Draft. Please do not quote.

Thinking about Interventions:

Optogenetics, Experiments, and Maker's Knowledge[1]

Carl F. Craver

Key Words: Experimentation, Intervention, Causation, Mechanisms, Applied versus Basic Sciences, Maker's Knowledge, Optogenetics.

1. Introduction. The biological sciences, like other mechanistic sciences, comprise both a modeler's and a maker's tradition. The aim of the modeler in biology, in the narrow sense intended here, is to describe correctly the causal structures, the mechanisms, that produce, underlie, maintain, or modulate a given phenomenon or effect seen in the living world.[2] Such models are expected to save the phenomenon tolerably well (that is, to make accurate predictions about it) and, in many cases, to correctly represent the components and causal relationships composing the mechanism for that phenomenon. The aim of a maker, in contrast, is to build machines that produce, underlie, maintain, or modulate effects we desire. Such maker's knowledge might be deployed in the service of modeler's knowledge, as when engineering triumphs become the next generation of experimental intervention and detection, or it might be deployed for good or ill to serve our needs.[3]

[1] Thank you to the Minnesota Causality and Biology workshop for comments on early drafts, especially John Beatty, David Danks, Christopher Hitchcock, Sarah Roe, Roberta Millstein, Ken Waters, and Marcel Weber. Thanks also to Andreas Hütteman's Research Group on Causality and Explanation at the Universität zu Köln for support and helpful discussion in Spring 2011. Finally, thanks to Charlie Kurth for discussing the epistemology of a single case and Pamela Speh for assistance with figures. Thanks also to William Newsome for first bringing optogenetics to my attention and to Marshall M. Weinberg and the Molecular, Cellular, & Developmental Biology Department at the University of Michigan for providing funds and a workshop invitation that justified setting time aside to think about these issues more carefully.

[2] I am using the term "modeler" in a more restrictive sense than typical. Not all models are mechanistic models (see Bogen 2005; Craver 2007; Kaplan and Craver 2011). Here I focus on mechanistic modelers, reverse engineers. Makers are engineers simpliciter.

[3] Martin Carrier (2004) discusses the relationship bdtween applied and basic science, which is related to the narrower distinction between modelers and makers I have in mind. At points in this essay, I address Carrier's challenge to catalogue some differences between the two kinds of knowledge. I take no stand on whether one kind of knowledge is more fundamental than the other and am content for the moment only to note that they are distinct when it comes to an epistemology of intervention.

The works of maker and modeler alike depend fundamentally on the ability to intervene into a system and make it work contrary to how it would work were it left to its own devices. The aim of this essay is to identify some dimensions progress (or at least difference) among different means of intervening into biological systems for these modeling and making objectives.

In what follows, I use an example to illustrate these diverse dimensions of progress and difference in intervention: optogenetics. Optogenetics is a kind of genetic manipulation that makes neurons responsive to light. Karl Deisseroth and colleagues published the first paper using optogenetics in 2005. In 2010, *Nature Methods* named optogenetics the Method of the Year. *Science* that year included it among the Top Breakthroughs of the Decade. At the time of writing, over a hundred papers using optogenetic interventions have been published in the highest profile journals in neuroscience. It is widely acknowledged, in other words, that optogenetics constitutes an advance in our ability to intervene into the brain. By looking at how researchers justify this new intervention technique, we gain some insight into the requirements that researchers place on interventions, the arguments by which intervention techniques are validated, and the dimensions along which one intervention technique might be said to improve upon another.

Optogenetics allows researchers to control electrophysiological properties of neurons with light.[4] Researchers insert bacterial genes for light-sensitive ion channels into target cells in a given brain region. They couple these genes to gene regulatory units that ensure the gene is expressed only in specific cell types. The virus by which this genetic construct is inserted into cells commandeers the cell’s protein synthesis and delivery mechanisms to assemble the channels and insert them into the cell’s

[4] Some use the term optogenetics to describe both intervention and detection techniques. Optogenetic detection techniques, for example, use gene constructs to cause cells to fluoresce when they express a protein, for example. I am concerned exclusively with intervention.

membrane. The researcher then inserts a fiber optic cable into the brain near the region of interest. Light delivered through the cable activates the newly inserted channels. The channels open, allowing ions to flow across the membrane. This ionic current can be used to raise or lower the neuron's membrane potential, and so to modulate or produce electrophysiological signals.[5]

To bring out the advantages of this new intervention technique, I first present a standard schema for thinking about causal experiments. Then I discuss twelve dimensions of progress or difference in the ability to intervene into brain function. For many of these dimensions, what counts as progress can be specified only within the context of a given experimental or practical objective. Nonetheless, by exploring some of the qualities that distinguish intervention techniques from one another, we get a feel for the epistemological principles that guide the assessment of progress in intervention. To catalogue such dimensions does not itself amount to an epistemology of intervention. For such an epistemology, this largely descriptive approach should be supplemented with a normative framework showing how these twelve dimensions of virtue make a difference to causal inference. Here I take some preliminary steps in that direction, but my primary objective is to simply frame some of the most salient dimensions of evaluation in a way that makes transparent where such justificatory arguments might be focused. I close by reflecting on some points of overlap and difference between the ways that makers and modelers think about the epistemology of intervention.

[5] For accessible reviews of optogenetics, see Dieseroth 2010; 2011; Fenno 2011). An early paper presenting the technique is Zhang (2006). For a discussion of its application in nonhuman primates, see Diester (2011)

2. Causal Experiments. Figure 1 represents a standard form of causal experiment.[6] A given causal or mechanism schema is instantiated in a *target system*. The target system is the subject, organism, system, or sample group in which one performs the experiment.[7] One *intervenes* on the system to change one or more *target variables* (*T*), and one *detects* the value of one or more putative *effect variables* (*E*).

The *intervention technique* is a means of introducing a change into one or more target variables in the mechanism. One literally sets the value of the target variable to the desired value. In other cases, one might find natural circumstances under which the target variable has been set to that value, but here I focus on interventions that are under the researcher's control.

The *detection technique* (iii) is a device or process that takes some feature or magnitude in the experimental system as an input and returns a reading or measure as an output. Detection techniques are indicators of the value of a variable. For example, litmus paper can be used to assess pH, an osmometer can detect ionic concentrations, and functional MRI can be used to detect dendritic field potentials.[8] A reliable detection technique indicates well: it is tightly correlated with the value of the

[6] This experimental framework is introduced more slowly and with a different example in Craver and Darden (forthcoming).

[7] Rheinberger uses the term "experimental system" to refer to something broader than I have in mind by a target system. Rheinberger's systems include for example, the experimental subject, the intervention and detection techniques, lab protocols, preparatory procedures, storage devices, data analysis techniques, and the like (see Rheinberger 1997; Weber 2005). Rheinberger describes experimental systems as the least unit of experimental analysis. However, I focus almost exclusively on intervention techniques, leaving the rest of the experimental system as background.

[8] Although I treat intervention and detection techniques as independent components of a typical causal experiment, this is not generally the case. Often, new intervention techniques require or at least spawn the development of new techniques to measure the hypothesized effects of the interventions. The synergy of experimentation and detection, as well as the virtues of combining techniques, are fascinating topics for future discussion.

variable in such a way that one can infer the value of that variable in the target system within some margin of error.

Franklin (1986, 1990, 2009) discusses several strategies by which scientists argue for the reliability of their detection techniques. They argue, for example, that the instrument reliably detects known magnitudes, that its reports are consistent with what one would expect on the basis of a well-confirmed theory, or that its readouts match tolerably well the measurements produced by another, more ore less causally independent instrument one already trusts.[9] They show that the instrument is designed on the basis of principles supported by well-confirmed theories (e.g., carbon dating), or that its readout cannot be explained by known sources of detection error.[10] In fact, Franklin (1986, 1990) defines the epistemology of experiment in terms of the validity of detection, in particular, the ability to distinguish 'between a valid observation or measurement and an artifact created by the experimental apparatus' (1990, 104; see 1986, 165, 192). One goal of this paper is to take similar steps forward in thinking about the epistemology of intervention.[11]

[9] Chang (2004) discusses aspects of detection validity in history of the thermometer.

[10] Franklin (1990) grounds his account of these strategies in Bayesian statistics. Mayo (1996) provides an alternative analysis grounded in error statistics and severity of testing. Weber (2012) reviews some of the advantages and disadvantages of these and other approaches to the epistemology of detection.

[11] Franklin (2009) focuses on two experiments in biology. He describes Kettlewell's experiments on moths and the Meselsohn-Stahl experiment demonstrating the semi-conservative nature of DNA replication. In the case of Kettlewell's experiments, the intervention is entirely out of the hands of the experimenters. Industrial pollution changes the coloration of tree bark, causing the number of dark (as opposed to light) moths to increase. Birds pluck moths from soot-covered branches. The researcher sits behind the camera. Likewise, the primary interventions in the Meselsohn-Stahl experiment are a) the radiolabeling of Nitrogen, and b) the replication of bacteria. The first of these interventions is entirely in the service of detecting the difference between parental DNA and offspring DNA (it is an eliciting condition, see below). The second of these interventions is up to the bacteria themselves; the experimenter decides only when the process comes to an end, not how the process unfolds. As a result, the examples do not afford much of an opportunity to think about interventions of the sort in our standard causal experiment, though they do prompt us to recognize

Two final components of the standard causal experiment are *intervention checks* and *eliciting conditions*. Researchers use *intervention checks* to detect the value of the target variable and to determine the extent to which the intervention has, indeed, succeeded in setting the target variable to the desired value. In animal lesion experiments, for example, one typically confirms the location of a lesion using CT scans or post-mortem inspection of brain tissue. In a pharmacological experiment, one might measure the concentration of the pharmacological agent in the blood. Intervention checks are used to confirm that the intervention technique works as it is supposed to work. They are also used to compensate for unreliable intervention techniques. If one has doubts about the reliability of one's interventions, one can simply measure the value to which the target variable has actually been set, and use the detected value, rather than the desired value, in one's analysis of the causal experiment.

varieties of experiments besides those considered in the standard scheme. See Craver and Darden (forthcoming; Chs. 6 and 7) for a fuller discussion of kinds of experiments.

Elliciting conditions (bottom right, Figure 1) are interventions required to prepare the effect variable for detection. In microscopic studies, experimenters treat tissues with stains. In classic PET studies, subjects are given radiolabeled glucose so that researchers can track regional changes in blood flow. Eliciting conditions are also interventions into the system (hence the causal arrow in Figure 1), but these interventions are directed ultimately at the putative effect variables rather than the cause variables.[12]

3. Dimensions of Progress in Intervention. Focus now just on interventions. Woodward (2003) offers an account of causation that relies fundamentally on the concept of intervention. On his view, the claim that *X* causes *Y* amounts roughly to the *claim* that there exists an *ideal intervention* on *X* that changes the value of *Y*. An intervention need not be ideal in this proprietary sense to deliver useful information about the causal structure of a system. Nonetheless, the concept of an ideal intervention provides a useful starting point for a discussion of how researchers make progress in the ability to manipulate biological systems for modelers and makers objectives.

Figure 2 summarizes the requirements on an ideal intervention. Unidirectional arrows represent causal relations. Bidirectional dotted arrows represent correlations. Bars across arrows indicate that the relation must be absent. The intervention in this case is designed to test whether the target variable, *T*, makes a difference to the effect variable, *E*. An intervention, *I*, fixes the value of *T* so that it is independent of all the other causal influences, *U*, on *T* (1). This intervention produces a change in *E* that is not mediated directly (2). Nor is the effect produced by directly changing a variable *S*

[12] Hacking (1983) uses staining as a primary example of the role of intervention in the growth of scientific knowledge. But eliciting conditions do not play the same epistemic role as do interventions into targets for the purpose of conducting a causal expeirment.

that is intermediate between *T* and *E* (3). Nor is it produced because *I* is correlated with some other variable *C* that is causally relevant to the value of *E* (4). Nor or is it produced by influencing the detection apparatus required to measure the effect of the intervention (5).[13] Such conditions assert, in a nutshell, that any observed change in the value of *E* can be attributed to the change to the target variable and not to some other change induced by the intervention.

Many of these conditions are exemplified by the standard use of placebos in drug trials and sham surgeries in lesion studies. The ideal is to contrive control conditions that mimic the intervention in all respects except for the change to *T*. This practice is in place to assess whether the changes observed in *E* are due to the changes in *T* or to some other unanticipated effect of *I* upon (or some correlation of I with) some other variable causally relevant to *E*. [14]

{Figure 3 Interventions are themselves complex mechanisms}

Consider now some ways that the schemas represented in Figures 1 and 2 might be complicated to accommodate experiments typical of a biological or neuroscientific research report.

First, in most interesting experiments, one can decompose *I* into a complex causal sequence (as in Figure 3) or a set of such causal sequences. As discussed further below, the experiment involves an experimental choice (*C*), an action or procedure on the part of the experimenter (*A*), a device or machine (*M*) that delivers

[13] Woodward's account of an ideal intervention does not include requirement (5), as it is not irrelevant to the semantics of causal claims. It is, however, relevant to how causal claims are tested. Likewise, in experiments crucially involving an intervention check, one would want to know if the intervention influences the intervention check independently of the change induced in the target variable T.

[14] I stress again that not all useful interventions are ideal in this proprietary sense: Woodward's goal to provide a semantics of causal claims (that is, an answer to the question, "What do we mean when we assert that X causes Y?"). He does not provide, nor does he intend to provide, an account of the constraints that an experiment must satisfy to reveal useful information about the causal structure of a system. Nor is his account intended as a model of all causal experiments.

the intervention to the system (*P*) in which the target variable (*T*) is located. Figure 3 is misleadingly simplified in representing the intervention as a single sequence.

Consider optogenetics, for example. The namesake intervention in optogenetics requires a causal connection between light and neural activity. That connection has to be constructed. One must build the gene construct, insert it in a virus, infect the relevant cells, and get the cells to express and traffic the protein. Additionally, one must drill through the skull and insert a fiber optic cable.[15] This complex intervention also typically takes place along side a number of "support" interventions that prepare the system for study: the animal is kept in constant temperatures, fed standard chow, maintained in regular light/dark cycles, and the like. And in many cases, there is more than one experimental manipulation. The system is frequently surgically prepared or pre-treated in ways required to make the experiment possible or convenient. The animal might have been trained on a task, for example. In one experiment discussed below, researchers first intervene with a traditional electrophysiological stimulus to stimulate the basolateral amygdala (BlA) in order to test the effects of the second intervention, optogenetic inhibition of BlA terminals in the central amygdala (CeA). It is in most cases a tremendous simplification to represent interventions as a single variable, I, on a target variable.

Second, and most importantly, many experiments fall short of the ideals expressed in Figure 2. In some cases, it might be impossible or undesirable to change *T* in a way that screens *T* off (makes the value of *T* independent of) from its other causes (as required in 1). One might wish, for example, to increase the amount of dopamine in a system while allowing endogenous dopamine to vary under the influence of its typical, physiological causes. Interventions are frequently ham-fisted,

[15] Channels responding to infrared light might allow researchers to skip this step.

changing many potentially relevant variables at once; so one can hope (but never truly know) whether such interventions in fact satisfy constraints such as 2 or 3. In other cases, experimenters simply might not know whether their interventions in fact satisfy the requirements on ideal interventions represented in Figure 2. They might not know about unmeasured variables or causal relationships among variables that run afoul of constraints 1-5.

The point is that Woodward's notion of an ideal intervention, useful as it is as a guide to the semantics of causal claims, does not, and was not intended to, provide a set of necessary or sufficient conditions on an adequate experimental test of causal relations. If one were to take Woodward's analysis as providing such a set of conditions, one would be forced to conclude that very few actual experiments succeed in testing causal claims. And this conclusion is clearly false. This essay is about progress in intervention, and presumes therefore that one can speak meaningfully of more or less ideal interventions, of one intervention being better or worse than another for testing a given causal claim.

But what are the dimensions along which one evaluates whether one intervention technique is better or worse than another?[16] Table 1 lists twelve dimensions of virtue in intervention. In the body of this section, I discuss each in turn, using optogenetics to illustrate how these dimensions contribute to the evaluation of techniques.[17]

{Table 1 Near Here}

[16] Note that this question is distinct from the one that motivates Hacking's (1983) interest in intervention. I am more interested in maker's knowledge itself than I am in the possibility that maker's knowledge provides a new reason to be a realist about entities.

[17] I use the term, dimension here for want of something better. Little would be served by representing a hyperspace of these dimensions. Still, the term captures the idea that the usefulness of an intervention for answering a particular question depends on a number of independently varying features of the intervention. The term "virtue" might serve as well or better, and I will sometimes use it, though it connotes inappropriately that it is always good or better to have more the dimension in question.

3.1 Number of Variables. Perhaps the most obvious dimension of progress in intervention concerns the number and diversity of variables a researcher can control. Suppose one believes the theta rhythm in the hippocampus is causally relevant to the function of the cells in the hippocampus. If one has no ability to manipulate the theta rhythm to see whether changes in its frequency and amplitude, for example, alter the function of the hippocampus, a question will remain whether the observational data correlating the theta rhythm and aspects of hippocampal function indicate that the theta rhythm contributes to hippocampal function. Similarly for distributed cortical representations. At the moment, at least for many areas of the cortex, researchers have no idea how to intervene on the cortex to produce physiologically relevant patterns of activation across widely distributed brain networks (perhaps involving millions of neurons). Clearly the development of such a technique would be a very significant advance and would help to place talk of "codes" and "representations" on a firmer epistemic footing in our models of the causal structure of the brain. So to begin with the most banal observation, we make progress in intervention to the extent that we have the capacity to manipulate more and more of the variables that potentially make a difference to how things work.

{Figure 4. Precision among variables}

3.2 Precision Among Variables. The second dimension, *precision among the variables*, is more relevant to the case at hand. Deisseroth lists precision as the primary advantage of optogenetics:

> What excites neuroscientists about optogenetics is control over defined events within defined cell types at defined times--- a level of *precision* that is most likely crucial to biological understanding beyond neuroscience." (Deisseroth 2010*)*.

Optogenetics achieves this end through the genetic mechanisms that allow researchers to determine which cells express the proteins required to construct the light-sensitive channels. Boolean combinations of promoters and inhibitors in the gene regulatory construct allow researchers to target specific cells with specific chemical signatures, specific morphologies, and specific topological connectivity. Only infected cells that have the right signatures to initiate expression become light-responsive.

This is a decided advantage over both electrophysiological and many pharmacological techniques. When one uses an electrode to deliver current to a brain region, the current spreads through the brain matter, falling off in intensity depending on the current delivered and distance. As a result, the technique tends to excite all the cells in a given region of cortex, running roughshod over known divisions among cell types. On the assumption that these differences in cell type in fact make a difference to the mechanisms in which they participate, a technique that fails to distinguish among them is incapable of evaluating a broad range of causal hypotheses.

A similar point can be made for optogenetics relative to pharmacological interventions. Pharmacology gives one considerably greater precision than does electrophysiological stimulation (as applied to large brain regions). Pharmacological agonists and antagonists can be used to interfere with or stimulate specific signaling pathways in the brain (such as ion channels and intracellular molecular cascades). However, pharmacological techniques are often non-specific in the sense that a given drug works on many systems at once (e.g., all dopamine cells). A technique more precise in disambiguating different putative causal variables is to be preferred over one that is less precise. Such an improved technique allows one potentially to discern a wider number of differences that might make a difference.

Let us be a bit more explicit about the relevant kind of precision. Suppose we are testing a causal hypothesis (H), that *T* makes a causal difference to *E*. The experimental goal is to intervene on *T* to change its value from what it would otherwise have been, and then to detect the consequences of this intervention (if any) on the value of *E*. Now we can imagine a set of alternative hypotheses, each of which asserts a different putative cause variable for *E*: *T**, *T***, *T****. An intervention technique that changes the value of T alone will be of greater value for discriminating among these hypothesis than will an intervention technique that changes *T*, *T**, *T***, and *T**** at the same time. Rather, one needs an intervention technique that is more precise in sorting among the variables that might be causally relevant to the effect in question. This follows from the abstract discussion of ideal interventions above: one ideal of experimental intervention is that the technique should allow one to *change the putative cause variable without changing other putative cause variables at the same time*.

In other words, the greater the precision of the intervention, the more one can eliminate potential confounds. A confound in this sense is an effect of the intervention other than the change to *T* that might explain the difference (or absence of a difference) in the effect variable subsequent to the intervention. The higher the precision in one's intervention techniques, the greater is one's ability to home in on a particular variable and distinguish it from others. In the domain of maker's knowledge, greater precision among the variables might allow one to devise interventions with reduced risk of inessential side effects. This is because side effects often result from the ham-fisted nature of the therapeutic intervention. (Think for example, of the extrapyramidal effects of certain tricyclic antidepressants).

One respect in which present-day optogenetics remains ham fisted is that optogenetic stimulation makes an entire population of neurons pulse synchronously with one another. Given that the neural code in many systems is likely not to be carried by synchronous activity in neuronal populations, there is a spatial dimension of coding beyond the reach of current generation optogenetics.

3.3. Grain and Range Among Values of a Variable. Suppose one wants to control the firing rate of neurons in the BlA. One might still wonder how, exactly, one wants those neurons to fire. An intervention technique for such a variable will be better to the extent that it allows one to explore the space of possible firing patterns and rates that might make a difference. First generation optogenetic techniques gave researchers control over the firing rates of action potentials with mean interspike intervals around 100-200 ms. By modifying the structure of the channel protein, researchers can now reliably produce action potentials with a mean interspike interval around 5 ms (see Gunaydin 2010). As Boyden, et al., put it, "This technology thus brings optical control to the *temporal regime* occupied by the fundamental building blocks of neural computation" (Byden et al., 2005; italics in original).

To be a bit more abstract, suppose that one has decided upon a given variable of interest (such as mean spiking rate or a given temporal pattern of spikes). One technique is better than another to the extent that it allows the researcher to explore the space of plausible *switch-points* for that variable (see Craver 2007). A switch-point in the value of a variable is a difference in the value of the variable that makes a difference to the effect variable. Zero degrees Celsius, for example, is roughly the switch point in temperature that makes a difference to whether water is liquid or solid. Such differences might be analog, where each increment in the cause variable (no matter how small) makes some difference to the effect variable (e.g., mass and gravitational attraction), or they might be digital (there is a *minimally effective difference* in the cause variable). In either case, a switch point is the border between values of the cause variable associated with one value of the effect variable and neighboring values of the cause variable associated with a different value of the effect variable.

It is often of considerable interest to set the putative cause variables to the widest *range* of relevant values. To test theories of heat, for example, one would want means of producing heat across a broad spectrum from extremely high to extremely low temperatures. (See Chang's (2005) for a discussion of detection in this context). Likewise, if one is interested in exploring the effects of mean firing rate on the behavior of a population of neurons, one would want control over the known spectrum of that firing rate, or perhaps even simply a plausible range of the known firing rates. The greater the range of the intervention technique the more it can be used to explore the space of possible switch-points. The switch-points thus discovered also become the buttons and levers that might be exploited in the pursuit of maker's knowledge.

An additional dimension of progress concerns the *grain* of the intervention: the slightest change that the technique can reliably induce in the cause variable. The appropriate grain of intervention varies with the system in question and with the pragmatic use to which the intervention is to be put. Pharmacological interventions, for example, give one the ability to influence the chemical environments of neurons in ways that can directly affect neuronal activity. However, the effect on the electrophysiological properties of neurons is imprecise in comparison to optogenetics. Pharmacological antagonists increase or decrease the probability of neuronal activity, but they do not allow one precisely to regulate the mean firing rate of the neurons in the area, let alone the precise temporal patterns of their activity. Systems-level pharmacology thus affords comparatively crude control over the electrophysiological behavior of the system. To the extent that one believes that precise rate or temporal coding is relevant to the behavior of the system, one would then have reason to prefer optogenetic interventions.

{Figure 5. Interventions with different grains and ranges}.

These ideas are represented in Figure 5. At the top left is an intervention into *T* that covers its entire range of values at a very fine grain. In the limit, *I* can be used to set *T* to any value it can take. To the right is an intervention technique that allows the researcher to set the putative cause variable to only part of its range of possible values. The grain, we might assume, is the same, but its range is diminished relative to the ideal just mentioned. On the bottom is a technique that covers the same range as the first, but does so at a cruder grain. The technique lacks the precision required to set *T* to a fine-grained value and can be counted upon only to set the variable to a value somewhere within a particular range.

3.4. Physiological/Ecological Relevance. In biology, the maker's tradition and the modeler's tradition often place different demands on a researcher. The modeler wants to understand how some biological system works, and typically this means that she wants to understand how it works when it works properly and in its standard operating conditions. In such investigations, one enforces a sharp distinction between how a system in fact works and how it might be made to work under extreme or otherwise unusual conditions. Modelers working on physiological systems will thus prize intervention techniques that can be used to manipulate physiologically relevant variables in ways and within levels that are similar to the ways and levels at which those variables change in the "typical" or "normal" course of their biological working. Modelers working on organisms in an environmental context, likewise, might try to devise interventions that are appropriate to the "typical" or "normal" context in which a given organism lives.

I use scare-quotes around these terms to reflect the fact that they stand for an unspecified teleological orientation that filters the typical and normal from the atypical and abnormal.[18] Makers care less than modelers about how something typically or normally contributes to well-being than they do about how something might be made to contribute to well-being, how it might be modified to do so better, or (as in the case of optogenetics) how it might be transported to a different context to do new work. Makers might be concerned for health, life, and the good of the species, but they think of these ends as both malleable in the face of future technological development and lacking a position of privilege in the space of possible ends. They recognize in biological systems a set of buttons and levers that might be pushed and pulled for a variety of purposes, some of which are irrelevant in evolutionary or physiological contexts (see Craver 2010; forthcoming).

Be that as it may, one measure of progress in intervention from the modeler's perspective is the ability to intervene in ways that more or less mimic the types of change one normally sees in the target system. A few different aspects of this assessment are, i) whether the intervention targets the *variables* operative in the system as it typically works, ii) whether the values to which the target variable can be set are within the *range* of values those variables take during the typical functioning of the target system, iii) whether the stimulus values change at *rates* or otherwise in manners comparable to how they change during the typical functioning of the system

[18] I see little point in attempting to ground such distinctions objectively (see Craver forthcoming). Judgments of typicality presuppose a choice of a reference class (think, for example, of the normal cancer cell or the typical mammalian cell expressing bacterial rhodopsin). Judgments of normality are inherently tinged by a preference for health, life, the good of the species, or some such valued end.

under investigation, iv) whether the changes in the stimulus values are induced via a *mechanism* similar to the mechanism by which the changes are produced in the typical operation of the system.

Many of the early papers on optogenetics are motivated by the desire to satisfy these criteria. The technique is targeted at electrophysiological properties known to be relevant to the function of the nervous system (i). As discussed in Section 3.3, the technique targets the physiological range of electrophysiological function. Given that the spike trains can be timed specifically, and given that spike trains have a high degree of replicability trial to trial (as discussed below), researchers are excited about having the ability to produce electrical signals in neurons that resemble closely the very signals that neurons produce when they work. Finally, although optogenetics does not produce action potentials by precisely the same mechanisms that would typically produce action potentials in cells, the mechanism is strongly analogous to such physiological mechanisms, involving the flux of ions through a membrane-spanning channel (iv). Compared to electrophysiological stimulation, such signals are considerably more physiologically relevant in this respect.[19]

This dimension of progress--- physiological relevance--- clearly matters more to the modeler than to the maker. Though the maker always has to take certain practical constraints into consideration, the maker is not constrained to intervene within the limits of normal or typical functioning in a teleological sense. The maker

[19] There are also respects in which optogenetic stimulation is unlike the typical action potential. In modified channel rhodopsins, for example, the temporal properties of the action potential are not precisely the same as in the target cells. And as mentioned above, and as in electrophysiological interventions, applied light influences the behavior of a population of neurons all at once, more or less synchronously. And whereas it would be reasonable to expect that populations of neurons interact in virtue of patterns of activation and inactivation across a population of neurons, the inability at this moment to produce such changes should be counted as a potential limitation as the method currently stands.

can look for any intervention that might be usefully commandeered for the purposes of making the system work for us. For the maker, functional biology offers only a blinkered view of the space of possible causal structures that might be put to new uses in engineering.

3.5. Reversibility. A further dimension for evaluating intervention techniques is the capacity to reverse the intervention. Think, for example, of the role of lesion techniques in experimental neuropsychology. In the simplest case, one opens the skull of a model organism and uses a scalpel, a vacuum, or an electrical stimulus to remove a piece of brain tissue. Though brains do tend to reorganize and recover over time, one cannot simply replace the brain tissue. One consequence of this limitation is that one must compare the performance of the organism on a cognitive task prior to the lesion with the performance after the lesion. One is not able, however, to test again, in the same organism, the effects of changing the target variable back and forth.

Advances in lesion techniques have given researchers more control over this situation. For example, some researchers produce functional lesions by simply inserting a probe that cools brain tissue to a temperature at which the tissue no longer functions. Other researchers lesion areas of the cortex functionally with magnetic coils that interfere with the mass electrical activities of neurons in that area. The advantage of these intervention techniques is that one can intervene to change the value of the target variable (turning an area from "on" to "off") and then intervene again to restore the target variable to its prior value. This allows one to run experiments and controls in the same animal and in whichever order one desires.

Pharmacological interventions into neural systems have the benefit of reversibility, but in many systems, the effects are reversible only very slowly. Optogenetics has a decided advantage in the rate at which researchers can reverse the intervention. In standard optogenetics applications, light changes the probability that the channel enters an open or a closed state. Removing the light removes the energy required to maintain the active state, and the channel goes to its rest condition. In more exotic applications involving *step function opsins* one wavelength of light activates the channel, allowing a steady flow of ions across the membrane, and a different wavelength of light turns it off. The reversal is, for most practical purposes, immediate.

The value of reversibility is illustrated dramatically in a set of experiments on the role of the BlA in anxiety. It is possible to chart, in one and the same animal, the effect turning the light stimulus to the BlA on and off. The effect of this intervention on anxiety is detected by watching how mice behave in open spaces. Mice typically avoid them. Placed in a large, open box, the mouse invariably sticks to the walls and corners, only rarely wandering across the open center of the box. One can make a mouse even more anxious (more wall-hugging) by stimulating its BlA, a sub-region of the limbic system that sends axonal projections to the CeA (among other structures). By inhibiting those projections in the CeA, one can effectively nullify the effect of the stimulation. Figure 6 shows the effect of electrophysiological stimulation in the BlA with or without optogenetic inhibition of terminals in the central amygdala.[20] The data represented in Figure 6 are made possible by the fact that one can turn the light stimulus on and off, comparing the experimental and control conditions in the same populations over time.

{Figure 6. Reversible optogenetic interventions on a mouse model of anxiety.}

The point is that an intervention that can move a variable away from baseline, bring it back to baseline, move it away again, and then bring it back gives one

[20] Note that this is an example of a causal experiment with two interventions.

considerably more experimental flexibility in one and the same experimental system. This avoids some of the confounds introduced by the need to compare an experimental group and a control group, given that one and the same target system (e.g., subject) is used for each. A reversible technique also allows one to choose the order of repeated experimental and control conditions, thus allowing one to detect and account for order effects.

3.6 Bivalence. A related, but distinct, dimension of progress in intervention is to move from interventions that can manipulate the value of a variable only in one direction to interventions that can manipulate the value of the variable in both directions. Lesion studies, for example, simply remove the part in question. They cannot stimulate it to action. Optogenetics, for reasons just described, gives researchers the capacity to both stimulate cells and to inhibit their activity. Bivalence is represented in figure 7.

{Figure 7. Bivalent interventions can increase and decrease the value of an intervention in one and the same target system.}

Why is bivalence a virtue? Bivalence plays a crucial role in the search for switch-points in the value of a cause variable relative to a given effect variable. A switch-point, again, is a difference in the value of the cause variable that makes a difference to the value of the effect variable. A bivalent intervention technique allows one to raise or lower the value of the putative cause variable from its baseline. Bivalent interventions allow one to explore whether the cause variable and the effect variable vary concomitantly across the spectrum of their values and perhaps to describe that concomitant variation mathematically. (Think, for example, of how Hodgkin and Huxley (1952) used the voltage clamp to explore how conductance changes with membrane voltage). Bivalent interventions also allow one to explore higher-order differences produced by different kinds of changes: to assess the effect of increases and decreases, different rates of increase or decrease, and different patterns of increasing and decreasing. Bivalent intervention techniques thus allow one to explore a wider range of values, and a wider range of changes to the value of the relevant target variables, than do univalent intervention techniques.

3.7. Efficacy. As noted in the introduction to this section, interventions are typically complex causal sequences. As shown in Figure 3, interventions typically involve experimenter choices (*C*), experimental actions (*A*), machines or instruments of varying complexity (*M*), and interactions with the target system itself (*P*). Just as an experimenter's efforts to *detect* a given variable can be complicated by, for example, failures of the detection apparatus or failures on the part of the experimenter to correctly record what the detection apparatus reports, experimenter's efforts to *intervene* into a target system might be complicated by failures anywhere along this causal chain.

An experimenter might choose to make a particular intervention but fail to take the appropriate action. Perhaps her hand wiggles at the wrong moment, or she grabs the wrong vile from the chemical storeroom. Moving through the causal sequence, she might make the appropriate action, but her device might fail to initiate the causal sequence that would actualize her intentions within the target system (perhaps the machine has a twitter). Likewise, the machine might succeed while the target system somehow foils the would-be intervention. In a drug trial, for example, the subject might forget to take the drug, or the patient might have an enzyme in his stomach that degrades the drug before it enters the blood stream. Because interventions are themselves complex causal processes, they are the kinds of things that might break in any number of ways. If we assume that the experimenter typically acts as she chooses to act, the efficacy of an intervention might be informally described as the reliability with which the intervention technique produces its desired outcome (or something close enough).

Efficacy is in part a matter of objective frequency: given that an experimenter decides to set *T* to a particular value (or, more accurately, to within a particular range of values), and given that she has signaled her instruments to deliver the appropriate intervention, what is the probability that *T* actually takes the chosen value (or falls within an acceptable range around the target value)?

In the case of optogenetics, two measures of efficacy have been particularly important. One measure is the extent to which the technique succeeds in causing the expression of rhodopsin channels in the membranes of all and only the target cells. Tsai et al (2009) describe the result as follows:

> Greater than 90% of the TH immunopositive cells were positive for ChR2-EYFP near virus injection sites and more than 50% were positive overall, demonstrating a highly *efficacious* transduction of the TH cells (2009).

TH, tyrosine hydroxylase, is a marker for dopaminergic cells. ChR2-EYFP is a marker for the expression of the Channel-Rhodopsin 2 gene. Tsai et al thus use intervention checks to establish the efficacy of their intervention. As impressive as these results are, they are measured against an ideal of 100% expression in the relevant cells. Indeed, one potential limit of optogenetics techniques is the extent to which different kinds of cells in different regions of the brain (in different experimental organisms) can, in fact, be prodded into expressing the channel rhodopsins.

Another objective frequency measure relevant to establishing the effectiveness of optogenetics as a technique is the extent to which individual pulses of light in the relevant wavelength in fact succeed in producing action potentials in the target cells. The first paper to describe optogenetic interventions (Boyden et al. 2005) is largely dedicated to establishing this dimension of control. They show that light induces stable ionic currents with very rapid activation kinetics. They explore different stimulus durations in order to fine-tune the technique to reliably deliver trains of action potentials. They characterize how many cells in a population produce action potentials. They show that the same patterns of action potentials can be repeated time and again in the same neuron and across different neurons by delivering the same pattern of light stimuli.

These distinct arguments about the efficacy of optogenetics address different parts of a complex intervention mechanism. The first (expression rates) concerns the efficacy of the relationship between delivery of the virus and channel expression in the target cell population. The second set of findings concerns the efficacy of the relationship between light and electrophysiological activity in the target cell population. All of these interventions ultimately result in the generation of action potentials (*T*), but this effect requires the confluence of many *tributary interventions*. What's more, the second intervention is dependent upon the first; if the cells did not express the channels, or did so only with very low efficacy, then light would not become a reliable means to generate action potentials in neurons. The tributary tasks of making the gene construct and inserting it into cells are required to build a mechanism that bridges the causal gap between light and electrophysiology. Such tributary tasks are what Fred Dretske calls *structuring causes* in his discussions of representation; in this case, however, the structuring causes build the mechanism that allows the experimenter to actualize her intentions in the target variable. Light is a *triggering cause* in the system thus assembled.

Efficacy might be understood in terms of the variance of the target variable given that one initiates an intervention to fix the value of that variable. Given that one has decided to set a magnitude *T* to a particular value *t*, one wants to know the probability that *T* takes a value within a range around *t*. One could then determine the efficacy of a technique by summing the squared distances between the intended and the actual value on a number of trials and dividing that sum by the number of trials. Efficacy is improved by shrinking the variance: bringing more of the actual values closer to the intended value. As interventions improve in this respect, one's confidence that the intended value of the variable (or something tolerably close) will be actualized is increased.[21]

An intervention need not be perfectly efficacious to be useful, either for the modeler testing a causal hypothesis or for the maker hoping to effect some practical outcome. Whether an intervention is efficacious enough for a given epistemic or practical objective depends on the variable in question and on the relevant switch-points for the effect variable. For the modeler, a terribly inefficacious intervention can be perfectly useful *so long as* appropriate intervention checks are in place to determine whether or not the intervention has produced the desired outcome in each particular case. For the maker, however, failures are failures, even if one can tell that the fault for the failure lies with the equipment. This is an important difference between modeling and making.[22]

3.8 Dominance. The dominance of an intervention is the extent to which it satisfies condition I4 in Woodward's view of ideal interventions (see U in Figure 2). In an ideal intervention, the intervention technique sets *T* to a value independent of all

[21] This paragraph builds on a suggestion from Frederick Eberhardt.
[22] Thanks to Christopher Hitchcock for this point.

T's other causes. Such an intervention screens off the value of *T* from the effect of all causes of *T* besides *I.*

Woodward includes I4 among the conditions on an intervention because if such a condition does not hold, one is not in a position to say, in the token case, that the intervention is responsible for the change observed in the putative cause variable. Perhaps, for example, the intervention does not set *T* to the desired value. Perhaps its effects are more than swamped by compensatory or coincidental changes induced by the other causes of *T*. The case is perhaps clearest when one intervenes in such a way to reduce the value of *T* and fails to notice a change in E. Perhaps other causes of *T* compensate for the intervention or coincidentally raise the value of *T* so as to erase the effect of the intervention. In that case, it would be wrong to conclude that the induced change is causally irrelevant to the effect.[23]

Along this dimension, interventions can be hard or soft (see Eberhardt 2009). A hard intervention removes all causal arrows besides the intervention into the target variable. This means that the value of the target variable is influenced, to the extent that it can be influenced, only by the intervention. A soft intervention does not fix the value of the target variable; rather, it changes the conditional probability distribution over the values of the target variable. Hard interventions are arrow-breaking, eliminating the influence of *T*'s other causes. Soft interventions are arrow-preserving, leaving at least some of the parental influence on *T* intact.

[23] One way to meet this kind of challenge is to use intervention checks to confirm that one has, in fact, set the value of the variable to the desired level. Consider again the experiment from Tye et al. 2011. In that experiment, they record from the CeA during the two major interventions of the study. The first intervention stimulates the BlA. The second intervention inhibits the cells of the CeA onto which the cells of the BlA project. The detection technique located in the CeA is designed to check that the interventions (electrophysiological stimulation in the BlA and light stimulation of the CeA) in fact have changed the activation of CeA cells as desired. By measuring the value of the target variable for the intervention, one can confirm the value of the target variable to rule out the problems with non-dominating intervention discussed above. See note 23.

Optogenetic interventions can be hard or soft depending on how the experimental system is prepared. A single neuron in a dish, for example, can be removed from its cellular context such that the light stimulus is the only exogenous determinant of cellular activity.[24] Researchers using step function opsins, in contrast, simply raise the membrane potential to make it more probable that an incoming signal will in fact lead to the generation of an action potential in the post-synaptic cell. This is a soft intervention.

Eberhadt and Scheines (2007) discuss some of the advantages of hard interventions for learning about a system's causal structure. First, if the intervention breaks all of the causal arrows into the target variable, then one can infer that the observed correlation between the target variable and the putative effect variable is not due to the action of a common cause. The effect of any such common cause on the target variable is, by definition, severed by the intervention. Second, hard interventions allow one to assess the direction of the causal influence between two correlated variables, i.e., to distinguish cases in which *A* causes *B* from cases in which *B* causes *A*. Finally, because hard interventions allow the experimenter to set the value of the cause variable, it gives the researcher information that might be used, for example, to characterize the strength of the causal relationship.

Often, however, such dominating interventions are impossible or undesirable. They are practically impossible when there is no known means to break all of the causal arrows into the target variable. They are undesirable in cases where one wants to leave the causal structure of the system intact, for experimental or practical reasons.

[24] Even under such circumstances, one might still expect occasional "noise" in the system. Ion channels are chancy machines. Thus, one might inhibit electrical activity in a cell significantly and still see the occasional spike. In the context of other experiments, however, the light pulses are added in addition to the inputs at dendrites that might generate spikes on their own.

The value of non-dominating interventions is perhaps most apparent in the domain of maker's knowledge. Many standard medical interventions are non-dominating. Insulin injections for diabetes, for example, augment rather than replace endogenous insulin production. L-Dopa treatments for Parkinson's disease augment rather than replace endogenous dopamine levels. In such cases, one intervenes to boost residual capacities of the system and/or to encourage it to compensate for its weaknesses, not to replace it with an artificial control system (as a heart and lung machine replaces hearts and lungs while they are off-line).

Modelers can also make considerable use of soft interventions. Eberhardt and Scheines show, for example, that there are conditions in which multiple simultaneous soft interventions on a system will suffice to determine the system's causal structure in a single experiment (unlike hard interventions). Furthermore, if only a single intervention is allowed per experiment, then the choice between hard and soft interventions makes no difference to the rate at which the data from the series of experiments will converge on the correct causal structure (Eberhardt and Scheines 2007). These results must be bracketed by the additional assumption that one knows in advance that the variables in one's understanding of the causal structure are jointly sufficient to determine the value of the effect variable in question. If one allows for the possibility of unmeasured common-cause confounds, as is likely the case in most biological experiments, then there is a clear advantage to dominating interventions if one wants to learn about the causal structure of the system.

In the domain of maker's knowledge, whether a hard or a soft intervention is desirable or necessary depends on a host of pragmatic factors: on precisely what one wants to do with the intervention, on the ethical constraints on possible interventions, and on the practical constraints on such interventions (such as cost and functionality in context). The move from soft interventions to hard interventions (or vice versa) is not, that is, an intrinsic form of progress in the ability to intervene into a system. Hard and soft interventions should rather be seen as distinct intervention strategies that can be used to answer different kinds of question or solve different kinds of practical problems.

3.9 Determinism. Some interventions are deterministic. Others are probabilistic. In a deterministic intervention, the intervention technique is used to set the value of the target variable to one and only one value. In indeterministic interventions, one sets merely the probability distribution over possible values of the target variable.

In several of the early experiments on optogenetic control of neural systems, the researchers intervened by stimulating cells with pulses of light spaced around a mean interspike interval of, for example, 100 ms. The precise timing of any given spike is allowed to vary so long as the entire train of spikes lands around a mean interspike of 100 ms. The timing is generated computationally to fit a normal distribution around a desired mean.

Whether one wants to determine the precise sequence of spikes or merely a probability distribution around a mean interspike interval depends on what, precisely, one wants to know. In these experiments, researchers were merely hoping to demonstrate that the technique could be used to generate action potentials at rates approaching physiological ranges. They also wanted to show that they could do so without presuming any particular pattern of stimulation. One might also hypothesize that mean interspike interval is the difference-making variable for the effect of interest. In that case, the experimenter chooses to intervene indeterministically (i.e., the indeterminacy falls early in Figure 3, at node *C*).

It is far more common, however, that interventions are simply *de facto* indeterministic because there is a stochastic element in the complex intervention mechanism. Most real-world interventions are probabilistic to some extent, if only due to differences in efficacy and dominance. People make mistakes. No intervention instrument (*M*) is perfect. Experimental subjects might have individual differences in how they respond to the intervention (*P*). Any formal treatment of the epistemology of interventions must accommodate the fact that interventions are often chancy affairs, and that chance can enter into a complex intervention mechanism at many different stages.

3.10 Replicability. Replicability of the intended change to the target variable is clearly desirable if one is to make inferences about the causal structure a system on the basis of more than one trial. In abstract analyses of experimental situations, replicability is presumed; but in fact it is typically a matter of degree.

The issue arises starkly in classic lesion experiments in neuroscience. Typically researchers enter the skull of a model organism, locating a vacuum tube or an electrode at a particular stereotaxic coordinate relative to the shape of the head. The researcher then suctions or burns the area until the desired amount of brain tissue has been removed. Skilled researchers acquire a knack for removing just the right amount for a given experimental intervention. But the method itself allows for significant variation from one lesion to the next. And individual brains might differ sufficiently one to the next that the same stereotaxic coordinates in fact direct the lesion to different brain structures. One might speak of the average lesion, or the average volume of tissue removed, and one can use intervention checks to characterize just where the lesions were and how they differ from one animal to the next. But there is always some difference between the lesions produced in two different brains or in two halves of the same brain.

One crucial test of optogenetics as a technique was to establish that the same patterns of neural excitation could be produced time and again in the same cell. On separate trials, the pattern of action potentials matched one another 95% at mean interspike intervals of 100 ms and 98% at intervals of 200 ms. Comparing across cells, the same optogenetic stimulus produced trains of action potentials that match trial to trial by 85% (Boyden et al. 2005).

Such replicability is also likely also desirable for makers, who must rely on repeatable interventions in to make the brain work for us. For both modelers and makers, what counts as a replication will be defined more or less loosely for different epistemological and practical objectives. Given the chosen contrast in the values of the effect variable, one can begin to ask after the relevant switch-points in the value of the cause variable. Such judgments, that is, must be made relative to a decision about the relevant effect contrast in question. Yet there remains an objective sense in which advances in replicability of intervention, to whatever arbitrary degree possible, count as a steps forward in intervention.

3.12. Summary. Table 1 lists twelve, epistemically relevant ways that one intervention might differ from another. In some cases, it is possible to speak meaningfully of a kind of progress being achieved by moving from one end of a single dimension to another. In other cases, any preference along that dimension depends fundamentally on the empirical or practical problem the researcher faces. Such context-relativity, however, is underwritten by a deeper sense that there are distinct kinds of causal question and distinct forms of intervention that are appropriately or most efficiently used to address them. This list of virtues is no doubt incomplete; and a proper epistemology of intervention awaits a normative framework for thinking about how these distinct virtues contribute to the ability to search the space of possible causal models and build new things. Nonetheless, this preliminary list captures many of the virtues that must be discussed by any adequate epistemology of intervention.

4. Conclusion: Makers and Modelers Revisited. Progress in the ability to intervene just is progress in the ability to control and produce phenomena with whatever reliability and precision we desire. In medical contexts, for example, the dream is to deploy maker's knowledge to cure and prevent diseases, to build prosthetic devices (such as artificial hearts and brain-machine interfaces), and to encourage recovery and reorganization in such systems after damage or disease. This growth in human know-how is one of the primary fruits that Bacon envisioned for organized science (1620; see Sargent 2001; 2012).

Maker's knowledge is a potent stimulus to the search for modeler's knowledge (see Carrier 2004). In the laboratory, especially in sciences such as neuroscience, growth of knowledge about causal systems depends fundamentally on the development of new ways not merely to detect how biological systems function but to intervene into those systems and make them work in new ways. Detection techniques, such as functional imaging and single unit recording, offer scientists a window on how the brain functions during the performance of tasks or in response to other stimuli. Such techniques are indispensible in the search for causes. Yet such techniques alone are demonstrably incapable of disambiguating certain causal structures (Eberhardt 2009). Such ambiguities can only be removed systematically through the judicious use of appropriate interventions. Advances in intervention, such as optogenetics, make an irreplaceable contribution to our ability to search the space of possible models for a given phenomenon.

More than that, advances in intervention often bring with them new conceptual tools, allowing researchers to envision previously occluded parts of the space of possible mechanisms. Just as the telescope allowed Galileo and Kepler to see new things, and the invention of many different kinds of telescope broadened the range of things one can detect about astronomical objects, developments in intervention techniques have the capacity to open new worlds to a researcher. New intervention techniques earn their keep by allowing researchers to test old causal hypotheses with new precision, by revealing previously unknown causal structures, and by encouraging researchers to think in creative ways about how the system might be causally organized. The very act of constructing experiments forces the experimenter to take explicit responsibility for the values of relevant variables and to vary those concrete commitments from one experiment to the next. So just as developments in the telescope has come to suggest new principles of organization of astronomy, optogenetics will likely transform in subtle ways how researchers think about the transmission of information through neural circuits.

I underscore, however, the serious differences between maker's knowledge of how to do things with biological systems and modeler's knowledge of how biological systems work. The search for modeler's knowledge, in the restricted sense intended here, starts with the task of understanding how a given biological mechanism works teleologically. One choses a phenomenon of teleological interest, and one tries to understand how the different parts of the mechanism are organized together such that they produce, underlie, or maintain that phenomenon. One includes in one's model of the mechanism all and only the variables that make a difference to that phenomenon.

Maker’s knowledge is not restricted in this way. Makers begin with an engineering problem. They assess the parts and tools at their disposal, and they set about designing a machine that solves the problem out of the available parts. The maker’s parts need not correspond at all to the modeler’s parts. Where the modeling neuroscientist asks, “How does the brain do this?” the maker asks “How can I get the brain to do this?” And for the maker, “this” might be something quite unusual when viewed from a physiological or evolutionary perspective (such as driving a robotic arm or playing PONG with EEG waves).

Perhaps more to the heart of the matter, makers and modelers take different attitudes on the epistemology of their projects. The solution to an engineering project wears success on its sleeve: the machine either works or it does not. It does not need to be pretty; it does not need to mirror a natural system; it just needs to work. There is nothing more to “getting it right,” for a maker, than producing the right input-output relationship within design constraints. Modelers cannot rest comfortable with simply *producing* the phenomenon in this way. They face the additional epistemic burden of saving the phenomenon and, in the cases with which I am concerned, describing the mechanism that typically or normally produces that phenomenon in physiologically and ecologically relevant conditions. Modelers, in other words, also take on epistemic commitments to mirroring tolerably well the causal structure of the system in question (Kaplan and Craver 2011).

The science of genetics has always included both maker’s and modeler’s traditions. Today, the science of genetics is arguably more about making than it is about modeling. Neuroscience certainly has its roots in medical science, but much of the work in cognitive neuroscience over the last thirty years has been driven by modeler’s aspirations. Such modeling is required to even begin to identify possible means for manipulating brain circuitry in potentially useful ways. The development of optogenetics, however, marks a significant step in the coming ascendancy of maker's knowledge of the brain, which ascendancy will no doubt continue to fascinate us in coming decades.

References

Bacon, F. (1620) *The New Organon.* In Graham Rees (Ed.). The Oxford Francis Bacon (Vol. 11). Oxford: Clarendon Press. Revised 2004.

Bacon, F. (1626) *The New Atlantis.* From *Ideal Commonwealths*, P.F. Collier & Son, New York.(c)1901 The Colonial Press, expired. Prepared by Kirk Crady from scanner output provided by Internet Wiretap. This book is in the public domain, released August 1993. http://oregonstate.edu/instruct/phl302/texts/bacon/atlantis.html

Bogen, J. (2005) ''Regularities and Causality; Generalizations and Causal Explanations,'' in C. F. Craver and L. Darden (eds.), ''Mechanisms in Biology,'' *Studies in History and Philosophy of Biological and Biomedical Sciences*, 36: 397–420.

Boyden ES, Zhang F, Bamberg E, Nagel G, Deisseroth K. (2005) Millisecond-timescale, genetically targeted optical control of neural activity. *Nature Neuroscience*. 8:1263-8.

Carrier, M. (2004) "Knowledge and Control: On the Bearing of Epistemic Values in Applied Science", in: P. Machamer & G. Wolters (eds.), *Science, Values and Objectivity*, Pittsburgh: University of Pittsburgh Press; Konstanz: Universitätsverlag, 275-293.

Chang, H. (2004) *Inventing Temperature: Measurement and Scientific Progress*. Oxford University Press, New York.

Craver, C.F. (2007) *Explaining the Brain: Mechanisms and the Mosaic Unity of Neuroscience.* Carlendon Press, Oxford.

Craver, C.F. (2010) Prosthetic Models *Philosophy of Science*, 77: 840-51.

Craver, C.F. (forthcoming) Functions and Mechanisms: A Perspectivalist Account. In P. Hunneman, ed. *Functions*: *Selection and Mechanisms*. Springer, Synthese Press.

Craver, C.F. and Darden, L. (forthcoming) *The Search for Mechanisms: Discoveries across the Life Sciences.* University of Chicago Press, Chicago.

Deisseroth, K. (2011) Optogenetics. *Nature Methods*. 8:26-9.

Deisseroth K. (2010). Controlling the brain with light. *Scientific American*. 303:48-55.

Diester I, KaufmanMT, Mogri M, Pashaie R, Goo W, Yizhar O, Ramakrishnan C, Deisseroth K, Shenoy KV. (2011). An optogenetic toolbox designed for primates. *Nature Neuroscience.* Epub Jan 30.

Dretske, Fred (1988) *Explaining Behavior: Reasons in a world of Causes.* Cambridge, Mass.: MIT Press. A Bradford Book.

Eberhardt, F. (2009). Introduction to the Epistemology of Causation. In *The Philosophy Compass*, 4(6):913-925.

Eberhardt, F. & Scheines, R. (2007). Interventions and Causal Inference. In *Philosophy of Science*, 74:981-995.

Fenno LE, Yizhar O, Deisseroth K. (2011) The development and application of optogenetics. *Annual Review of Neuroscience*. 34:389-412.

Franklin, A, (2012). Experimentation in Physics. *Stanford Encyclopedia of Philosophy*. Available online at http://plato.stanford.edu/entries/physics-experiment/.

Franklin, A. (1990). *Experiment, Right or Wrong*. Cambridge: Cambridge University Press.

Franklin, A. 1986. *The Neglect of Experiment*. Cambridge: Cambridge University Press.

Gunaydin LA, Yizhar O, Berndt A, Sohal VS, Deisseroth K, Hegemann P. (2010) Ultrafast optogenetic control. *Nature Neuroscience*.13: 387-92.

Hacking, I. 1983. *Representing and Intervening*. Cambridge: Cambridge University Press.

Hodgkin, A.L. and Huxley, A. F. (1952). ''A Quantitative Description of Membrane Current and its Application to Conduction and Excitation in Nerve,'' *Journal of Physiology*. 117: 500–44.

Kaplan, D.M. and Craver, C.F. (2011) "The Explanatory Force of Dynamical Models" *Philosophy of Science* 78: 601-627.

Mayo, D. G., 1996, *Error and the Growth of Experimental Knowledge*, Chicago: University of Chicago Press.

Rheinberger, H.-J. (1997). *Toward a History of Epistemic Things*. Stanford: Stanford University Press.

Sargent, R.-M. (2012) "Bacon to Banks: The Vision and the Realities of Pursuing Science for the Common Good." *Studies in History and Philosophy of Science* 43: 82–90.

Sargent, R.-M."Baconian Experimentalism: Comments on McMullin's History of the Philosophy of Science" *Philosophy of Science* 68 (2001), 311–17.

Tsai HC, Zhang F, Adamantidis A, Stuber GD, Bonci A, de Lecea L, Deisseroth K. (2009) Phasic Firing in Dopaminergic Neurons Is Sufficient for Behavioral Conditioning. *Science*. 324: 1080-84.

Tye KM, Prakash R, Kim SY, Fenno LE, Grosenick L, Zarabi H, Thompson KR, Gradinaru V, Ramakrishnan C, Deisseroth K. (2011) Amygdala circuitry mediating reversible and bidirectional control of anxiety. *Nature*. 471: 358-62.

Weber, M. (2005). *Philosophy of Experimental Biology.* Cambridge: Cambridge University Press.

Weber, M. (2012) Experimentation in Biology. *Stanford Encyclopedia of Philosophy* available online at http://plato.stanford.edu/entries/biology-experiment/.

Woodward, J. (2003). *Making Things Happen*. New York: Oxford University Press.

Zhang F, Wang LP, Boyden ES, Deisseroth K. (2006) Channelrhodopsin-2 and optical control of excitable cells. *Nature Methods*. 3: 785-92.