about what will happen. Now, this is certainly a way to use Coase's model. But it is clearly not how Coase reasoned with his model. The purpose was not to show what would happen should one end up in a situation in which transaction costs were zero or approached zero. Quite the contrary, Coase used his model to illuminate the role certain institutions play when transaction costs are *not* zero. Given this, we do not think Coase's model is best understood as an instance of case-based reasoning.

3.6 *How-Possibly*.

On another account of reasoning with economic models, some argue that models are used in economic reasoning by offering *how-possibly* explanations. In contrast to an actual explanation of some phenomenon – which requires "the identification of true (sufficient parts of) causes that brought about the explanandum" – how-possibly explanations "identify elements of possible causes for an explanandum" (Grüne-Yanoff 2013: 854). As an example, if model M containing features f_1, f_2, and f_3 were to act as an actual explanation of outcome O, then it must be the case that features f_1, f_2, and f_3 were the actual cause of outcome O in the real world. But, if we understand M as giving us a how-possibly explanation, then it need only be the case that features f_1, f_2, and f_3 could have been the cause of outcome O.

But how does a how-possibly explanation help us learn about the world? According to Till Grüne-Yanoff (2013: §4) there are at least five ways in which this can happen. First (*i*) how-possibly explanations can show us how an explanation of an outcome we thought was impossible is not, in fact, impossible; second (*ii*) if there are multiple possible mechanisms hypothesized to cause an outcome, how-possibly explanations can change our distribution of credences across which of these mechanisms is the correct explanation; third (*iii*) how-possibly explanations can show that certain kinds of causal mechanisms *simply cannot* produce certain kinds of outcomes; fourth (*iv*) how-possibly explanations can show that systems with certain properties are capable of

producing outcomes with certain properties; and fifth (*v*) how-possibly explanations can offer explanations for an outcome that was either unexplained or explained differently.

We do not think the how-possibly framework can capture Coase's reasoning. The reason why is that, according to the how-possibly framework, there needs to be a symmetry between (1) the outcome *O* of a model (containing features f_1, f_2, and f_3) and (2) what we see and experience in the real world. Then, in can reason with the model in the ways highlighted in the paragraph above: perhaps we now think that it is in principle possible for f_1, f_2, and f_3 to cause *O*. Or maybe we now think it is more likely that f_1, f_2, and f_3 cause *O* compared to an alternative hypothesis in the literature – that f_4, f_5, and f_6. Importantly, though, this symmetry between the outcome of the model and real world is absent in Coase's case. Here, the model yields a result that is completely foreign to what we actually experience in the real world. In the model world, property rights are irrelevant, and all disputes are resolved in the same efficient manner; in the real world property rights matter greatly in how disputes are resolved, and many resolutions fail to be efficient. As such, Coase's model cannot give a how-possibly explanation of why property rights influence the resolution of externalities, because its outcome shows the exact opposite to be the case. In other words, the current framework says we can give a how-possibly explanation of real-world outcome *O* by building a model showing how f_1, f_2, and f_3 cause *O*. Coase, however, wants to explain the relationship between some feature *f* and outcome *O* by building a model in which *f* is absent and some other outcome, *O`*, obtains. This is, we do not think, qualifies as a how-possibly explanation.

3.7 *Caricatures*.

Above we noted that Allan Gibbard and Hal R. Varian offer one way of reasoning with models: models can act as approximations, telling us about the world by offering a reasonably accurate depiction of it. Besides acting as approximations, Gibbard and Varian also argue that models can be used in economic inquiry as *caricatures*: "often the assumptions of a model are chosen not to approximate reality, but to exaggerate or isolate some features of reality… a model that is a better approximation to reality may make for a worse explanation of the role of some particular feature of reality" (Gibbard and Varian 1978: 673). Clearly, this describes Coase's model: it is a gross caricature of reality in that it removes a pervasive and important aspect of

economic life. But even so, does Coase's inferential strategy match the inferential strategy proposed by Gibbard and Varian? In fact, what *is* the inferential strategy Gibbard and Varian propose to accompany a model that caricatures reality?

Gibbard and Varian are not particularly clear on this. One example of how models as caricatures are used in reasoning is by proving a robustness analysis. According to the authors, we can be more confident in a model's conclusions when "(1) the assumptions of the model caricature features of the situation, and (2) the conclusions are robust under changes in the caricature" (Gibbard and Varian 1978: 675). This, of course, is similar to our analysis of derivational robustness reasoning in §3.3 above. There we saw that the robustness framework does not capture Coase's reasoning. Beyond this, however, Gibbard and Varian do not specify the way in which one draws inferences from a caricatured model. So though Coase's model *is* clearly a caricature in the sense Gibbard and Varian describe, they do not tell us enough about how such models are used in economic reasoning. We remedy this lacuna in the next section.

4. Models as Foils

Intuitively, it seems as though Coase's informal model of a world without transactions costs teaches us something deeply important about the world. But our best philosophical accounts of economic modeling cannot accommodate this intuition. We now remedy this failing by sketching a new account of reasoning with economic models that we call *models as foils*. As we said in the introduction, we do not claim that this is the only way models gain value as tools of economic inquiry. Indeed, many models are valuable precisely because they produce testable hypotheses, are credible worlds, induce isolations, and so on. We merely wish to lay out one way in which models can be used in economic inquiry that, we think, has yet to be carefully articulated.

To begin it is helpful to go back in time. In the introduction of this paper we noted that, though contemporary accounts of economic modeling cannot account for Coase's inferential strategy, his method of reasoning is similar to John Stuart Mill's *method of difference*, from his *A System of Logic, Ratiocinative and Inductive*. Though we do think there are significant similarities here, Mill's method of difference does not fully capture Coase's reasoning. But even so, examining the way in which Mill's method of difference *almost* captures Coase's reasoning will be instructive

for articulating our account of models as foils. So, we begin by examining how Mill's method of difference comes close but ultimately fails to capture Coase's reasoning. Then, learning from these failures, we articulate our new account of reasoning with models.

Mill describes the method of difference as follows:

> If an instance in which the phenomenon under investigation occurs, and an instance in which it does not occur, have every circumstance in common scarce one, that one occurring only in the former; the circumstance in which alone the two instances differ, is the effect, or the cause, or an indispensable part of the cause, of the phenomenon (Mill 1843: 452).

The above is a bit dense but Mill (1843: 451) offers an example that helps clarify this inferential strategy. Suppose, in Case 1, what Mill calls "elements" A B C produce outcome *a b c*. Then, suppose in Case 2 that elements B C produce outcome *b c*. Clearly, Case 1 and Case 2 have different outcomes; in Case 1 *a b c* occurs, but in Case 2 *b c* occurs. Since the only difference between Case 1 and Case 2 is the presence of A, there is reason to think that element A causes outcome *a*. This is the method of difference.

Trying to offer a bit more clarity, we believe the following sequence of reasoning is an accurate characterization of Mill's method of difference.

1. In Case 1, features $f_1, f_2, \ldots, f_n$ produce outcome *O*.
2. In Case 2, features $f_1, f_2, \ldots, f_{n-1}$ produce outcome *O`*, where *o\** represent those features or properties outcomes *O* and *O`* differ across.
3. Since the only difference between Case 1 and Case 2 is feature f_n, there is reason to think that f_n causes *o**.

Stated like this, we believe that Mill's method of difference is clear enough, and a legitimate way of drawing inductive inferences. The question is whether this strategy of inference mirrors Coase's reasoning with his model of a world absent transaction costs.

Initially one might think that this does capture Coase's reasoning. For consider:

1. In Case 1 (the real world), different allocations of property rights result in externalities being resolved in different ways (some or all of which may be inefficient).

2. In Case 2 (the real world *minus* transaction costs), all externalities are bargained to the same efficient point regardless the initial allocation of property rights.
3. Since the only difference between Case 1 and Case 2 is the presence of transaction costs, there is reason to think that transaction costs cause different allocations of property rights to result in externalities being resolved in different ways (some or all of which may be inefficient).

While this application of Mill's method of difference reaches the conclusion that Coase does from his model – unlike other accounts of economic inference examined in §3 above – there are nonetheless problems with it.

In the above schematization, step (2) says that Case 2 consists of all those features of the real world, minus transaction costs. The outcome of this case is that all externalities are bargained to the same efficient point regardless of the initial allocation of property rights. Case 2 in the above schematization is supposed to represent Coase's model world. It is true, we know, that the outcome of this model world is indeed the irrelevance of property rights for the resolution of externalities. The problem, though, is that Coase's model world does not consist of *all* those features of the real world *except* for transaction costs. In fact, there are many features of the real world that Coase's model leaves out. Indeed, this should not be surprising – almost by definition, models abstract away from almost all features of reality in their effort to simplify.

Coase's model, for example, leaves out the fact that the sky is blue. More relevant than this, the model leaves out the fact that persons often lack perfect information, that they often have asymmetric information, that they often behave irrationally, and so on and so forth. As such, it is misleading to say that Case 2 consists of all those features of the real world *except* transaction costs. Coase's model abstracts away from many features of the real world, transaction costs being but one. This fact, though, means that Mill's method of difference is inapplicable to Coase's reasoning. According to Mill's method of difference, we can only draw a causal inference between two cases so long as these two cases "have every circumstance in common scarce one" (Mill 1843: 452). Problematically, Coase's model and the real world do not have every circumstance in common scarce one. There are many features across which they differ, transaction costs being one of many. As a result, Mill's method of difference does not capture Coase's reasoning.

Another way of thinking about the problem here is that Coase wants to use something like Mill's method of difference, but the two cases he wants to use are (*i*) what we observe in the real world, and (*ii*) what we observe in a model world. Problematically, though, a model world will never be able to include all those features present in the real world scarce one. Because of this, one might conclude that not only does Mill's method of difference fail to capture Coase's reasoning, but that it will fail to capture any reasoning that employs a comparison between two distinct cases, where one of the cases consists of real-world observation and the other a model world. Our account of models of foils not only captures Coase's reasoning but allows for these sorts of comparisons more generally.

The method of difference does not capture Coase's reasoning, but it does seem quite close. In particular, both Coase and the method of difference (*i*) examine two cases, (*ii*) these two cases have different outcomes, and (*iii*) these different outcomes are pinned on a particular feature that causes the relevant difference. For Mill, it is easy to pin down the relevant feature. By hypothesis, there is only one difference between the two cases, so the differentiating feature is the causal one. Coase still wants to pin down the difference between the two cases on the presence or absence of a particular feature. But he cannot do this so easily as Mill wishes to with the method of difference, because there are many features that the real world and model world differ across, not just one.

This does not mean, however, that it is impossible to show that one feature does cause the relevant difference. Consider a toy example. Suppose features $\{f_1, f_2, f_3, f_4\}$ cause some outcome O, and features $\{f_1, f_2\}$ cause $O`$, where $o*$ represent those features or properties outcomes $O$ and $O`$ differ across. Here, Mill's method of difference is inapplicable because the two cases differ across multiple features – f_3 and f_4. As such, it is not immediately obvious whether f_3 is the primary cause of $o*$, or f_4. But just because this is not immediately obvious it does not follow that this cannot be established. We know, at least, that it is either f_3 or f_4 that is to account for $o*$. And here, it is sometimes (but not always) possible that some kind of argument can be given for why it is f_3 over f_4 (or vice versa) that explains $o*$. Perhaps there is a theoretically plausible mechanism showing how f_3 induces $o*$, yet there is no such mechanism for f_4. Or perhaps in reasonably similar circumstances, it is f_3 that produces the outcome, not f_4. Should such an argument be established, then we do have *some* reason to insist that f_3 is the primary cause of $o*$, rather than f_4, *even if* the

justification of our belief here is not as strong as would be the case were the only difference between the two cases to be f_3 (in which we would be allowed to use Mill's method of difference).

Coase does something like this. He does not simply present two cases – the real world where property rights matter in the resolution of externalities and the model world where they do not – and then conclude that transaction costs are the relevant causal feature. Rather, he carefully works through examples in the model world where property rights are allocated differently and argues that it is precisely because transaction costs are absent that these different allocations of rights simply do not matter. He does this by examining cases of conflict and showing that there exist efficient bargains the parties *could* reach, *if* they had limitless time to acquire information, engage in such bargains, and so on. And it just so happens that these efficient bargains result in the same distribution of resources regardless the initial allocation of property rights. Coase shows this at length by examining disputes between the cattle farmer and crop raiser, the confectioner and the doctor, and many, many more (Coase 1960/1988: §§3-5). We believe that carefully working through these cases and pointing out the existence of such efficient bargains constitutes an argument for why it is the absent transaction costs in the model world that is driving the difference in outcomes between the real world and the model world. So Coase uses something like Mill's method of difference, only (*i*) there is more than one difference between the two cases (because one of the cases consists of real-world observation and the other a model world) and, because of this, (*ii*) some kind of argument needs to be given concerning why one differentiating feature – rather than all the other differentiating features – is responsible for the different outcomes. A bit more precisely:

1. In Case 1 (the real world), the set of features *F* produces outcome *O*.
2. In Case 2 (the model world), the set of features *F*`, which is a proper subset on *F*, produces outcome *O*`, where *o** represent those features or properties outcomes *O* and *O*` differ across.
3. Letting *F** be the complement of *F*` (indexed to the universe *F*), and letting $f_n$ be an element of *F*\*, there exists an argument for why $f_n$, compared to all the other elements in *F*\*, best explains the divergence between *O* and *O*`.
4. As such, there is some reason to believe that f_n causes *o**.

Call this inferential strategy *models as foils*, for a model world is being used as a foil against reality to gain insight on the causal impact of some feature of the real world. In particular, we gain insight on feature *f* of the real world by comparing the real world (step one) to the world of a model where *f* is not present (step two), showing that *f* is most plausibly responsible for the divergence between the real world and the world of the *f*-less model (step three).

Some things to note about models as foils. Models as foils is a less reliable inferential strategy than Mill's method of difference. With the method of difference, there is only one culprit to pin the relevant explanation on, for there is by definition only one difference between the two cases. We are thus reasonably sure that this single differentiating feature is causally efficacious. With models as foils, though, because the two cases being compared are the real world and a model world, there will be more than one culprit capable of explaining the relevant difference, for there will be multiple differentiating features present. Here, the theorist needs to give some argument concerning why feature f_n (a member of set F^*) rather than f_m (also a member of set F^*) best explains the divergence between the two outcomes. Of course, the theorist's argument cannot show this definitely – even though she thinks that there is compelling reason to think it is f_n that causes the difference, in reality it could be f_m that is doing all the work. Because of this, we think that a theorist who uses models as foils as an inferential strategy is less epistemically justified in her conclusion than someone who uses Mill's method of difference. But being less justified does not imply that one lacks any sort of justification at all. We do think that models as foils as a method of inference can establish credible beliefs about the world.

The whole purpose of this paper is to give an account of the way in which Coase reasons with models. We believe our account of models as foils does this, and we now wish to end the paper by highlighting this. Consider the following reasoning:

1. In Case 1 (the real world), different allocations of property rights result in externalities being resolved in different ways (some or all of which may be inefficient).
2. In Case 2 (the model world), all externalities are bargained to the same efficient point regardless the initial allocation of property rights.
3. Compared to all of the elements present in the real world but absent in the model world, there exists an argument for why transaction costs best explain the divergence in outcomes.

4. As such, there is reason to think that transaction costs cause different allocations of property rights to result in externalities being resolved in different ways (some or all of which may be inefficient).

The conclusion stated in (4) is indeed the conclusion that Coase reached with his model. And, moreover, we belief the structure of reasoning in the schematization is similar to Coase's, at least as we articulated it in §2 above. To put it more informally and less cumbersomely: because property rights wouldn't matter in a world without transaction costs – precisely because these costs are absent – but because we live in a world where both transaction costs are present and property rights *do* matter, we have some reason to believe that transaction costs are the reason why these property rights do in fact matter. We have reason to believe this through using the model world as a foil against our real one.

Works Cited

Alexandrova, Anna. 2008. "Making Models Count." *Philosophy of Science* 75: 383-404.

Bernstein, Sara. 2015. "The Metaphysics of Omissions." *Philosophy Compass* 10: 208-218.

Bertrand, Elodie. 2010. "The Three Roles of the 'Coase Theorem' in Coase's Works." *European Journal of History of Economic Thought* 17: 975-1000.

Boettke, Peter J. 1997. "Where Did Economics Go Wrong? Modern Economics as a Flight from Reality." *Critical Review* 11: 11-64.

Cartwright, Nancy. 2009. "If No Capacities then No Credible Worlds. But can Models Reveal Capacities?" *Erkenntnis* 70: 45-58.

Clarke, Kevin A. and David M. Primo. 2012. *A Model Discipline: Political Science and the Logic of Representation*. Oxford: Oxford University Press.

Coase, Ronald H. 1960/1988. "The Problem of Social Cost," in *The Firm, the Market, and the Law*: 95-156. Chicago: University of Chicago Press.

Coase, Ronald H. 1988a. “Notes on the Problem of Social Cost,” in *The Firm, the Market, and the Law*: 157-185. Chicago: University of Chicago Press.

Coase, Ronald H. 1988b. “The Firm, the Market, and the Law,” in *The Firm, the Market, and the Law*: 1-32. Chicago: University of Chicago Press.

Friedman, Milton. 1953. *Essays in Positive Economics*. Chicago: University of Chicago Press.

Gibbard, Allan and Hal R. Varian. 1978. “Economic Models.” *Journal of Philosophy* 75: 664-677.

Gilboa, Itzhak, Andrew Postlewaite, Larry Samuelson, David Schmeidler. 2014. “Economic Models as Analogies.” *The Economic Journal* 124: F513-F533.

Grüne-Yanoff, Till. 2009. “Preface to ‘Economic Models as Credible Worlds or Isolating Tools?’” *Erkenntnis* 70: 1-2.

Grüne-Yanoff, Till. 2013. “Appraising Models Nonrepresentationally.” *Philosophy of Science* 80: 850-861.

Hayek, F.A. 1946/2014. “The Meaning of Competition.” In *The Collected Works of F.A. Hayek Volume Fifteen*, edited by Bruce Caldwell: 105-116. Chicago: University of Chicago Press.

Hayek, F.A. 1961/2014. “A New Look at Economic Theory.” In *The Collected Works of F.A. Hayek Volume Fifteen*, edited by Bruce Caldwell: 374-426. Chicago: University of Chicago Press.

Hoffman, Elizabeth and Matthew L. Spitzer. 1982. “The Coase Theorem: Some Experimental Tests.” *The Journal of Law & Economics* 25:73-98.

Hsiung, Bingyuan. 2004. “An Interpretation of Ronald Coase’s Analytical Approach.” *History of Economics Review* 39: 12-32.

Hsiung, Bingyuan and J. Patrick Gunning. 2002. “Ronald Coase’s Method of Building More Realistic Models of Choice.” *Review of Political Economy* 14: 227-239.

Kirzner, Israel M. 1973/2013. *Competition and Entrepreneurship*. Indianapolis: Liberty Fund.

Knuuttila, Tarja. 2009. “Isolating Representations versus Credible Constructions? Economic Modeling in Theory and Practice.” *Erkenntnis* 70: 59-80.

Kuorikoski, Jaakko and Aki Lehtinen. 2009. “Incredible Worlds, Credible Results.” *Erkenntnis* 70: 119-131.

Kuorikoski, Jaako, Aki Lehtinen, and Caterina Marchionni. 2010. “Economic Modelling as Robustness Analysis.” *British Journal of the Philosophy of Science* 61: 541-567.

Libecap, Gary D. and Dean Lueck. 2011. “The Demarcation of Land and the Role of Coordinating Property Institutions.” *Journal of Political Economy* 119: 426-467.

Mäki, Uskali. 1992. “On the Method of Isolation in Economics.” *Ponzan Studies in the Philosophy of the Sciences and the Humanities* 26: 19-54.

Mäki, Uskali. 1998. “Is Coase a Realist?” *Philosophy of the Social Sciences* 28: 5-31.

Mäki, Uskali. 2004. “Theoretical Isolation and Explanatory Progress: Transaction Cost Economics and the Dynamics of Dispute.” *Cambridge Journal of Economics* 28: 319-346.

Mäki, Uskali. 2005. “Models are Experiments, Experiments are Models.” *Journal of Economic Methodology* 12: 303-315.

Mäki, Uskali (ed). 2009. *The Methodology of Positive Economics: Reflections on the Milton Friedman Legacy*. Cambridge: Cambridge University Press.

Medema, Steven G. 2011. “A Case of Mistaken Identity: George Stigler, ‘The Problem of Social Cost,’ and the Coase Theorem.” *European Journal of Law and Economics* 31: 11-38.

Mill, John Stuart. 1843. *A System of Logic, Ratiocinative and Inductive: Being a Connected View of the Principles of Evidence and the Methods of Scientific Investigation.* eBook reproduced by Gale, Cengage Learning.

Mireles-Flores, Luis. 2018. “Recent Trends in Economic Methodology: A Literature Review.” *Research in the History of Economic Thought and Methodology* 36A: 93-126.

Mises, Ludwig von. 1949/2007. *Human Action: a Treatise on Economics*. Indianapolis: Liberty Fund.

Morgan, Mary S. 2012. *The World in the Model*. Cambridge: Cambridge University Press.

Stigler, George. 1966. *The Theory of Price*. 3rd Edition. New York: Macmillan.

Sugden, Robert. 2000. "Credible Worlds." *Journal of Economic Methodology* 7: 1-31.

Wallis, John J. and Douglass North. 1986. "Measuring the Transaction Sector in the American Economy, 1870-1970." In *Long-Term Factors in American Economic Growth*, edited by Stanley L. Engerman and Robert E. Gallman: 95-162. Chicago: University of Chicago Press.

Ylikoski, Petri and N. Emrah Aydinonat. 2014. "Understanding with Theoretical Models." *Journal of Economics Methodology* 21: 19-36.