Notes on Chapter 1 of *Making Things Happen*

Branden Fitelson
02/15/11

§1.1 Woodward begins by contrasting what he is doing with various other projects in the general vicinity. First, he *narrows* his attention to "causal" approaches to scientific explanation (as opposed, *e.g.*, to Churchland's conception of scientific explanation, which is much *broader*).

Woodward also contrasts his approach with other "causal" approaches, which he takes to be *too narrow*. For instance, he mentions the "causal process" accounts of Salmon (and Dowe), according to which causal explanations must involve the *tracing* of certain sorts of *physical processes* from cause to effect. Woodward also highlights Sober's distinction between "causal" and "equilibrium" explanations of (*e.g.*) the following explanandum:

(*E*) a certain collection *g* of a gas is in a certain (*type of*) physical state *s* at a certain time *t*

(where this *type of* state *s* is characterized in terms of pressure, temperature, and volume).

On Sober's view, a (properly) *causal* explanation of *E* will involve *tracing the actual sequence of events leading up to E*. Presumably, this will involve (at the fundamental level) describing the dynamical evolution of *the particles in g*, from some "initial condition", up to *t*. Sober contrasts this type of "causal" story with an "equilibrium" explanation of *E*, which involves showing that it is in some sense "highly probable" that *g* would (eventually) end-up in a state of type *s*. On Woodward's view, "causal" includes *both* of Sober's kinds of explanation. According to Woodward...

> ...any explanation that proceeds by showing how an outcome depends (where the dependence in question is not logical or conceptual) on other variables or factors counts as causal. ... the distinguishing feature of causal explanations...is that they are explanations that furnish information that is potentially relevant to manipulation and control: they tell us how, if we were able to change the value of one or more variables, we could change the value of other variables. According to this conception, both derivations involving the ideal gas law and Sober's equilibrium explanations count as causal explanations.

The emphasis on *manipulation and control* is important here. Presumably, Woodward does *not* include *Sober*'s sense of "causal" here, since he thinks the microscopic story about the dynamical evolution of the particles comprising a collection of an ideal gas is *not* particularly relevant to *manipulation and control of the state of the gas* (described in terms of temperature, pressure, and volume). On the other hand, the ideal gas law states just the sort of (functional) relationship between "variables" that Woodward considers to be important for such manipulation and control. I'm less clear on why Sober's "equilibrium" explanation should be "causal" in Woodward's sense.

Perhaps this is just because Sober's "equilibrium" explanation will involve probabilistic relations between the salient variables (*i.e.*, the ones that *are* "handles" for "manipulation and control" of the type of state of an ideal gas). The precise *meaning* of "manipulation and control" (and "handles for") will be a recurring theme in our discussions of Woodward's book.

§1.2 Woodward claims that a standard (meta-philosophical) view in this literature is that there are only two things one might be doing when one tries to come up with an account of "causation":

(1) attempting to provide a "conceptual analysis" of "cause" and related locutions, where this is a matter of describing ordinary and/or scientific usage.

(2) engaging in what Dowe calls "empirical analysis", where this is to "discover what causation is in the objective world."

Woodward distinguishes his own project from both (1) and (2). With respect to (1), Woodward points out that he is not merely interested in how people (actually) use words. He is interested in "larger practices of causal inference and explanation in scientific and nonscientific contexts", where these "larger practices" are not merely linguistic practices. Moreover, his aim is to make some *new distinctions* among various sorts of causal and explanatory claims — distinctions that are often overlooked by "actual practitioners". Finally, and most interestingly, Woodward takes himself to be involved in something that has *pre*scriptive (not merely *de*scriptive) significance. He wants to be able to offer *recommendations* about "what we *ought to mean*" by various causal and explanatory claims (or "how we *ought to engage*" in the larger practices he has in mind). This involves clearing up confusions and ambiguities, as well as offering accounts that have a clear set of *aims* and *methods*. One way of putting this is that Woodward wants a way of engaging in causal and explanatory activities that has a certain kind of *rational structure*.

Woodward grants that *prior* (*actual*) *usage* of causal and explanatory language places *important constraints* on any adequate account. But, there are various other constraints that Woodward thinks are perhaps more fundamental. These desiderata are discussed in chapter 1, and then summarized toward the end of the chapter (§1.9). But, here, I would say the main thing Woodward is focusing on is the claim that our causal and explanatory practices have certain *aims* or *purposes*. And, accounts of "causation" and "explanation" (and related concepts) should — ultimately — be answerable to how effective they are at allowing us to achieve these aims.

He mentions various historical precedents for the sort of "partially revisionary" project he has in mind. He discusses mathematical accounts of intuitive geometrical notions like

"continuity". In those cases, prior usage was often plagued by various (sometimes subtly hidden) unclarities, which were then "cleaned up" by the mathematical accounts. Woodward seems to think there has been a lack of philosophical discussion of the sort of project he is engaging in.

On this score, I find it surprising that he does not discuss Carnap's notion of "explication". For Carnap, *explication* is not "mere conceptual analysis" either. It, too, has a "prescriptive component", which aims to "clean up" ordinary (or scientific) usage of a concept (or family of concepts). Indeed, I think it can be plausibly argued that Carnap's *primary* way of approaching philosophical analysis was *via* explication. The process of explication involves, first, clarifying the *explicandum* (*i.e.*, the informal concept, the prior usage of which we aim, in part, to "clean up"). Once the explicandum has been sufficiently clarified, we then go on to construct precise (usually formal) candidate *explicata* for our explicandum. For instance, we may offer various mathematical definitions of "continuity" as candidate *explciata* for some "continuity explicandum" (which can be extracted/gleaned from prior usage in various ways). Then, we will offer various *desiderata*, which are philosophical constraints on our candidate *explicata*. Finally, by applying our *desiderata*, we will select a "best *explicatum*" for our *explicandum*. For instance, we may settle upon a particular mathematical definition of "continuity" as the best *explicatum* for a given "continuity *explicandum*". Interestingly, Woodward himself offers several desiderata in this first chapter (see below). Presumably, those are to be used to narrow down sets of candidate explicata for various explicanda among the family of concepts surrounding "cause", "explanation", *etc*. [Next week, in our readings for Cricky's visit, we have papers by Strawson and Carnap, which involve a debate about the philosophical value and significance of the sort of precise, formal explications Carnap offered. Also, I have added Carnap's "On Explication" to our readings.]

§ 1.3 This section is meant to provide (an informal and preliminary) sense of the kind of "manipulability" conception of causation that Woodward has in mind (think of this as part of his "clarifying an *explicandum*"). The key contrast here is between "explanatory information" and "descriptive information" (and between "explanation" and "mere description"). The idea is supposed to be that, roughly, the two kinds of "knowledge" or "information" contrast thusly:

(a) *Explanatory* (or causal) information is information that is relevant to manipulating, controlling, or changing nature, in an "in principle" sense (to be clarified in chapter 3).

(b) *Descriptive* (or non-causal) information is information that, although it may provide a basis for prediction, classification, or more or less unified representation or systematization, does not provide information potentially relevant to manipulation.

A lot of work will have to be done to flesh-out "manipulability in principle", *etc.* But, even at this rough level of description, we can see an important *intuitive* distinction (*i.e.*, at the level of "explicandum"). Remember back to our Berkeley graduate school example from last time. Why did we think that the partition <department$_i$> was *more explanatorily probative* than the partition <"F"–"K" surname, non-"F"–"K" surname>? One thing that seems (*intuitively*) to be the case here is that *changing the first letter of one's surname does not seem to be a way to manipulate/control whether (or the probability that) you will be admitted into graduate school.* So, at least at this intuitive level, we can see some hope that that this kind of "causal information" was the "X-factor" (or one of the factors) that was *missing* from the S–R account of scientific explanation.

One complication that arises immediately is the fact that it is often the case that manipulation/control are *actually not possible* (for one reason or another). This is where the "in principle" caveat becomes important. Woodward begins to clarify what (a) means, as follows:

(a*) *Explanatory* information (regarding a phenomenon) is information involving factors such that *if* (perhaps contrary to fact) manipulation of these factors *were* possible, then this would afford a way of manipulating or altering the phenomenon in question.

Getting a grip on what this might mean will involve first working out a manipulationist account that makes sense in cases where manipulation *is* possible, and then *extending* this account to cases where manipulation is not possible. Woodward gives the example of a causal explanation of the extinction of the dinosaurs, which appeals to a (hypothesized) asteroid impact (as explanans). Supposing we cannot manipulate the past, this explanation can only be given a manipulationist gloss if we have some way of making sense of (a*). The idea is that there causal explanation involves answering certain "what-if-things-had-been-different questions" — from a "manipulationist" perspective. And, sometimes (as in the dinosaur extinction explanation), such questions involve thinking about "manipulations" of factors that are not (actually) manipulable. [In this connection, Woodward mentions the interesting hypothesis that if we had been unable to manipulate nature (in Woodward's sense), then we would not have developed (Woodwardian) notions of "cause" and "explanation". Dummett discusses such a possibility in his "Bringing About the Past". He also contrasts this view of science — as deeply concerned with manipulation and control and causal explanation — with a "passive empiricism" that is only concerned with description or prediction.]

§1.4 In this section, we start to get a more concrete sense of what Woodward *means* by his key terms *causation*, *invariance*, and *intervention*. For illustration purposes, he introduces an example of (a Newtonian model of) a block sliding down an inclined plane.

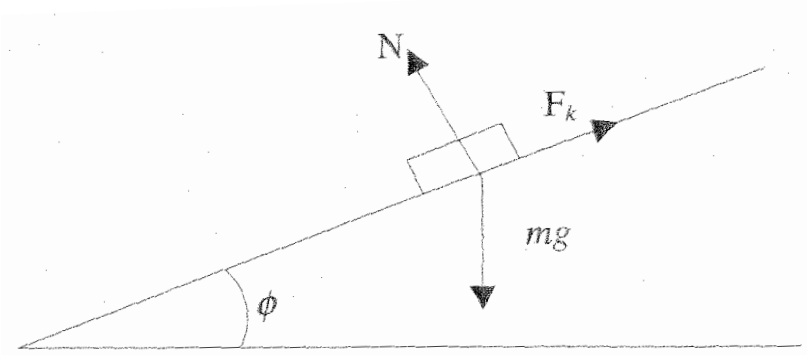

Figure 1.4.1

Woodward offers the "textbook explanation" of "what accounts for the motion of the block". According to this explanation, there are three forces operating on the block: a gravitational force F_g due to the weight of the block, a normal force N, which is perpendicular to the plane, and a frictional force F_k which opposes the motion of the block. Woodward then writes down some (functional) equations (from Newton's theory), which tell us the magnitudes of these forces:

(1.4.2) $F_k = \mu_k N$

(1.4.3) $F_g = m g$

Here, μ_k is a coefficient of kinetic friction, m is the mass of the block, and g is the (assumed constant) acceleration due to gravity (near the earth's surface). The gravitational force is directed toward the center of the earth. The component of the gravitational force directed down the plane is given by $F_g \sin \phi = m g \sin \phi$, and the normal force (which is equal and opposite to the component of the gravitational force that is perpendicular to the plane) is $F_g \cos \phi = m g \cos \phi$. Thus, the frictional force is $F_k = \mu_k m g \cos \phi$. Finally, the *net* force acting on the block is just the sum of the two forces acting along the direction of the plane. That is:

(1.4.4) $F_{net} = m g \sin \phi - \mu_k m g \cos \phi$

And, then, the net acceleration on the block is:

(1.4.5) $a = g (\sin \phi - \mu_k \cos \phi)$

Now, this will certainly allow us to *predict* the subsequent motion of the block (provided that we know its initial position and velocity). But, Woodward thinks that it also furnishes an *explanation* of the motion of the block. This is because the equations (1.4.2)–(1.4.5) encode various

counterfactual relationships between the variables, which will allow us to answer various "what-if-things-had-been-different" questions. Presumably, however, there are going to be some explanatory/causal *asymmetries* here. For instance, while we can use (1.4.2)–(1.4.5) to "predict" the value of ϕ from knowledge of (all) the values of (all) the other variables, I take it that we do *not* think that changing the mass of the block (m) is a way of changing the value of ϕ. In other words, it seems that m is *not* a "handle" for "controlling" ϕ. So, clearly, more will need to be said concerning the proper "causal interpretation" of such systems of functional equations. We'll get to that below (*invariance*). For now, the important thing to keep in mind is that Woodward *will* have an account of such explanatory/causal asymmetries. For instance, Woodward discusses the barometer/storm example. Let S be a (dichotomous) variable, which takes the value 1 if a storm is approaching, and the value 0 if a storm is not approaching. And, let B be a variable whose values are given by a (properly functioning) barometer. Woodward explains that:

> (1.4.7) If the value of B were changed as a result of an intervention, then the value of S would change.

is a *false* counterfactual (on his theory of non-backtracking counterfactuals in causal models). And, as a result, on his account, B will *not* end-up counting as a cause of S. *Roughly*, Woodward's account of causal/explanatory relevance will come down to the following counterfactual:

> X is a cause of Y (or X figures in some causal explanation of Y) iff the value of Y would change under some intervention on X in some background circumstances (which may include interventions on *other* variables *besides* X and Y).

We'll learn *a lot* more about how this (rough) idea gets precisified in subsequent chapters (and Cricky will be here next week talking about some interesting examples of Woodwardian causal models).

Another crucial notion in the Woodward book is that of an *invariant generalization*. His notion of invariant generalization does not line-up exactly with any existing notion of "law", but Woodward thinks that invariance does many of the (explanatory) jobs that "lawhood" was *supposed* to do. Roughly, Woodward's invariance idea can be expressed as follows:

> A generalization G (relating two variables X and Y) is *invariant* iff G would continue to hold under some intervention that changes the value of X in such a way that — according to G — the value of Y would change (where "continue to hold" means that G correctly describes how Y would change under the interventions on X in question).

On this understanding of invariance, the correlation between B and S does *not* result in an invariant generalization concerning the values of S *as a function of the values of* B, since this

correlation *breaks down* under (all) *interventions* on B. On the other hand, the equation (1.4.2) *is* an invariant generalization, since there are some interventions on N for which (1.4.2) continues to hold (as a relationship between F_k and N). When a relationship is invariant under at least some interventions, it is (thereby) *potentially* useful for manipulation and control, in the sense that even if an actual intervention on X is not possible, it remains true (on Woodward's account) that *if* a (suitable) intervention on X *were* performed, *then* the value of Y would change (in accordance with G). Note: the invariance of a generalization G is *not* undermined by the fact that there may also be some *other* interventions on X under which G *does* break down. For Woodward, "laws" (in any traditional sense) are just going to be *one kind* of invariant generalization.

§ 1.5 In this section, some important distinctions among different kinds of causal explanations are drawn. First, there is the distinction between *levels* of explanation. The explanation of the extinction of the dinosaurs is a *token* causal explanation: the *explanandum* is a *token* event. In contrast, the explanation offered in the sliding-block example is an explanation that applies to a *type* of event or a *phenomenon* (that is, *any* block under similar circumstances will be subject to the Newtonian explanatory strategy employed there). Another important distinction here is that between *net* causes *vs direct* or *contributing* or *component* causes. In the block example, we have a *net* cause of the block's motion (the net force), and we also have *contributing* or *component* causes of the block's motion (*e.g.*, the frictional force). The final topic discussed here involves the "depth" of various causal explanations. Consider the following causal claim:

(1.5.1) Depressing the gas pedal in my car causes it to accelerate.

This causal claim *can* feature in certain explanations of the car's acceleration. But, intuitively, these would be rather "shallow" explanations compared to an explanation based on knowledge of the internal mechanisms or workings of the car (that, say, a mechanic may be able to offer). An account of causal explanation should be able to say something about this sort of "depth". Woodward will account for this by appealing to (a) different *degrees of invariance* of the generalizations the feature in the explanations, and (b) differences in the *range of what-if-things-had-been-different questions* that we can answer using the generalizations involved.

§ 1.6 In this section, Woodward urges that there is an important place for (some sort of) causal explanations in "ordinary life". Many of these explanations (and the "folk causal knowledge" that comes along with them) are rather mundane and "shallow" (in the above sense). But,

nonetheless, there does seem to be *something* important (in *practical* terms) about engaging in "the everyday activity of causal explanation". He calls this "the practical point or payoff" of causal explanation (eventually, he'll argue that *the point is* "manipulation and control").

Woodward also urges in this section that there should be some sort of "continuity" between the everyday/practical causal explanations we encounter, and the more rarified forms of explanation we see in science (and philosophy thereof). Specifically, he thinks we should have:

(1) *Substantive* continuity. Our account of scientific explanation should *build on* the folk practices surrounding causal explanations in ordinary life. [Here, he contrasts his view with those philosophers who view folk and commonsense beliefs as *fundamentally mistaken*.]

(2) *Methodological* continuity. We should expect causal explanations in different areas of science to share at least some structural features with causal explanations in more ordinary contexts. We should see scientists who construct such explanations as attempting to satisfy some of the explanatory goals of ordinary or folk explanation (but achieving these goals using knowledge that is more detailed and systematic).

In this sense, Woodward thinks that scientific explanation is a kind of "improved" or "better" version of "ordinary" causal explanation. So, he thinks that many others who write about scientific explanation have things *backward*. Hempel (and others) have suggested that scientific explanations involve (a) explicit chains of deductive inference, and (b) laws of nature (in a technical, philosophical sense). They view "ordinary" or "folk" explanations as legitimate only to the extent that they can be seen as somehow "tacitly" relying on (a) and (b). Woodward is skeptical about the "backing" relation that (allegedly) obtains between "laws" (in any traditional sense) and "garden variety causal claims". He also thinks that "deductive validity" and "law" are "sophisticated products of a specific philosophical tradition". And, he recommends that we *start* with the "folk" or "ordinary" causal claims/explanations (understood in manipulationist terms), and then think of the (specialized, scientific) explanations involving chains of deductive inference, laws, *etc*., as *special cases* of (manipulationist) causal explanations (in a broad sense).

§ 1.7 This section is about reductionism. Woodward is (unabashedly) *non*-reductionist. For Woodward, reductionism has various aspects. A *reductive* account of concepts in some family *F*:

- should be *non-circular* — it should analyze concepts in *F* using only concepts *C* that are *not* contained in *F*.

- should be *M&E-kosher* — the concepts *C* it uses should satisfy certain epistemological and metaphysical constraints (*e.g.*, "empiricist" or "Humean" constraints).

Woodward is not *opposed* to reductionism (in principle). He just thinks that no adequate (or illuminating, *etc.*) non-reductive account of causation (or causal explanation) is (in fact) forthcoming (but he would welcome one). Moreover, he thinks that his own preferred non-reductionist accounts *are* adequate (at least, that they satisfy *his desiderata*, which do *not* include reductionism). Of course, we'll have to judge for ourselves how this all works out in the details...

One interesting (and important) point Woodward makes in this section is that he thinks the "standard" philosophical analyses of counterfactuals are *inadequate* for grounding our causal and explanatory practices. Moreover, he offers an alternative account of counterfactuals, and their proper role in explanations (only hinted at in chapter 1), which he thinks is superior in this regard. He could be right about *this* point, *even if* he is wrong about reductionism. This is because there is a way of understanding his criticisms that is *orthogonal* to the reductionism debate — it has only to do with the *precise way* in which the various concepts in the "causal/explanatory circle" are *interconnected*. There are *trivial* or *viciously* circular ways of interconnecting, and ones which aren't. Woodward's account of *C* causes *E* is *circular*, but (he claims) *not viciously* so. His account of *C* causes *E* will involve *interventions I* on *C* (and what happens to *E* under such *I*s). The notion of an intervention *I* on *C* is *itself* a causal notion (Woodward will speak of interventions *I* that *cause* changes in *C*). The idea here is that the causal relations needed to characterize what it is for *I* to be an intervention on *C* are different from the causal relationship between *C* and *E*. Recall:

> *C* is a cause of *E* (or *C* figures in some causal explanation of *E*) iff the value of *E* would change under some intervention *I* on *C* in some background circumstances (which may include interventions on *other* variables *besides C* and *E*).

Here, the "would change" *itself* involves causal relations, but not the same causal relation we're explicating. Woodward claims this is *non*-vicious (at least not in the way "*C* causes *E* iff *C produces E*" is, where "production" is *primitive*, and we're not given a good story about what "production" is and how it relates to "causation"). How compelling this claim is will depend on the details.

§ 1.8 Of course, the metaphysics and epistemology of causation (explanation, *etc.*) are *conceptually distinct*. But, it would nonetheless be nice if we had a story about the metaphysics of causation (explanation, *etc.*) that is accompanied by some (not altogether implausible) story about how human beings *might be in a position to know* that causal/explanatory claims are true (or how human

beings might be able to *acquire evidence that bears on* the truth of such claims). And, this story should dovetail with things we think we already know about *paradigm cases* of testing causal hypotheses (*e.g.*, *controlled experimentation* of various kinds).

According to Woodward, causal relationships are "out there" in nature. But, explanation has to do with the "discovery and provision" of *information*, *by* human beings. As such, another epistemological requirement is that explanatory information must be "epistemically accessible" (in some sense). That is, explanatory information must be information that can be "recognized, surveyed, and appreciated" — it must be able to "contribute to understanding". Woodward suggests here (and discusses further in chapter 4) that this requirement is *not* met by some existing accounts of scientific explanation. He views this as an advantage of his account of scientific explanation over those alternatives.

§ 1.9 Here, the various *desiderata* Woodward has in mind are summarized. They include:

- *Description*. An account of scientific explanation must capture relevant features of paradigmatic explanations in *both science and ordinary life*. It should provide some insight into how such explanations *work* — how they manage to "convey understanding".

- *Unification*. Our account should *unify* different kinds of "causal explanation" — it should be able to reveal what it is that such explanations *have in common*.

- *Evaluation*. Our account should allow us to *evaluate* explanations. It should (a) allow us to distinguish explanatory claims from merely descriptive ones, (b) allow us to distinguish are *better* and *less good* explanations, and (c) it should enable us to understand the *grounds* on which the normative assessments of explanations are made.

- *Supersession*. Our account should *supersede* previous accounts of scientific explanation. That is, it should elucidate the successes (and failures) of previous accounts, and it should be able to solve/avoid problems that previous accounts of explanation have faced.

- *Accessibility*. Our account must explain how it is that explanatory information is (in some sense) *epistemologically accessible*, and how it can "contribute to understanding".