Précis of
Modality and Explanatory Reasoning

The aim of *Modality and Explanatory Reasoning* (*MER*) is to shed light on metaphysical necessity and the broader class of modal properties to which it belongs. This topic is approached with two goals: to develop a new and reductive analysis of modality, and to understand the purpose and origin of modal thought. I argue that a proper understanding of modality requires us to reconceptualize its relationship to causation and other forms of explanation such as grounding, a relation that connects metaphysically fundamental facts to non-fundamental ones. While many philosophers have tried to give modal analyses of causation and explanation, often in counterfactual terms, I argue that we obtain a more plausible, explanatorily powerful and unified theory if we regard explanation as more fundamental than modality. The function of modal thought is to facilitate a common type of thought experiment—counterfactual reasoning—that allows us to investigate explanatory connections and which is closely related to the controlled experiments of empirical science. Necessity is defined in terms of explanation, and modal facts often reflect underlying facts about explanatory relationships. The study of modal facts is important for philosophy not because these facts are of much metaphysical interest in their own right, but largely because they provide evidence about explanatory connections.

1. The nature of ontic modality

My discussion starts from the plausible idea that a proposition is necessary if its truth is in some sense very secure, invariable, or unconditional. I argue in chapter 2 that this notion of secure truth is the same one we use when we ask of a true proposition *how easily it could have been false*. The less easily it could have been false, the more secure its truth. How easily something could have been the case depends on how great a departure from the way things actually are is required for it to be the case. Suppose your team would have won the game if the goalkeeper had stood half an inch further to the left during the last minute. Then it's true to say that they could easily have won. The same

claim is false if they would have won only if they had done countless things differently during the last ten years. Similarly, how easily a true proposition *P* could have been false depends on how great a departure from actuality is required for *P* to be false. The greater the departure required, the more secure *P*'s truth.

It is often assumed that necessity and possibility are all-or-nothing matters. But how easily a proposition could have been false is clearly a matter of degree, and I argue on that basis that we should think of necessity and possibility themselves as coming in degrees. To say that *P* could more easily have been true than *Q* is to say that *P* has a higher degree of possibility than *Q*. And to say that *Q* could more easily have been false than *P* is to say that *P* has a greater degree of necessity than *Q*.

While talk about degrees of possibility is ubiquitous in ordinary life, the idioms we use are often not overtly modal but instead use the metaphors of distance, security or fragility. We say that Fred *nearly* missed the train, or got *within a hair's breadth* of disaster, to communicate that the realization of a certain situation requires only a minimal departure from actuality. The peace between two nations during some past period can be called *fragile* or *secure*, depending on how easily their tensions could have escalated into war. A sentence like "Smith came *closer* to winning than Jones did" compares two unrealized scenarios by their proximity to actuality. Such comparisons also underlie counterfactual judgments, since a counterfactual $\ulcorner A \mathrel{\Box\!\!\rightarrow} C \urcorner$ is true iff some scenarios where $\ulcorner A \,\&\, C \urcorner$ is true depart less from actuality than any scenario where $\ulcorner A \,\&\, {\sim}C \urcorner$ is true (Stalnaker 1968, Lewis 1973a).

An analysis of necessity in terms of an ordering of non-actual scenarios by their closeness to actuality may seem circular at first blush, since the very property of being a non-actual situation is often thought to be modal. Many philosophers take "non-actual situations" to be just another expression for *unrealized ways things could have been* (unactualized metaphysically possible situations). However, I think it is a mistake to identify the space of unactualized scenarios with the class of unrealized metaphysically possible scenarios. To mention just one of the difficulties for this view, on the assumption that all scenarios are metaphysically possible, the account of counterfactuals described in the previous paragraph entails that all counterfactuals with metaphysically impossible antecedents have the same truth-value. That seems wrong. It is metaphysically impossible

for Hillary Clinton to be Antonin Scalia's daughter. And yet, it seems plausible that the counterfactuals "Clinton would be a conservative if she were Scalia's daughter" and "Clinton would be a liberal if she were Scalia's daughter" have different truth-values. The obvious remedy, proposed by a number of philosophers, is to formulate the account of counterfactuals not in terms of possible worlds, but in terms of worlds more generally, including both possible and impossible worlds (see, e.g., Nolan 1997). Worlds are maximally specific ways for reality to be, and they include both ways reality could be and ways reality could not be. As discussed below, I take worlds to be definable in non-modal terms.

We can appeal to worlds to sharpen the account of modality sketched above. How easily *P* could have been true (*P*'s degree of possibility) is determined by how close the closest *P*-worlds are to actuality. The class of worlds within a certain distance from actuality may be called a "sphere" around the actual world. The ordering of unactualized worlds by their closeness to actuality generates a system of nested spheres. For each sphere there is a grade of necessity that attaches to just those propositions that are true at *every* world in that sphere, as well as a grade of possibility attaching to all propositions that are true at *some* world in the sphere. The larger the sphere, the greater the associated grade of necessity. One sphere, described in more detail below, corresponds to metaphysical necessity: for a proposition to be metaphysically necessary is for it to hold at every world in that sphere. Another, smaller sphere corresponds to nomic necessity. There are many other spheres as well, which give us yet further grades of necessity, some lower than nomic necessity, some between nomic and metaphysical necessity, and some greater than metaphysical necessity. I argue in chapters 2 and 3 that this theory does a good job of capturing our core beliefs about what necessity is, and that it illuminates and explains various well-known features of modality and modal discourse.

Kripke famously distinguished between epistemic necessity (a prioricity) and metaphysical necessity. Unlike some philosophers (e.g., Frank Jackson, David Chalmers, Robert Stalnaker), I take his findings to reflect a fundamental distinction between two forms of modality that arises at the levels of worlds and propositions as well as sentences. My account traces the distinction back to the difference between two cognitive practices of thinking about alternative situations that serve different purposes, and which employ

quite different sets of criteria for classifying and comparing scenarios. In one practice we employ epistemic criteria (such as whether a situation can be ruled out a priori), while in the other we compare scenarios by their mutual overall similarity (closeness) in non-epistemic respects (e.g., in their histories and the laws that govern them). The first practice gives rise to epistemic modal notions, the second is connected to a form of graded modality that is metaphysical rather than epistemic, and which I call 'ontic modality.' Metaphysical necessity, nomic necessity, and counterfactual dependence are examples of ontic modal properties and relations. Ontic modality is the subject matter of *MER*.

While we employ different standards of closeness in different contexts, I follow David Lewis (1979) in thinking that there is a specific standard that we use as our default (absent special features of the context). The ontic modal properties and relations discussed in *MER*, including metaphysical and nomic necessity, are defined in terms of the closeness relation determined by these standards (which I call the "standard closeness relation"). In chapters 8–9 I give a non-modal analysis of these standards in terms of explanation. The next section contains an outline of this account, prefaced with a brief summary of my background assumptions about explanations.

To complete my analysis of modality, I offer a non-modal account of worlds (chapters 4–5). In essence, worlds are classes of Russellian propositions that describe reality in maximally detailed and logically consistent ways. (The notion of logical consistency can be understood non-modally in terms of the logical structure of Russellian propositions.) Chapter 4 provides novel motivation for the view (also endorsed by a number of other philosophers) that many worlds exist contingently, and it gives a new account of the existence conditions of worlds. I argue that worlds are even more modally fragile than previously thought. For example, there are worlds, some very close to actuality, that fail to exist at themselves—if they had been actualized, then they would not have existed. In addition, which propositions are true at a given world *w* can vary between different possible worlds where *w* exists, and the very property of being a world is a contingent feature of many worlds (some worlds could have been non-maximal situations rather than worlds). These results have noteworthy implications for our understanding of iterated modality and of what it is for a world to be actualized. They also impose significant

constraints on a definition of worlds. In chapter 4 I offer a definition intended to meet these constraints. For an account like mine that views worlds as classes of propositions, it is a difficult challenge to afford a sufficiently rich space of worlds without falling prey to paradox. Chapter 5 proposes a technical solution to this problem.

2. Necessity and Explanation

In *MER* I use "explanation" for a metaphysical relation, not an epistemic one: to say that *x* explains *y* is to say that *x* is the reason why *y* obtains, or that *y* is due to *x*. Causal relationships are one paradigmatic example of explanation, but there are other forms of explanation as well. For example, effects are typically explained not by their causes alone, but by these together with certain facts about the laws of nature. More generally, I endorse an anti-Humean "governing" view of laws according to which (to simplify somewhat) the fact that it is a law of nature that *P* explains the fact that *P* and also explains the individual instances of the proposition that *P*. For instance, the fact that it is a law that any two bodies attract each other explains the general fact that any two bodies attract each other, and it also explains the fact that the specific bodies *b* and *b** attract each other. This is an example of non-causal explanation: the fact that a certain principle is a law partly explains certain goings-on but it doesn't cause them. Yet another form of explanation is the relation of *grounding* (discussed in chapter 6). Grounding is importantly analogous to causation. Under determinism earlier states of the universe causally generate later ones in accordance with the natural laws. Similarly, metaphysically more fundamental facts ground less fundamental ones in accordance with certain principles I call "metaphysical laws." The metaphysical laws include the essential truths, which state conditions for being a certain entity or for instantiating a certain property or relation. The metaphysical laws play an explanatory role similar to that of the natural laws. To take an example, the following may be an essential truth about the property of being a gold atom:

(1) All and only atoms with atomic number 79 are gold atoms.

The fact that (1) is essential to gold atomhood explains the fact that all and only atoms with atomic number 79 are gold atoms. Moreover, the fact that *a* is a gold atom is

explained by its ground—the fact that *a* is an atom with atomic number 79—together with the fact that (1) is essential to gold atomhood. This is another instance of the governing conception of laws, but applied to a metaphysical law rather than a law of nature.

Attempts to analyze causation and explanation in counterfactual terms are motivated by the observation that judgments about causal and explanatory relationships are often guided by counterfactual beliefs. However, it is equally true that counterfactual beliefs are often informed by preexisting judgments about explanatory relationships, a pattern that I illustrate in chapters 8–9 by describing numerous examples (some taken from the literature and some new). Just as one can use the first phenomenon to motivate an analysis of explanation in counterfactual terms, one could use the second phenomenon to support an analysis of closeness and counterfactuals, and of ontic modality more generally, in terms of explanation. To decide between the two directions of analysis, we need to determine in more detail which approach can better account for the complex relationship between counterfactuals and explanation. Counterfactual analyses of causation and explanation face well-known problems in this area (which are briefly reviewed in chapter 10). Chapters 8–12 aim to show that an account of ontic modality in terms of explanation can give a better account of the data.

Such a theory is developed in chapters 8–9, where I survey numerous data (both old and new) and propose an account of the standards of closeness that explains them. Roughly speaking, the comparative closeness to actuality of two worlds is determined by weighing their various similarities to actuality against each other. Not all similarities carry weight—some count for nothing. The *first* part of my theory is a rule, called the "explanatory criterion of relevance" (ECR), that singles out the relevant similarities. Roughly speaking, if a fact *f* obtains both at the actual world and at another world *w*, then this similarity between the two worlds is relevant to the closeness ordering iff all factors that contribute to explaining *f* at the actual world obtain at *w*. The *second* part of the account specifies the relative weights of different kinds of relevant similarities. The weightiest such similarities concern the metaphysical laws. To simplify somewhat, worlds that have the same metaphysical laws as actuality and perfectly conform to these laws are closer to actuality than worlds that don't meet these conditions. The former

worlds therefore form a sphere, *S*, around actuality. Metaphysical necessity is the grade of necessity corresponding to that sphere (chapter 7). The second weightiest kind of similarity is match in the natural laws. Of the worlds in sphere *S*, those with the same natural laws as actuality form a second, smaller sphere within *S*. Nomic necessity is the grade of necessity corresponding to this second sphere. Similarities between the histories of two worlds matter to the closeness ordering as well, although to a lesser extent.

Since ECR is the centerpiece of the proposed analysis of ontic modality in terms of explanation, I spend a significant portion of chapters 8–12 on elaborating and supporting the principle. My main argument for ECR is that it provides a simple and unified explanation of a wide variety of data. For example, ECR plays a key role in explaining the differences in degree of necessity (described in the previous paragraph) between facts about the metaphysical laws, facts about the natural laws, and facts about the course of history. (See Lange's comments and my reply.) Moreover, when combined with independently plausible assumptions about explanation, ECR can account for the difference in counterfactual-supporting power between the natural laws and accidental regularities (see my response to Lange), while also explaining the datum (*MER*: 8, 209–210) that laws support some counterfactuals but not others (see my reply to Sullivan). ECR together with the temporal asymmetry of causation can explain the temporal asymmetry of counterfactual dependence, i.e. the fact that the future counterfactually depends on the past to a much greater extent than the past depends on the future (chapter 8; for more detail, see Kment 2006a: sct. 5). ECR also explains the frequently observed fact (Edgington 2003) that in counterfactual reasoning we tend to hold fixed the outcome of a post-antecedent chance process only if this outcome is causally independent of the antecedent. (See the lottery example in section 1 of Sullivan's comments.) There are further data that can be explained by ECR as well, some of which are discussed in Sullivan's comments and my reply.

3. The function and origin of modal thought

In Chapters 10–12 I argue that modal thinking developed, at least in part, because of the utility of counterfactual reasoning in evaluating claims about explanatory relationships. This procedure is an extension of John Stuart Mill's method of difference, which is

central both to ordinary causal thought and to the experimental methodology of empirical science. Idealizing considerably and focusing on deterministic contexts, we can describe the method of difference as follows. The agent observes Scenario 1, in which *A*–*D* are present and are followed by *E*, and Scenario 2 (the "control condition"), in which *B*–*D* are present but *A* isn't and in which *E* doesn't obtain at the next moment. If she believes that *A*–*D* include all factors in Scenario 1 that are causally relevant to the presence of *E*, then she can infer that *A* is a cause of *E* in Scenario 1.

On the reconstruction I offer in chapter 10, the method of difference relies on a thesis I call the "determination idea." When applied to causation under determinism, this is the thesis that the causes of *E* and the laws involved in *E*'s explanation together determine *E*, where determination is understood non-modally in terms of logical entailment between Russellian propositions. (Other versions of the determination idea apply to probabilistic causation (*MER*: 326) and to grounding (*MER*: 168).) The determination idea is not an analysis of causation. It merely states a condition that is necessary, though not sufficient, under determinism for certain factors to include all of *E*'s causes: these factors and the laws involved in *E*'s explanation must together determine *E*. The determination idea provides a straightforward explanation of how the method of difference works. Since *B*–*D* obtain in Scenario 2 but *E* doesn't, the agent can conclude that *B*–*D* and the laws don't determine *E*. But by the determination idea, the factors that caused *E* in Scenario 1 and the laws must together determine *E*. So, *B*–*D* can't include all of the causes of *E* in Scenario 1. Given the assumption that *A*–*D* do include all of these causes, it follows that *A* must be a cause of *E* in Scenario 1.

The method of difference is limited in scope. If we have observed *A* followed by *E*, and we want to show that *A* was a cause of *E*, we have to find or create another situation where *A* doesn't obtain but which otherwise matches the scenario we have observed in all causally relevant ways. That is often impossible in practice. And the method is useless when our goal is to find out not what caused *E*, but which laws were involved in *E*'s explanation. For the laws never vary between different scenarios that actually obtain. If my reconstruction of Mill's method is on the right track, however, then there is a straightforward extension of it that remedies these shortcomings. On my account, the sole function of Scenario 2 is to show that *B*–*D* and the laws don't determine *E*. But given a

realistic amount of background knowledge about the laws, we can show the same by mental simulation (section 10.6.2). We represent to ourselves an unactualized scenario where *B*–*D* obtain but *A* doesn't, and where history then unfolds in accordance with the actual laws. If *E* fails to obtain in this situation, then *B*–*D* and the laws don't determine *E*. Using the determination idea, we can again infer that in the actual scenario *B*–*D* don't include all causes of *E*. Given our background assumption that *A*–*D* do include all of *E*'s actual causes, we can conclude that *A* is actually a cause of *E*. The mental simulation I described is a simplified version of the reasoning by which we determine whether *E* depends counterfactually on *A*: we imagine a scenario where *A* is absent, holding fixed various other facts that actually obtain (*B*–*D* and the laws), and we then determine whether *E* obtains in that situation. The situation imagined serves the same purpose as Scenario 2 (the "control condition") in the method of difference, and by holding fixed the right facts we achieve the same as by controlling for background conditions in an experiment. The same type of mental simulation can also be used to show that a certain law *L* is involved in explaining *E*, only in that case we need to imagine a scenario where *L* isn't a law but where other relevant factors are the same as they actually are.

Chapter 11 explains why a sophisticated version of counterfactual reasoning requires a closeness ordering of unrealized scenarios that is governed by the specific standards described in chapters 8–9. Roughly speaking, this ordering gives us an easy way of deciding, for any fact *A*, which unrealized scenarios we need to consider if we want to test whether *A* partly explains a certain other fact: of all scenarios where *A* is absent, we should consider those that are closest to actuality in the ordering. The background facts that we need to hold fixed are just those that obtain in these closest scenarios. As mentioned before, our standards of closeness accord great weight to similarities in the natural laws, and even greater importance to match with respect to the laws of metaphysics. Whenever possible, we should hold fixed which metaphysical laws are in force, and if possible, we should also hold fixed what the natural laws are. The rationale for these rules is closely connected to the distinctive explanatory roles of the metaphysical and natural laws that are described in chapter 6. The purpose of our various modal notions, including those of metaphysical and nomic necessity, consists in the fact that they make it easier to apply these rules of counterfactual reasoning. On this account,

ontic modal thinking, and the concepts of metaphysical and nomic necessity, are integral parts of a familiar framework for studying explanatory relationships that forms the backbone of the scientific study of causation and is ubiquitous in ordinary life.

Since the relation of closeness used in counterfactual reasoning is defined in terms of explanation, we typically cannot establish an explanatory claim by counterfactual reasoning unless we already have some knowledge about explanatory relationships. But there is no threat of circularity, since the explanatory knowledge required for our reasoning differs from the explanatory knowledge we gain as a result of it (*MER*: sct. 11.5). Counterfactual judgments mediate the inference from old items of explanatory knowledge to new ones, thereby allowing us to extend our stock of such knowledge. I hold that that is one of the most important functions of ontic modal thought, and I also suggest (more speculatively) that we developed the capacity for ontic modal thinking in large part because of its utility for this purpose.

My account explains why counterfactual reasoning is a reliable method of testing causal and other explanatory claims across a wide range of circumstances. But I argue in Chapter 12 that the view also predicts and explains why the method doesn't work in certain other cases, like those of causal overdetermination and preemption. These are the examples that have dogged counterfactual accounts of causation. My theory can account for them.

4. Modality in metaphysics

During the last half century or so, modal notions have played a pivotal role in many metaphysical theories. Examples include supervenience formulations of physicalism, counterfactual analyses of causation, and the modal conception of essence. On the whole, such approaches have not fared very well. This fact has recently motivated a number of philosophers to doubt that modal concepts deserve the central place they have occupied in metaphysical theories, and to hold that such notions as grounding and essence are more apt to play this role. As explained in more detail in Kment 2015, the account of *MER* underwrites this shift of focus from the modal to the explanatory domain. Modal facts concern a relation of comparative closeness between certain classes of propositions (the worlds) that is not metaphysically deep or fundamental in any sense. Facts about explanatory connections, and

facts about essences and the metaphysical laws, are more fundamental than modal facts and better suited to form part of the subject matter of metaphysics. At the same time, the theory of *MER* can explain why modal considerations have figured so prominently in many philosophical debates whose ultimate concern is with explanation. For it entails that modal facts, e.g. counterfactual dependencies, supervenience relationships, and facts about which propositions are necessary, often *reflect* explanatory connections or facts that are of interest because of their central explanatory role (such as facts about the essences of things). Modal facts therefore constitute an important set of data in the study of explanatory relationships. For example, a hypothesis about essence, grounding, or metaphysical fundamentality can be evaluated in part by its consistency with the modal facts and its ability to explain them. In these cases, the modal facts are not themselves the ultimate targets of the investigation, but are of interest solely in their role as evidence.