Robustness and Modularity

Trey Boone

ABSTRACT

Functional robustness refers to a system's ability to maintain a function in the face of perturbations to the causal structures that support performance of that function. Modularity, a crucial element of standard methods of causal inference and difference-making accounts of causation, refers to the independent manipulability of causal relationships within a system. Functional robustness appears to be at odds with modularity. If a function is maintained despite manipulation of some causal structure that supports that function, then the relationship between that structure and function fails to be manipulable independent of other causal relationships within the system. Contrary to this line of reasoning, I argue that functional robustness often attends feedback control, rather than failures of modularity. Feedback control poses its own challenges to causal explanation and inference, but those challenges do not undermine modularity—and indeed, modularity is crucial to grappling with them.

1. Introduction

Functional robustness refers to a system's ability to maintain some function despite substantial variation in relevant component parts and processes that sustain that function.[1,2] In other words, this sort of robustness refers to functional stability despite underlying causal heterogeneity. Over the past twenty years, robustness has come to play a prominent role in analyses of complex systems, particularly in genetics, systems biology, multiple levels of analysis in neuroscience, and network analyses of social systems.[3]

Standard methods of causal inference involve intervening on component parts and processes within a system and observing changes that occur in response – that is, concomitant variations. Functional robustness poses obvious and unique challenges to such methods of causal investigation. If complex systems respond to intervention in such a way as to flexibly maintain functions, then methods of concomitant variation may be unreliable in unearthing the causal factors that sustain those functions. In this vein, Sandra Mitchell (2008, 2009) has argued that systems that exhibit functional robustness fail to be modular, in the sense of having causal relationships that are independently disruptable. This argument is consequential because modularity is a crucial component of both difference-making accounts of causation, particularly interventionism, as well as the methods of causal investigation associated with those accounts (Hausman and Woodward 1999, Woodward 2003, Woodward 2008).

---

[1] Many terms have been introduced that have similar and related meanings to what I refer to as "functional robustness"—e.g., "biological robustness" (Kitano 2004), "distributed robustness" (Wagner 2005), "degeneracy (Edelman and Gally 2001, Whitacre 2010), "canalization" (Waddington 1942). See §II below for further discussion and clarification.

[2] The relevant sense of robustness here is ontological, in the sense of referring to a feature of systems in the world that are the targets of explanations and causal inference, thus contrasting with the epistemic notions of robustness analysis or robustness of models (cf. Weisberg 2006). See Boone (2018) for further discussion of the significance of this ontological notion of robustness, particularly as it relates to multiple realization.

[3] See Felix and Barkoulas (2015) for genetics, Marder and Goaillard (2006) for cellular and systems neuroscience, and Boldi et al. (2011) for social networks.

My aim in this paper is to explore the consequences of functional robustness for accounts of causation and causal inference. In contrast to Mitchell, I argue that modularity retains a crucial role in causal inference in systems that seem to exhibit functional robustness. Mitchell's argument relies on anomalous genetic knockout results, in which some gene knockout produces no apparent phenotypic effect. I show that closer inspection of genuine instances of functional robustness reveals that these phenomena are often indicative of feedback loops driving systems toward particular outcomes. That is, robustness does not attend failures of modularity but rather failures of acyclicity. Causal inference in cyclic systems presents its own set of challenges, but those challenges do not support general skepticism of difference-making accounts of causation or support a call for radically different methods of causal investigation. Indeed, I show that modularity is often a crucial component of unearthing the causal structure of cyclic systems.

I proceed as follows. In §I, I provide general background on interventionist accounts of causation and explicate the role of modularity in those accounts. In §II, I reconstruct Mitchell's argument that functional robustness in genetic knockout experiments is incompatible with modularity. I argue, to the contrary, that modularity plays a crucial methodological role in knockout experimentation. In §III, I argue that, rather than failures of modularity, functional robustness often attends feedback control—that is, the presence of cyclic causal structure. I conclude by exploring some of the challenges feedback control poses to theories of causation and causal inference and show how modularity may be crucial to overcoming those challenges.

2. Modularity and Interventionism

The interventionist account of causation interprets the meanings of causal claims as counterfactual statements about how the world would be different given certain changes, where those changes are understood as hypothetical idealized experimental manipulations. So, to say X causes Y is just to say

that, if other relevant features are held fixed, changing the value of X will result in a change in the value of Y. For instance, to say that smoking cigarettes causes lung cancer is just to say that, given appropriate controls (or conditioning on the other relevant variables like diet, exercise, and alcohol consumption), adjusting cigarette smoking will alter the risk of developing lung cancer.[4]

Interventionism is attractive as a theory of causation because it comports well with experimental practice and causal inference techniques—such as those based on directed acyclic graphs, or DAGs (Pearl 2000, Spirtes et al. 2000)—and it is well-suited to offering interpretations of causal claims in the special sciences. In particular, it has the advantage, over so-called process theories of causation of allowing straightforward interpretation of causal claims that do not involve obvious physical connection between events—say, the claim that inflation causes interest rates to increase.[5] Additionally, interventionism trades the standard framing of natural laws qua universal, exceptionless generalizations, perhaps with ceteris paribus clauses, for different, graded forms of relative stability of causal generalizations.

Central among these forms of stability are the notions of invariance and modularity.[6] To say that a generalization is invariant, as opposed to universal, is to allow that it may hold only for restricted ranges of variable values. Take, for instance, a snap mousetrap (see *figure 1* for reference). The causal operation of snap mousetraps is roughly as follows. The trap is set by pulling the hammer, against the force of the spring, in an arch over the platform; the holding bar is then placed over the hammer and held by the catch on the opposite side of the platform. When the catch is tripped, the holding bar is

[4] Note that the claim is not that it must be possible to perform this intervention. For instance, we can understand the claim that the moon's gravity causes oceanic tides as a counterfactual claim about how the tides would be affected if the moon didn't exist or were closer or further from Earth without actually being able to perform those interventions. (See Woodward, 2003: pp. 129-131 for discussion.)

[5] Cf. Dowe 2000, Salmon 1984.

[6] For extended discussion of invariance, see Woodward 1997, 2003 (Ch. 6); for modularity, see Hausman and Woodward 1999, Woodward 2003 (pp.327-342), and Woodward 2008.

released and the potential energy of the spring's force against the hammer is converted to kinetic energy, causing the hammer to slam onto the opposite side of the platform.

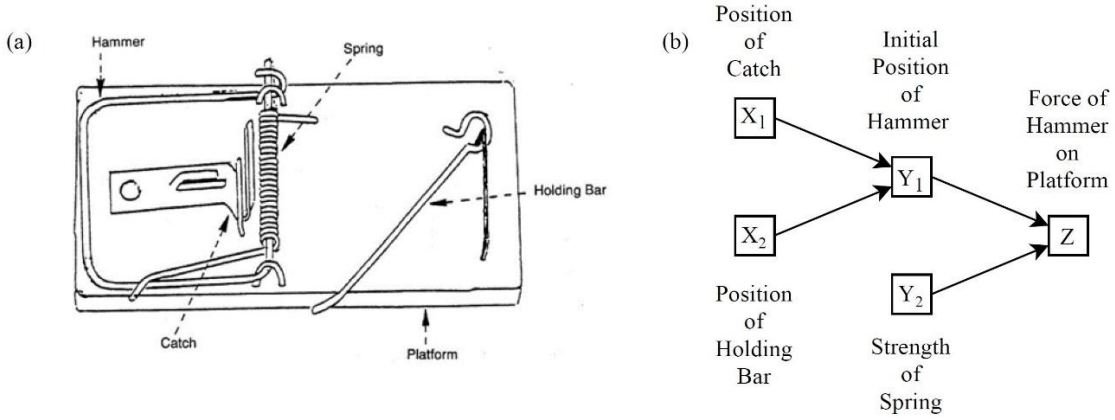

Figure 1: (a) Illustration of a snap mousetrap with (b) a causal diagram of relevant factors bearing on the force exerted by the hammer on the platform upon release of the catch.

This causal story does not hold universally—instead, it is invariant over a range of values that the relevant variables in this causal chain can take. For instance, if the spring is too strong relative to the sensitivity of the catch or the strength of the holding bar, the system will cease to function in accord with this set of causal relationships.

Modularity refers to a distinct form of stability that is an important, though controversial, component of the interventionist framework.[7] In short, modularity refers to the independent manipulability of causal relationships within a system.[8] Independent manipulability here means that changing one functional relationship within the system does not cause changes to other functional relationships within the same system. Woodward (2008) offers the following definition:

[7] In less cautious moments, interventionists have described modularity as intrinsic to the concept of causality (e.g., Hausman and Woodward 1999, p.550) and have attempted to use modularity in the service of deriving the Causal Markov Condition (Hausman and Woodward 1999, Hausman and Woodward 2004). See Cartwright 2002 and Steel 2006 for objections to the latter efforts.

[8] See also the notion of autonomy from Aldrich (1989), as incorporated by Pearl (1999).

> Modularity: A system of equations is modular iff (i) each equation is invariant under some range of interventions on its independent variables and (ii) for each equation, it is possible to intervene on the dependent variable in that equation in such a way that only the equation in which that dependent variable occurs is disrupted while the other equations in the system are left unchanged. (Woodward 2008, p.221)

Note that the first condition, which is a restatement of the invariance condition just discussed, holds within individual equations. By contrast, the second condition stipulates a relationship that holds between equations in the same system.

To gain a clearer sense of what modularity entails consider, again, the snap mousetrap. Suppose we are interested in understanding the relevant causes that determine the force the hammer exerts on the platform, as represented in the associated causal diagram (1b). That force is a function of the initial position of the hammer and the strength of the spring. The initial position of the hammer is a function of the positions of both the catch and the holding bar. These causal relationships can be represented by the following system of equations, where U and V are error terms (typically interpreted as representing omitted causes).

(1.1) $Y_1=aX_1+bX_2+U$

(1.2) $Z=cY_1+dY_2+V$

Modularity entails, in the first place, that both equations are invariant under some range of interventions on their independent (right-hand side) variables. In this case, that is just to say that equation (1.1) correctly describes how the initial position of the hammer will vary given changes in the positions of the catch and holding bar, and equation (1.2) correctly describes how the force of the hammer will vary given changes in either the initial position of the hammer or the strength of the spring. The second condition then entails that intervening on one of the dependent (left-hand side) variables will only disrupt the equation that describes that variable, leaving the other causal relationship(s) intact. Suppose, for instance, we exogenously set the value of Y_1—say, by pulling the

hammer back manually to an initial position independent of the locations of the catch and holding bar. Doing so would disrupt equation 1.1, breaking the relationship between the positions of the catch and holding bar and the initial position of the hammer, but it would leave equation 1.2 intact.

Now note that we could substitute equation (1.1) into equation (1.2) to yield (1.3), where $e=ad$, $f=bd$, and $W=dU+V$.

(1.3) $$Z= eX_1+fX_2+dY_2+W$$

Equation (1.3) yields the reduced-form equation, describing the force of the hammer on the platform solely as a function of the three exogenous variables—the positions of the catch and holding bar and the strength of the spring. On one hand, equations (1.1)–(1.2) and equation (1.3) are logically equivalent, in the sense that algebraic transformations can allow you to derive (1.1)–(1.2) from (1.3) and vice versa. However, a central tenet of interventionism, enforced by the requirement of modularity, maintains that the content of equations (1.1)–(1.2) and that of equation (1.3) are in fact distinct, entailing different counterfactual claims about the causal system in question. The initial position of the hammer, Y_1, does not occur as a separate dependent variable in equation (1.3). From an interventionist standpoint, this implies that there is no separate intervention that could be performed on the initial position of the hammer apart from adjusting the initial position of the catch and holding bar. Equations (1.1)–(1.2), by contrast, include Y_1 as a separate dependent variable. This inclusion entails a distinct set of counterfactuals describing what would happen under different experimental manipulations—namely, in this case, the intervention mentioned above where the initial position of the hammer is set exogenously.

Modularity entails that at most one of these sets of equations provides a correct description of the system. Other accounts of explanation and causation fail to make this distinction. For instance, positivist accounts of explanation that take explanation to consist in logical derivation do not have the resources to distinguish between the explanatory value of logically equivalent claims (or systems of

equations). In a separate vein, Cartwright (2003) commits to the idea that causal laws must be "numerically transitive", which she explains to mean, "causally correct equations remain causally correct if we substitute for any right-hand-side factor any function in its causes that is among nature's causal laws" (p. 206). Modularity thus plays a crucial role distinguishing interventionism from alternative accounts of explanation and causation.

Apart from this somewhat technical point about logically equivalent systems of equations, modularity also plays a more basic and intuitive role in the interventionist framework. Specifically, it ensures the stability and reliability of causal inferences made on the basis of interventions. Interventions effectively set the value of some putative causal variable. Causal relationships can then be inferred based on resultant changes in the values of other variables. The bedrock of this approach lies in the idea that interventions on causal variables will only induce changes in other variables that are actually causally related to that variable—that is, causal "descendants" of the variable intervened upon. This is precisely what modularity ensures: when you intervene upon a causal variable, you disrupt all and only the causal pathways into that variable, leaving the other causal relationships in the system intact. To the extent that modularity holds, interventions can reliably be used to infer causal pathways from the intervened upon variable to the variables that change as a result of the intervention. To the extent that modularity fails, there is no guarantee that changes brought about by interventions provide reliable information about the causal structure of the system, because those changes may be the result of incidental alterations of other relationships in the system that are not causally connected to the intervened upon variable.

Take, for instance, what it would mean for modularity to fail for the snap mousetrap. Suppose that intervening on the initial position of the hammer changed the relationship between the strength of the spring and the force of the hammer on the platform. There is no clear physical interpretation

of such a violation in this system.[9] What *is* clear is that such a situation would profoundly obfuscate causal investigation into the workings of the mousetrap: There would be no way to disentangle the effects of changes in the initial position of the hammer (on the force of the hammer on the platform) from the effects of changes in the strength of the spring. It is unclear how one would even approach causal decomposition of such a mousetrap or systematic investigation into the causal relationships underlying its operation.

3. Mitchell's Challenge: Robustness and Modularity

Sandra Mitchell (2008, 2009) has argued that complex systems, particularly those found in biology, often fail to be modular in much the sense just described.[10] Mitchell bases her arguments on cases of functional robustness in biology. She argues that modularity fails when robustness is achieved through some form of reorganization of causal structure in response to localized experimental manipulations. Her support for this conclusion draws on evidence of genetic robustness—specifically, experiments that show that elimination, or "knock out", of individual genes often results in no discernible phenotypic change in organisms. She concludes from such cases that interventionist theories of causal explanation and their attendant methods of causal inference are not viable in complex systems and that instead new methods of causal inference and new theories of explanation are needed.

As mentioned at the outset, functional robustness refers to a system's ability to maintain some function despite wide range of perturbations in the relevant causal structures that support performance of that function (Kitano 2004, Wagner 2005, Whiteacre 2010). This concept of robustness should be

[9] Note that the claim here must not be that there is some unknown mechanism connecting the initial position of the hammer with the strength of the spring, which just amounts to the claim that the causal diagram in figure 1b is incorrect, omitting a directed edge from Y_1 to Y_2.

[10] A parallel line of argument can be found in Wimsatt's (1972) critique of Simon's (1962) claim that complex systems tend to exhibit near-decomposability. Near decomposability approximates to the interventionist notion of modularity. Wimsatt argues that complex systems typically exhibit significant interactional complexity between the component parts and processes they are analysed in terms of. This interactional complexity in turn generates functional overlap that seriously confounds efforts to understand such systems in terms of hierarchical, modular organization.

distinguished from the more general concept of functional stability (*figure 2a*). For any function there is some normal range of variation in the underlying parts and processes supporting it, over which it may be stable. For instance, spark-ignition engines can combust a range of air-fuel mixture ratios (roughly, between 8:1 and 18:1) that are regulated by a carburetor. The function of the engine is thus stable over this range of ratios. But this form of stability is weaker than or at least distinct from that implied by the concept of robustness of interest to biologists. Robustness in biological systems is a subclass of functional stability that involves some form of reorganization of a system in order to maintain function in the face of perturbations.

A further distinction can be drawn between (at least) two ways functional robustness can occur: redundancy and reorganization.[11] Redundancy occurs when a system maintains some function via some structurally and functionally similar component that fills in for a perturbed component. For instance, suppose a spark-ignition engine were designed in such a way that if its carburetor fails, a backup carburetor automatically comes online, allowing the engine to continue running. In such a case, the redundant, structurally similar component plays the same causal role in the system, so changes to other causal relationships within the system need not occur. By contrast, functional robustness via reorganization occurs when some number of different components play a range of different causal roles in ways that compensate for effects of perturbations. Though there is no easy analog for engines, it might be something like a spark-ignition engine having the capacity to reorganize itself into a compression-ignition engine in response to spark plug failure, with some kind of internal refinery coming online to convert gasoline into diesel. In such a case, the function is preserved as a result of changes to several other causal relationships in the system, rather than those relationships simply

[11] There is, in fact, something of a terminological morass here, with several further distinctions that can be and have been drawn with several cross-cutting terms: See, for instance, canalization (Waddington 1942), degeneracy (Edelman and Gally 2001), and distributed robustness (Wagner 2005). Degeneracy is frequently used in a way that approximates to what I refer to here as reorganization. However, degeneracy is typically defined at a base level as "structurally different components performing the same function" (e.g., Edelman and Gally 2001). This definition of degeneracy edges closer to redundancy than reorganization (see below).

remaining intact while a functionally equivalent component is substituted into the system, as in cases of redundancy. It may sound ridiculous in the context of engines, but something like this seems to be rather common in certain biological systems.

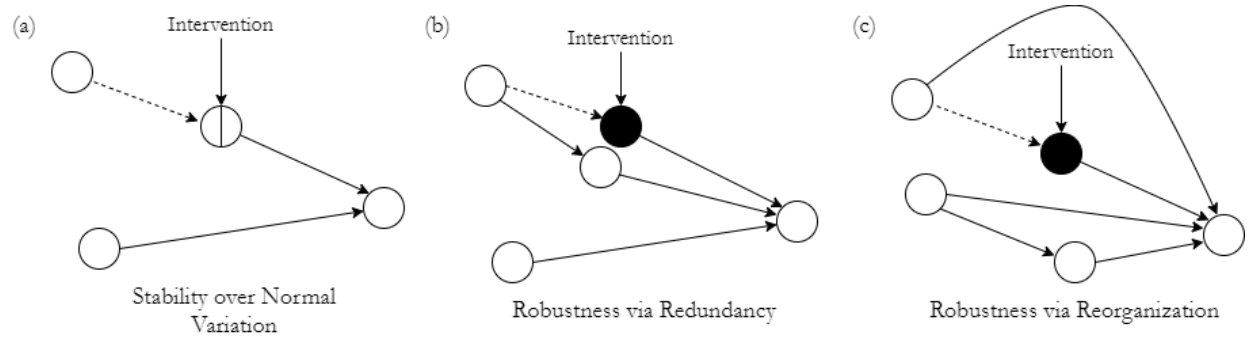

Figure 2: Examples of distinct causal structures that may give rise to functional stability.

Put a bit more precisely, we can think of the function of interest as some effect variable and the various causes supporting that function as causal variables with edges directed into that effect variable. Functional robustness via redundancy occurs when some new causal variable comes online in the face of an intervention that would otherwise disrupt the function (*figure 2b)*. In this case, an otherwise identical set of causal relationships maintains the function in the face of the perturbation. Functional robustness via reorganization, on the other hand, entails the formation of new causal relationships or changes to existing causal relationships in the face of such an intervention (*figure 2c)*. Prima facie, this form of robustness is in tension with modularity because an intervention brings about changes to other causal relationships, not directly related to the variable or causal relationship targeted by the intervention.

Accordingly, Mitchell's argument against modularity is rooted in this latter phenomenon of functional robustness via reorganization. Her argument is based on anomalous gene knockout experiments, which she argues indicate a form functional robustness in genetics. Mitchell argues that genetic robustness provides cases in which intervening on some particular causal relationship reconfigures other causal relationships in order to maintain some phenotype in the face of the

intervention. She concludes that genetic systems violate modularity and thus that new theories of causation are needed to account for the complex causation found in biological systems (see, in particular, Mitchell 2009, Ch4).

To understand Mitchell's argument, it is first necessary to provide some background. Genetic knockout is a targeted form of intervention that involves inserting artificial, nonfunctional DNA into the chromosomes of embryonic stem cells in vitro. The embryos are then transplanted into a female uterus of some particular model organism, most typically a mouse, and allowed to develop. The resulting mouse pups are heterozygous knockouts, which can then be bred to create homozygous knockouts. These techniques provide a precisely targeted means of manipulating genes to determine their roles in supporting different phenotypes. One might thus expect knockout experiments to be exemplars of interventionist accounts of causality—precisely controlled manipulations of dependent variables (genes) lead to changes in independent variables (phenotypes), and a causal link is inferred between those variables. And indeed, many knockout experiments proceed in just this way. Schofield et al. (2012) note in a review that of the ~25,000 mouse genes with protein sequence data, ~8,200 have identified phenotypes.[12] That is a significant amount of success in establishing causal links between genes and phenotypes, especially when considering that knockout techniques had only been in practice for 20 years when their review was published.

However, not all genetic knockouts have associated phenotypic changes. Roughly 15 percent of gene knockouts are developmentally lethal. Aside from those, and to the point crucial for Mitchell's argument, some sizable percentage of knockout experiments show little to no evidence of phenotypic change. These are the so-called "anomalous" knockout experiments, and Mitchell cites their proportion at roughly 30% of all knockout experiments (Mitchell 2008, p.700, Mitchell 2009, p.68).

[12] See also the massive databases available online collating information on the mouse genome (www.informatics.jax.org) and phenome (phenome.jax.org).

Other estimates fall in the 10-15% range (Barbaric et al. 2007), but the accuracy of this number is quite hard to gauge, for reasons that will be explored further below. Nonetheless, in some subset of these cases, Mitchell argues that systems exhibit a form of functional robustness via reorganization.

Her argument, offered primarily in her own words, proceeds as follows,

> [I]n up to 30% of double knockouts there is little or no evident phenotypic consequence of knocking out a gene. The cases where the knockout produces no substantive phenotypic difference may point to the dynamical plasticity of the genetic network. Robustness due to redundancy or degeneracy will make it difficult to make inferences about the normal causal structure from an intervention or perturbation of the system. (Mitchell 2008, p.700)

In the knockout case, redundancy simply involves a back-up copy of a gene being substituting for knocked out gene. When robustness is achieved through redundancy, the other causal relationships within the system need not change. Thus, these cases pose only a minimal challenge to modularity. Specifically, in cases of redundancy, an intervention on some variable (here, a gene) also changes exactly one additional causal relationship in the system, whereby a new, identical variable is introduced in the same causal relationship. This challenge can be overcome by identifying that redundant causal variable and intervening on it in tandem. Mitchell continues her reasoning in accord,

> The absence of a phenotypic change even when all redundant copies of a single genetic component are knocked out could indicate that the network itself has reorganized to compensate for the loss of the gene. If so, parts of the network that in the normal state would be described by one set of functional relationships *change* their interactions in response to the experimental intervention to produce a product similar to that of the unperturbed system. (Mitchell 2009, p.71, her emphasis)

> It appears that a degenerate or robust system where a genetic network reorganizes when some piece of it is knocked out is not independently disruptable. That is, one gene in the network functions as a causal contribution to the phenotypic effect under normal internal conditions, but functions in a different way when another part of the

> network is removed… Thus Woodward's condition of modularity is not met. (Mitchell 2009, p.77)

Mitchell thus argues that functional reorganization provides the best explanation for the results of some anomalous knockout experiments when redundancy can be ruled out. And it is these instances of robustness that pose a significant challenge to the interventionist notion of modularity. Her conclusions from here are sweeping: New concepts of causation are necessary and no uniform methodological prescriptions—for instance, regarding experimentation or how to conduct causal investigation—will be adequate for such complex causal systems.

While I am sympathetic to many aspects of Mitchell's argument, it moves too quickly to support such sweeping conclusions. Further, with respect to the more measured conclusion regarding the prospects for modularity in systems that exhibit robustness, I believe that she mislocates the challenge that robustness poses to theories of causation and causal inference. Spelling this out requires first taking a step back.

Results from anomalous knockout experiments are not straightforwardly attributable to functional robustness. Mitchell is well aware of this. In the setup for her argument (2009, p.68), she cites a famous quote from Mario Capecchi, who received the Nobel prize in 2007 for his research in the field: "I don't believe in complete redundancy. If we knock out a gene and don't see something, we're not looking correctly." She also offers a quote from the opposite end of the ideological spectrum from Robert Weinberg, a pioneer of research on the genetics of cancer: "The big surprise to date is that so many individual genes, each of which has been thought important, have been found to be nonessential for development."[13] Other researchers (e.g., Greenspan 2001, Edelman and Gally 2001), with whom Mitchell sides, offer a distinct possibility: phenotypes may be impervious to gene variation as a result of robustness in genetic networks. There are thus three broad possibilities that must be considered in

[13] Both quotes can be found in Travis 1992.

the interpretation of anomalous knockout experiments: (1) there is actually a phenotypic difference that simply has not yet been discovered, (2) the targeted DNA is nonessential, (3) the system exhibits functional robustness, either via redundancy or reorganization. Interpreting anomalous results to imply (3) is thus not uncontroversial.

I take (2) to be a conclusion of last resort, as it effectively ends inquiry despite the presence of other live options. However, (1) merits further scrutiny. Barbaric et al. (2007) provide an excellent review, detailing possible explanations of anomalous knockout results. They offer the following set of options.

> If inactivation of a gene does not lead to an observed abnormal phenotype, there are three possibilities: (i) the abnormal phenotype is present under the conditions currently being used but is yet to be discovered, (ii) the abnormal phenotype will only become evident under environmental conditions that have not yet been tested, or (iii) there is no abnormal phenotype. (Barbaric et al. 2007, p.92)

The first two options refer to distinct ways phenotypic differences can be obscured. First, (i) there may be a phenotypic difference that is present in experimental conditions but has not yet been discovered. One major methodological issue that creates this possibility is the lack of standardization in phenotyping protocols. Some large-scale efforts have been made to standardize protocols. In only its first few years of operation, one such effort, the German Mouse Clinic (Gailus-Durner et al. 2005), analysed more than 80 knockout lines and discovered previously uncharacterized phenotypes in 95% of those lines (Fuchs et al. 2009). The ongoing discovery of new phenotypes for different mutant strains should give pause to the idea that anomalous results should be taken at face value. However, as protocols become more standardized and more exhaustive, this should become less of an issue. It is difficult to estimate the impact this will have on the proliferation of anomalous knockout results, but it is likely to be significant.

Second, (ii) many phenotypes are only apparent in specific environmental contexts. Thus, it may be that anomalous knockout results are due to experimental conditions that do not provide the environmental conditions necessary for the phenotype to manifest, rather than reflecting an actual lack of phenotypic variation. For instance, Chen et al. (1997) discovered an exocrine gland dysfunction resulting from melanocortin 5 receptor (MC5-R) knockout in mice simply by chance due to an unrelated phenotyping assay designed to test stress response. They write,

> No readily visible phenotype was apparent in MC5-RKO mice… Appearance, behavior, growth, muscle mass, adipose mass, reproduction, basal and stress-induced corticosterone, glucose, and insulin levels in these animals were indistinguishable from wild-type littermates. More subtle physiological phenotypes of the knockout were studied by examination of responses to exogenous melanocortin peptides in biological assays… None of these assays produced identifiable differences between the wild-type and knockout animals. During a stress-induced analgesia assay in which the mice are made to swim for 3 min to activate the hypothalamic-pituitary-adrenal axis, it was observed that the knockout animals remained wet for a longer period of time than littermate controls. This effect was then identified to result from nearly double the water retention in the coat of the MC5-RKO, resulting in severe thermoregulatory defects in the animal as well. (Chen et al. 1997, p.794)

In other words, MC5-R knockout was previously characterized as an anomalous knockout gene, showing no phenotypic variation despite a wide range of phenotyping assays. Fortuitously, during an assay to test for abnormal stress response in which mice are forced to swim, Chen et al. noticed an unrelated abnormality—the coats of MC5-R knockout mice took significantly longer to dry than did the coats of wild-type mice. It turns out this abnormality reflects an exocrine gland dysfunction due to the MC5-R knockout.

Again, it is difficult to estimate what proportion of anomalous knockout experiments will ultimately be accounted for in this way. My point in raising these possibilities is not to come up with an estimate but rather to reflect on appropriate methodology. Mitchell argues, on the basis of

anomalous knockout results, that new methods of causal inference and new accounts of causal explanation are needed. In particular, she argues that when faced with anomalous knockout results, researchers should not assume modularity but should instead assume that some more complex causal structure is in play. On one hand, even given the preceding discussion, it is probably reasonable to expect that at least some anomalous knockout experiments are indicative of more complex causal structures. However, it is clearly premature to infer from this that standard forms of experimental design, captured by the interventionist framework and assumptions of modularity, should be abandoned. As a methodological principle, modularity pushes researchers toward discovery of new phenotypes rather than accepting anomalous knockout results at face value.

4. Answering Mitchell's Challenge: Robustness and Cyclicity

I have just argued that, from a methodological perspective, Mitchell is wrong that modularity should be abandoned as a principle guiding causal investigation in genetics. However, my discussion does not support the conclusion that all anomalous knockout results are due to undiscovered phenotypes. This thus leaves open the (likely) possibility that some results will be best explained by reference to functional robustness. So, my argument that modularity is the right methodological principle to retain in genetics is beside the point of this main issue: what are the implications of functional robustness for modularity?

Consider the following, from a discussion of robustness from O'Leary (2018): "If an insect loses a leg, it may or may not lose the ability to walk. But the biomechanical relationships between the remaining legs will be fundamentally altered" (O'Leary 2018, p.182). Autotomy (self-amputation) of appendages, in fact, occurs in many taxa, including vertebrates, echinoderms, crustaceans, and arachnids (Wrinn and Uetz 2007). Spiders are particularly interesting cases because their legs play vital roles not only in locomotion but also as sensory organs—the tiny hairs on spider legs are capable of

detecting minuscule vibrations. Yet they can lose two to three legs, and often do as a result of autotomy, and nonetheless retain their abilities to walk and detect prey.

This case gets to the heart of the issue with the relationship between modularity and functional robustness. On one hand, appendages are generally regarded as exemplars of modularity in biology (e.g., Williams and Nagy 2001). Granted, the notion of modularity in biology is distinct from, though it bears similarity to, the concept of modularity at play in interventionism.[14] And appendages do seem to satisfy modularity in the interventionist sense: they provide isolated causal contributions to the capacities they are involved in. However, in certain organisms, like spiders, removal of appendages leads to changes in other causal relationships within the system—particularly, the biomechanical and sensory relationships between the remaining legs. Taken together, these observations seem to suggest that appendages at once satisfy and violate modularity.

The tension here can be resolved by considering more carefully how the changes in other causal relationships within the system occur. In the immediate aftermath of leg removal, a spider's ability to walk is actually seriously impaired. This is because the biomechanical relationships between the remaining legs do not adjust automatically. The remaining legs are still controlled by motor patterns predicated on the organism having all eight of its legs. Thus, it is more accurate to say that the biomechanical relationships adjust over time in response to feedback that enables recovery of function. This is significant because it shows that a system can exhibit functional robustness while

[14] In the interventionist framework, modularity denotes independent manipulability of causal relationships, whereas in biology (among other fields like cognitive science), modularity refers to organization of a system into discrete subsystems. Neglecting some nuance, modularity in the biological sense implies modularity in the interventionist sense. If a system is organized into discrete subsystems, it should be possible to intervene on those subsystems without altering the causal relationships other subsystems are involved in. (Removing a few fingers from your dominant hand will affect your ability to type or grasp an object but will generally leave your capacity to run, metabolize complex carbohydrates, etc. intact.) The converse, however, need not hold: Modularity in the interventionist sense does not imply that a system be organized into discrete subsystems. The lack of (straightforward) decomposition of a system into subsystems does preclude that system from exhibiting independently manipulable causal relationships. An important caveat is that causal decomposition of a system into subsystems often provides a valuable (in some cases, plausibly, indispensable) guide to variable selection, a critical first step of interventionist analysis into the causal operations of a system.

nonetheless satisfying modularity. The components of a causal system can figure in isolable causal relationships, which can be assessed on short timescales immediately following an intervention that disrupts one of those causal relationships. And then, on longer timescales, other causal relationships may change as a result of feedback within the system in response to that intervention, and those changes may enable recovery of function—that is, functional robustness.

The upshot is that in systems where functional robustness is achieved via feedback control, there is a temporal gap between the intervention and the recovery of function.[15] This gap creates room for a productive notion of modularity to play a role both as a criterion on adequate causal explanations and as a principle guiding causal investigation. Indeed, modularity is often critical to understanding how feedback control enables robustness. This can be seen again in the example of the autotomous spider. To understand how the biomechanical relationships between the remaining legs need to change to preserve locomotion, it is necessary to first understand the modular causal contributions of the destroyed leg. That leg will have played different roles in different motor behaviours—for instance, prey capture or web navigation. Characterizing those different roles—the isolated causal contributions of the leg—is necessary to understand how those motor behaviours can be maintained in the absence of the leg. For instance, whether it is a front, back, or middle leg that is missing will have consequences for the new motor patterns that need to be learned, and hence new causal relationships that need to be adopted, to maintain those behaviours.

These considerations are not limited to this example. Feedback control seems to be one of the primary mechanisms responsible for enabling functional robustness across a range of systems, particularly in both neural and genetic networks. Mitchell is, of course, aware of this. However, she

[15] By feedback control here, I simply mean presence of cyclic causal structure, with that structure maintaining some stable output (akin to a centrifugal governor or thermostat). I do not intend to conjure any noncausal (for instance, strong teleological) notions that may otherwise be associated with the concept. For further discussion of a strictly causal notion of control mechanisms, see Bechtel (2018), and for further discussion of feedback mechanisms in biological systems, see Bechtel and Abrahamsen (2013).

concludes that feedback is inconsistent with modularity. She writes, "[a]ny physical system with complex feedback mechanisms will be one in which we can expect modularity to fail. But we should not conclude that such systems don't involve true component causes" (Mitchell 2009, p.82). Her idea seems to be that feedback involves alteration of other causal relationships within the system, so systems that exhibit feedback must not be modular. But this is not right: as I have just argued, feedback control and modularity are not incompatible, and to the contrary, modularity is an important component of adequate explanations of and investigations into systems that achieve robustness via feedback control. So, while it may be true that feedback control changes other causal relationships in the system, there is no independent reason to suppose those changes are not causally downstream from the intervention that triggers the feedback response—and this is what is required for modularity to be satisfied.

This point can be developed further by comparing the results of intervention techniques that probe robustness (and the mechanisms supporting it) on different timescales. Cellular neuroscience provides illuminating cases. The functions of individual neurons are generally characterized in terms of their response (input-output) properties. A neuron's response properties are determined by the combined effect of several currents that result from proteins (ion channels) allowing ions to permeate the cell's membrane. Whereas a neuron may live for decades, these proteins have shorter lifespans, typically turning over on the order of hours, days, or weeks.[16] As a result, the conductances that determine a neuron's response properties also vary over time. This variance raises questions regarding how stable those response properties are, and, to the extent that they are stable, how neurons maintain that stability. A range of studies has shown, for a variety of cells, that ion channel densities can in fact vary severalfold between cells that nonetheless have effectively identical response properties (Golowasch et al. 2002, Schulz et al. 2006, Ransdell et al. 2013). This research suggests that neuronal

[16] See, e.g., Hanwell et al. 2002; and Marder and Goaillard 2006, for relevant review.

response properties are highly robust and has motivated further research into the mechanisms that support that robustness.

Burst firing in Purkinje cells is a prime example of this sort of robustness of response properties to variation in the underlying conductances. Purkinje cells are a type of neuron found in the cerebellum that are relatively large with sprawling dendritic trees that receive tens of thousands of inputs. They play key roles in motor behaviours and, particularly, in motor learning. Climbing fibers, projections from the inferior olivary nucleus in the medulla oblongata, provide a strong source of excitatory input to Purkinje cells. According to a longstanding model, the inputs from climbing fibers convey a motor error signal that is integral to motor control and motor learning.[17] Depolarizing stimulation from climbing fibers evokes a stereotyped all-or none burst firing pattern from Purkinje cells. This burst firing is a crucial function that enables plasticity in adjacent circuits, in the form of both long-term potentiation and long-term depression. Purkinje cell burst firing is the result of a relatively small net inward current after an initial action potential. And Purkinje cells are among those that have been shown to have similar electrophysiological profiles despite massive variance in surface conductances. That is to say, Purkinje cells with very similar burst profiles exhibit substantial variation in underlying conductances of different types of ions. This similarity in burst profiles is surprising because the bursts are triggered by a small net influx of currents relative to the variability in any particular conductance.

In a remarkable study, Swensen and Bean (2005) investigated the mechanisms that support robustness of burst firing in Purkinje cells. Their experiments illustrate some of the unique challenges robustness poses to functional inference and how those challenges can be overcome (using techniques that assume modularity). Swensen and Bean used two different intervention techniques to investigate the robustness of Purkinje cell burst firing and the mechanisms that support that robustness. The first

[17] This is the Marr-Albus-Ito model, one of the most influential computational-cum-experimental models in neuroscience (Marr 1969, Albus 1971, Ito et al. 1982, Ito and Kano 1982, Ito 1989). For relevant review of the model, see Strata 2009.

technique involved pharmacological blockade of sodium conductance with TTX, which is transient and occurs on very short timescales. The relevance of the short duration of the intervention is that it rules out second messenger processes occurring within the cell that may alter ion channel expression on the cell membrane. The second intervention involved a genetic knockout similar to those discussed in the previous section. In particular, the mice used were knockouts for the $Na_v1.6$ gene, which codes for the protein that constitutes a particular subtype of voltage-gated sodium channel.

Swensen and Bean conducted additional analyses to determine the mechanisms responsible for robustness in each condition, and they found evidence of distinct mechanisms that operate on two different timescales. In their pharmacological intervention, Swensen and Bean first recorded the action potential waveforms for each Purkinje cell in current-clamp, and then switched to voltage-clamp and played the action potential waveform back to the cell.[18] This technique allowed them to observe any compensatory changes in other conductances that facilitated continued burst firing. They discovered that the acute decrease in sodium conductance due to TTX produced a decrease in the height of the action potential and a hyperpolarizing shift in postspike membrane potential. These changes in action potential waveform effected a small decrease in calcium conductance and, more notably, significant reductions in potassium current—both voltage-dependent and calcium-activated potassium current saw reductions ~50% in comparison to their levels with no TTX block. Sodium exists in high concentrations outside the cell, whereas potassium exists in high concentrations inside the cell. Thus, when TTX-sensitive sodium channels open, sodium rushes into the cell, depolarizing the membrane; and when voltage-activated and calcium-activated potassium channels open, potassium rushes out of the cell, hyperpolarizing the membrane.

[18] Current-clamp is a technique in which experimenters record from cells while injecting stimuli (usually 1ms depolarizing electrical currents) and recording the cell's response (so the cell's membrane potential is the dependent variable). Voltage-clamp, by contrast, is a technique in which the experimenter controls the cell's membrane potential and observes how aspects of the cell's electrophysiology respond (so the cell's membrane potential is the independent variable).

Both voltage-dependent and calcium-activated potassium channels are sensitive to changes in the action-potential waveform. The primary driver of decrease in voltage-dependent potassium is the hyperpolarizing shift in interspike potential. Swensen and Bean hypothesized that this decrease is due to the voltage sensitivity of deactivation of the Kv3-type potassium channel in the observed voltage range. They attribute the decrease in calcium-activated potassium current to three main factors: (1) the decrease in calcium entering the cell (despite the changes in calcium current being relatively small), and the hyperpolarizing postspike shift serves to (2) promote deactivation of BK-type potassium channels, and (3) decrease the potassium driving force. The causal dynamics here are obviously fairly complex, which on one hand makes it all the more remarkable that the system achieves such stability. On the other hand, the complex causal dynamics combine to produce a relatively simple causal structure, as shown in *figure 3* below.

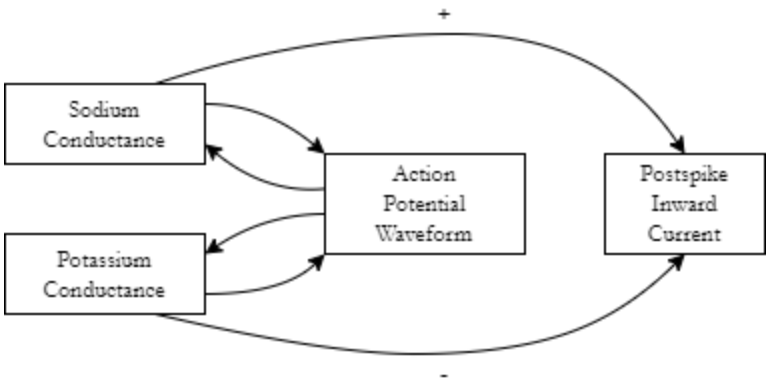

Figure 3: Simplified causal diagram of the acute mechanism supporting Purkinje cell burst firing robustness.

Intervening on sodium conductance disrupts the main inward current during the interspike interval. However, it also affects the action potential waveform. This leads to a form of acute feedback control, operating analogously to a flywheel governor, whereby potassium conductance, which is the main source of outward current flow in the interspike interval (Swensen and Bean 2003), also decreases as

a result of those changes in the action potential waveform. The net effect is that the postspike inward current is stable and burst firing persists.

Note that this case has all the features that make functional robustness seem, prima facie, incompatible with modularity. It appears as though burst firing is robust to variation in sodium conductance because other causal relationships within the system (particularly between potassium conductance and postspike inward current) change in tandem with the TTX intervention to preserve burst firing. (Note the symmetry here with the cases of genetic robustness discussed in the previous section and the case of the imagined "modularity-violating mousetrap" at the end of §I.) What Swensen and Bean's study elegantly illustrates is that there is a specific mechanism that explains how those apparent changes in other causal relationships are causally linked to the intervention. Intervening on sodium conductance does not change the relationship between potassium conductance and postspike current; rather, it alters potassium conductance via an acute feedback mechanism that compensates for changes in sodium conductance via changes to the action potential waveform. Recall that modularity requires that interventions on causal variables not give rise to changes in other (causally disconnected) relationships in the system. In other words, modularity requires that the changes brought about by an intervention are causally "downstream" from the intervened upon variable. Swensen and Bean's TTX intervention shows how functional robustness can occur on short timescales strictly via acute feedback control mechanisms causally downstream from sodium conductance.

Interestingly, in their second condition, using genetic knockout instead of pharmacological blockade, Swensen and Bean found evidence of a completely different mechanism operating over much longer timescales. Rather than the acute mechanism operating via changes in action potential dynamics, this mechanism involves compensating for the lack of sodium conductance through changes in the expression of other ion channels on the cell membrane—a process known as activity-

dependent synaptic plasticity. They found that the main difference driving robustness of bursting in $Na_v1.6$ knockouts was an increase in calcium conductance. This is surprising for two reasons. First, as noted above, potassium conductance was the main current that changed in response to acute reductions in sodium conductance in the TTX experiment. One might thus expect that in the knockout case the changes in channel expression would largely consist in a decrease in potassium channels proportional to the decrease in sodium channel expression. Yet this is not what Swensen and Bean found.

Second, one of the main drivers of *increased* potassium conductance in the TTX experiment was calcium-activated potassium current. Calcium influx can have, on net, either hyperpolarizing or depolarizing effects on membrane potential. While calcium entering the cell has a depolarizing effect, the hyperpolarizing response of calcium-activated potassium current is capable of overwhelming those depolarizing effects. However, Swensen and Bean found that this does not occur in the case of $Na_v1.6$ knockouts. This indicates plasticity in the coupling between calcium influx and calcium-activated potassium current. Thus, there are two main changes that account for robustness of burst firing in the knockout case—increased calcium ion channel expression on the cell membrane and a decrease in the responsiveness of calcium-activated potassium current to the presence of calcium within the cell (*figure 4*). Both these changes are likely the result of activity-dependent feedback control mechanisms driving the cells toward burst firing in response to the decrease in sodium conductance due to the knockout. Though the precise mechanisms responsible for both these forms of plasticity have yet to be characterized, they are likely tied to changes in the amount of calcium entering the cell due to the presence or absence of burst firing.[19]

[19] Burst firing triggers distinctive patterns of calcium influx and calcium is often implicated in activity dependent plasticity due to its roles in gene expression (see, among others, Grasselli et al. 2016, Puri 2020, Zucker 1999).

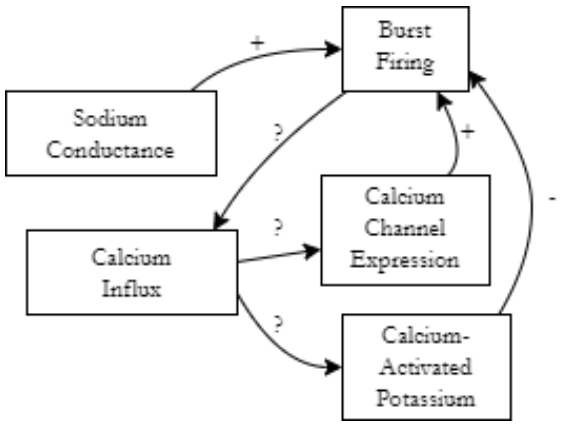

Figure 4: Simplified causal diagram of the chronic mechanism supporting Purkinje cell burst firing robustness. Question marks indicate hypothesized but unconfirmed causal relationships.

This knockout case reinforces two aspects of my arguments to this point. First, note that the causal relationships illuminated by the $Na_v1.6$ knockout satisfy modularity. Similar to the case of the TTX blockade, the activity-dependent plasticity both in calcium conductance and in the relationship between calcium influx and calcium-activated potassium are causally downstream from the $Na_v1.6$ knockout. That is, there is no reason to suppose these changes in other causal relationships in the system are anything other than effects of causal processes initiated by the intervention; indeed, precisely the opposite seems to be the case.

Second, Swensen and Bean's knockout experiment reinforces an important aspect of my argument from §II. Note that if we zoom out to the level of the whole organism, the phenotypic change associated with the Na1.6 gene knockout in cerebellar function would likely be obscured. If the function of the $Na_v1.6$ in that context were not already well-characterized prior to the experiment, the robustness of burst firing in the face of $Na_v1.6$ knockout could just as easily be construed as an anomalous knockout result—that is, a case where there is no apparent phenotypic effect associated with the gene knockout due to the functional robustness of Purkinje cells. If we were to follow Mitchell's prescriptions about causal investigation and explanation, we would thus misconstrue the effect of $Na_v1.6$ knockout and miss the means by which robustness is actually achieved. In other

words, this case points to another way in which anomalous knockout results can be misleading. Here phenotypic robustness occurs at an intermediate scale within an organism, and hence if researchers are not focused in on the appropriate scale, the effect of gene deletion could easily be missed.

This last point suggests important implications functional robustness may have for a particular aspect of causal discovery in the context of graphical modeling—namely, the Causal Faithfulness Condition (Spirtes et al. 2000, Zhang and Spirtes 2008, Forster et al. 2018, Weinberger 2018).[20] Causal discovery involves inference between the causal structure of a system and the joint probability distribution over all variables in that system. Two principles have played prominent roles in guiding the development of causal discovery algorithms – that is, algorithms designed to infer causal structure from probabilistic data. First, the Causal Markov Condition (CMC), states that, conditional on its direct causes, each variable within a system is independent of all other variables (other than its effects). Second, the Causal Faithfulness Condition (CFC), the converse of the CMC, states that the only independencies present in the joint distribution are those implied by the CMC. Failures of the CFC occur when two variables that are causally related are, nonetheless, probabilistically independent (conditional on the direct causes of the upstream causal variable).

In acyclic causal systems, failures of faithfulness occur when alternate causal pathways precisely counterbalance changes in some causal variable such that some effect variable does not change in response to variation in that causal variable (Zhang and Spirtes 2008, Andersen 2013, Weinberger 2018). For instance, grant that fish consumption increases both omega-3 fatty acid and mercury concentrations in the body. Now suppose omega-3s are inversely related to risk of developing dementia and that this inverse correlation is precisely counterbalanced by a positive relationship between presence of mercury and risk of developing dementia. Such a case would present a violation

[20] I thank an anonymous reviewer for pressing me to highlight this connection between robustness and causal faithfulness.

of the CFC: Risk of developing dementia will be independent of fish consumption despite the presence of causal pathways linking the two in the true causal structure.

Violations of faithfulness pose challenges to causal discovery due to this presence of independencies that obscure underlying causal relationships. Note that the challenge posed by anomalous gene knockout results is importantly similar to the challenge posed by violations of the CFC. With anomalous knockout results, it is unclear whether the independence between some gene and any detectable phenotypic change is due to the absence of a causal relationship or instead to some other factor, which in turn explains the apparent absence of phenotypic change. Note, further, that the above example regarding fish consumption relevantly parallels cases of functional robustness. The effect (risk of developing dementia) is "robust" to variations in some cause (fish consumption) due to precisely counterbalanced causal pathways (through mercury and omega-3 concentrations).

The most common and influential line of justification for the CFC maintains that violations should be vanishingly unlikely.[21] This is reasonably intuitive in the example of fish consumption because it is incredibly unlikely that the negative relationship between omega-3s and dementia risk should be the exact inverse of the positive relationship between mercury and dementia risk. However, the parallels between cases of functional robustness and violations of the CFC suggest that examples of functional robustness may provide cases where these sorts of "conspiracies of the evidence" to obscure causal relationships are not only *not* vanishingly unlikely but may, in fact, be overwhelmingly likely.[22]

There are, however, several important caveats here. First, the CFC was developed, and its implications for causal discovery have been explored, primarily in the context of acyclic causal models (particularly directed acyclic graphs – DAGs). The implications of violations of the CFC in cyclic causal models is more complex and less well-explored, and as the examples and arguments above

[21] More precisely, relative to all possible parameterizations, those that entail violations of the CFC will be Legesgue measure zero. See Spirtes, Glymour, and Scheines (2000) Theorem 3.2 pp.41-42 and pp.383-384 for the proof.
[22] For arguments to similar conclusions, see Cartwright (1999, 2007), Hoover (2001), and Andersen (2013).

show robustness often attends feedback control, which entails cyclic causal structure. Second, many cases of functional robustness are not contexts where causal discovery is at issue. In the Purkinje cell case, the causal structure of the system was well characterized prior to Swensen and Bean's work illuminating the mechanisms by which robustness is achieved. Indeed, the robustness of Purkinje cell burst firing to variation in sodium conductance is surprising precisely because sodium conductance is known to be *the* major contributor of post-spike inward current supporting burst firing. In such cases, there is thus no issue of inferring causal structure from joint probability distributions. Anomalous gene knockout experiments and other contexts in neuroscience, like effective connectivity in fMRI research, may provide cases where underlying causal structure is not already well known and functional robustness is likely to be present. In these cases, the parallel challenges to causal inference posed by the CFC and functional robustness may point to serious challenges to causal discovery and raise questions about the reliability of discovery algorithms that rely on the CFC.

Third, and perhaps most significantly, not all instances of functional robustness constitute violations of the CFC. Violations of the CFC entail that two causally related variables be probabilistically independent, provided appropriate conditionalizations. But in cases of functional robustness, it need not be the case, and often is not the case, that the function is probabilistically independent of the relevant causal variable(s). Instead, functional robustness typically occurs over some range of values for the relevant causal variable(s). Outside of that range, the function may be disrupted. As a result, the function and relevant causal variable(s) will not, in fact, be independent (assuming the joint distribution includes values beyond the range for which robustness holds). Consider the Purkinje cell case as an example. It is not the case that Purkinje cell burst firing persists for all values of sodium conductance. Rather, the function is stable for a (remarkable) range of sodium conductance – up to a ~50% decrease. When perturbed beyond this range, a dependence between sodium conductance and burst firing will emerge and reveal the underlying causal relationship (if this

were a context of causal discovery where the causal structure was not already characterized). There are certainly interesting and important challenges that arise with inferring causal structure in these sorts of cases—for instance, if the function is robust across all values of some independent variable (as may be the case with anomalous gene knock-out results), or if there are reasons to believe some data set is restricted to only values where robustness holds. Full exploration of these issues is beyond the purview of this paper, but these brief remarks hint at the rich potential for future work exploring the relationship between robustness and the CFC.

5. Conclusion

The preceding arguments show that functional robustness is compatible with modularity and furthermore that modularity is often a crucial methodological assumption guiding research into the feedback control mechanisms that enable robustness. However, it would be a mistake to conclude that feedback control and functional robustness do not carry significant consequences that impel revisions to the interventionist framework, particularly with respect to the notions of modularity, invariance, and the notion of intervention itself.

Different experimental interventions operate on different timescales. In the case of Purkinje cells, the transient, short-timescale intervention of pharmacological blockade has a totally different effect on the system than the persistent, longer-timescale intervention of the gene knockout despite both techniques targeting the same causal variable—namely, sodium conductance. This observation suggests that the concept of an intervention may require some inherent timescale-sensitivity—for instance, with respect to issues of variable choice and how to properly characterize controls. Furthermore, intervention techniques that target different timescales are, in turn, capable of illuminating distinct sets of causal relationships that also operate on different timescales, as evidenced by the contrast between the two mechanisms enabling robustness in Purkinje cell burst firing. Such

results suggest that the invariance of causal relationships may be constrained, not only by the values of the variables involved in those relationships, but also by the timescales on which those relationships operate.

Finally, for the notion of modularity a parallel issue arises wherein a particular set of causal relationships within a system may be modular only on some specific timescale. Feedback control that occurs in the wake of an intervention may alter causal relationships in distal parts of a system. On longer timescales than those on which feedback control operates, systems may thus appear to violate modularity despite the effects of feedback control being causal descendants of the intervened upon variable. Though I have argued that this set of conditions need not pose problems in principle for the concept of modularity, it does raise considerable practical challenges for causal inference techniques based on modularity.

First, it provides reason to believe that systems that exhibit functional robustness may be unsuitable for analysis with causal discovery algorithms that rely on the assumption of acyclicity. While it may be informative to treat such systems as acyclic on particular timescales, methods that assume acyclicity may fail to capture the full dynamics of systems that involve feedback control, especially when researchers are interested in understanding the behaviour of that system on much longer timescales than those on which relevant feedback control mechanisms operate. Second and relatedly, there is a measurement problem that arises here: if systems achieve robustness via feedback control operating on short timescales relative to the temporal resolution of some measurement instrument, that system may well appear to violate modularity. As a result, it will be impractical to employ causal inference techniques that assume modularity on data acquired in such conditions. In my view, these considerations should not be construed as posing fundamental problems for the interventionist framework so much as illuminating a set of issues where further work is needed.

Acknowledgements

Thanks to Edouard Machery, Mazviita Chirimuuta, Bob Batterman, Sandy Mitchell, Joseph McCaffrey, Zina Ward, Lauren Ross, Felipe De Brigard, Ashley Jardina, and especially Jim Woodward for helpful discussions and feedback on previous drafts. I am also grateful for the careful readings and feedback from two anonymous reviewers and one editor of this journal, which significantly improved the final version of this paper.

Imagination and Modal Cognition Lab
Duke University
Durham, NC USA
treyboone@gmail.com

References

Albus, James S. 1971. "A Theory of Cerebellar Function." *Mathematical Biosciences* **10** (1): 25–61. https://doi.org/10.1016/0025-5564(71)90051-4.

Aldrich, John. 1989. "Autonomy." *Oxford Economic Papers* **41** (1): 15–34.

Andersen, H. (2013). "When to expect violations of causal faithfulness and why it matters." *Philosophy of Science*, **80**(5), 672–683.

Barbaric, Ivana, Gaynor Miller, and T. Neil Dear. 2007. "Appearances Can Be Deceiving: Phenotypes of Knockout Mice." *Briefings in Functional Genomics & Proteomics* **6** (2): 91–103. https://doi.org/10.1093/bfgp/elm008.

Bechtel, W., & Abrahamsen, A. A. (2013). Thinking dynamically about biological mechanisms: Networks of coupled oscillators. Foundations of Science, **18**(4), 707-723.

Bechtel, W. (2018). The importance of constraints and control in biological mechanisms: Insights from cancer research. Philosophy of Science, **85**(4), 573-593.

Boldi, Paolo, Marco Rosa, and Sebastiano Vigna. 2011. "Robustness of Social Networks: Comparative Results Based on Distance Distributions." In *Social Informatics*, edited by Anwitaman Datta, Stuart Shulman, Baihua Zheng, Shou-De Lin, Aixin Sun, and Ee-Peng Lim, 8–21. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-642-24704-0_7.

Boone, W. (2018). "Multiple Realization and Robustness." In *Biological Robustness: Emerging Perspectives from within the Life Sciences*, edited by Marta Bertolaso, Silvia Caianiello, and Emanuele Serrelli, 75–94. History, Philosophy and Theory of the Life Sciences. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-01198-7_4

Cartwright, Nancy. 1999. *The Dappled World: A Study of The Boundaries of Science*. Cambridge: Cambridge University Press.

———. 2002. "Against Modularity, the Causal Markov Condition, and Any Link between the Two: Comments on Hausman and Woodward." *The British Journal for the Philosophy of Science* **53** (3): 411–53.

———. 2003. "Two Theorems on Invariance and Causality." *Philosophy of Science* **70** (1): 203–24. https://doi.org/10.1086/367876.

———. 2007. *Hunting Causes and Using Them: Approaches in Philosophy and Cconomics*. Cambridge: Cambridge University Press.

Chen, W., M. A. Kelly, X. Opitz-Araya, R. E. Thomas, M. J. Low, and R. D. Cone. 1997. "Exocrine Gland Dysfunction in MC5-R-Deficient Mice: Evidence for Coordinated Regulation of Exocrine Gland Function by Melanocortin Peptides." *Cell* **91** (6): 789–98. https://doi.org/10.1016/s0092-8674(00)80467-5.

Dowe, Phil. 2000. *Physical Causation*. Cambridge University Press.

Edelman, Gerald M., and Joseph A. Gally. 2001. "Degeneracy and Complexity in Biological Systems." *Proceedings of the National Academy of Sciences* **98** (24): 13763–68. https://doi.org/10.1073/pnas.231499798.

Félix, Marie-Anne, and Michalis Barkoulas. 2015. "Pervasive Robustness in Biological Systems." *Nature Reviews Genetics* **16** (8): 483–96. https://doi.org/10.1038/nrg3949.

Forster, M., Raskutti, G., Stern, R., & Weinberger, N. (2018). "The frugal inference of causal relations." *British Journal for the Philosophy of Science* **69** (3): 821-848.

Fuchs, H., V. Gailus-Durner, T. Adler, J. A. Aguilar Pimentel, L. Becker, I. Bolle, M. Brielmeier, et al. 2009. "The German Mouse Clinic: A Platform for Systemic Phenotype Analysis of Mouse Models." *Current Pharmaceutical Biotechnology* **10** (2): 236–43. https://doi.org/10.2174/138920109787315051.

Gailus-Durner, Valérie, Helmut Fuchs, Lore Becker, Ines Bolle, Markus Brielmeier, Julia Calzada-Wack, Ralf Elvert, et al. 2005. "Introducing the German Mouse Clinic: Open Access Platform for Standardized Phenotyping." *Nature Methods* **2** (6): 403–4. https://doi.org/10.1038/nmeth0605-403.

Glymour, C., Zhang, K., and Spirtes, P. (2019). "Review of Causal Discovery Methods Based on Graphical Models." *Frontiers in Genetics*, **10**: 524. https://doi.org/10.3389/fgene.2019.00524.

Golowasch, Jorge, Mark S. Goldman, L. F. Abbott, and Eve Marder. 2002. "Failure of Averaging in the Construction of a Conductance-Based Neuron Model." *Journal of Neurophysiology* **87** (2): 1129–1131.

Grasselli, Giorgio, Qionger He, Vivian Wan, John P. Adelman, Gen Ohtsuki, and Christian Hansel. 2016. "Activity-Dependent Plasticity of Spike Pauses in Cerebellar Purkinje Cells." *Cell Reports* **14** (11): 2546–53. https://doi.org/10.1016/j.celrep.2016.02.054.

Greenspan, R. J. 2001. "The Flexible Genome." *Nature Reviews. Genetics* **2** (5): 383–87. https://doi.org/10.1038/35072018.

Hanwell, David, Toru Ishikawa, Reza Saleki, and Daniela Rotin. 2002. "Trafficking and Cell Surface Stability of the Epithelial Na+ Channel Expressed in Epithelial Madin-Darby Canine Kidney Cells." *The Journal of Biological Chemistry* **277** (12): 9772–79. https://doi.org/10.1074/jbc.M110904200.

Hausman, D. M., and J. Woodward. 1999. "Independence, Invariance and the Causal Markov Condition." *British Journal for the Philosophy of Science* **50** (4): 521–583. https://doi.org/10.1093/bjps/50.4.521.

Hausman, Daniel M., and James Woodward. 2004. "Modularity and the Causal Markov Condition: A Restatement." *The British Journal for the Philosophy of Science* **55** (1): 147–61.

Hempel, Carl G. 1945. "Studies in the Logic of Confirmation (I.)." *Mind* **54** (213): 1–26.

Ito, M. 1989. "Long-Term Depression." *Annual Review of Neuroscience* **12**: 85–102. https://doi.org/10.1146/annurev.ne.12.030189.000505.

Ito, M., and M. Kano. 1982. "Long-Lasting Depression of Parallel Fiber-Purkinje Cell Transmission Induced by Conjunctive Stimulation of Parallel Fibers and Climbing Fibers in the Cerebellar Cortex." *Neuroscience Letters* **33** (3): 253–58. https://doi.org/10.1016/0304-3940(82)90380-9.

Ito, Masao, Masaki Sakurai, and Pavich Tongroach. 1982. "Climbing Fibre Induced Depression of Both Mossy Fibre Responsiveness and Glutamate Sensitivity of Cerebellar Purkinje Cells." *The Journal of Physiology* **324**: 113–34.

Kitano, Hiroaki. 2004. "Biological Robustness." *Nature Reviews Genetics* **5** (11): 826–37. https://doi.org/10.1038/nrg1471.

Marder, Eve, and Jean-Marc Goaillard. 2006. "Marder, E. & Goaillard, J.M. Variability, Compensation and Homeostasis in Neuron and Network Function. Nat. Rev. Neurosci. **7**, 563-574." *Nature Reviews. Neuroscience* 7 (August): 563–74. https://doi.org/10.1038/nrn1949.

Marr, David. 1969. "A Theory of Cerebellar Cortex." *The Journal of Physiology* **202** (2): 437-470.1.

Mitchell, Sandra D. 2009. *Unsimple Truths: Science, Complexity, and Policy*. University of Chicago Press.

Mitchell, Sandra D. 2008. "Exporting Causal Knowledge in Evolutionary and Developmental Biology." *Philosophy of Science* **75** (5): 697–706. https://doi.org/10.1086/594515.

O'Leary, Timothy. 2018. "Can Engineering Principles Help Us Understand Nervous System Robustness?" In *Biological Robustness: Emerging Perspectives from within the Life Sciences*, edited by Marta Bertolaso, Silvia Caianiello, and Emanuele Serrelli, 175–87. History, Philosophy and Theory of the Life Sciences. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-01198-7_9.

Pearl, Judea. 2000. *Causality: Models, Reasoning, and Inference*. USA: Cambridge University Press.

Puri, Basant K. 2020. "Calcium Signaling and Gene Expression." *Advances in Experimental Medicine and Biology* **1131**: 537–45. https://doi.org/10.1007/978-3-030-12457-1_22.

Ransdell, J. L., S. S. Nair, and D. J. Schulz. 2013. "Neurons within the Same Network Independently Achieve Conserved Output by Differentially Balancing Variable Conductance Magnitudes." *Journal of Neuroscience* **33** (24): 9950–56. https://doi.org/10.1523/JNEUROSCI.1095-13.2013.

Salmon, Wesley. 1984. *Scientific Explanation and the Causal Structure of the World*. Princeton University Press.

Schofield, Paul N., Robert Hoehndorf, and Georgios V. Gkoutos. 2012. "Mouse Genetic and Phenotypic Resources for Human Genetics." *Human Mutation* **33** (5): 826–36. https://doi.org/10.1002/humu.22077.

Schulz, David J., Richard A. Baines, Chris M. Hempel, Lingjun Li, Birgit Liss, and Hiroaki Misonou. 2006. "Cellular Excitability and the Regulation of Functional Neuronal Identity: From Gene Expression to Neuromodulation." *Journal of Neuroscience* **26** (41): 10362–67. https://doi.org/10.1523/JNEUROSCI.3194-06.2006.

Simon, Herbert. 1962. "The Architecture of Complexity." Proceedings of the American Philosophical Society, **106**(6): 467-482.

Spirtes, Peter, Clark N. Glymour, and Richard Scheines. 2000. *Causation, Prediction, and Search*. MIT Press.

Steel, Daniel. 2006. "Comment on Hausman & Woodward on the Causal Markov Condition." *The British Journal for the Philosophy of Science* **57** (1): 219–31.

Strata, Piergiorgio. 2009. "David Marr's Theory of Cerebellar Learning: 40 Years Later." *The Journal of Physiology* **587** (Pt 23): 5519–20. https://doi.org/10.1113/jphysiol.2009.180307.

Swensen, A. M. 2005. "Robustness of Burst Firing in Dissociated Purkinje Neurons with Acute or Long-Term Reductions in Sodium Conductance." *Journal of Neuroscience* **25** (14): 3509–20. https://doi.org/10.1523/JNEUROSCI.3929-04.2005.

Swensen, Andrew M., and Bruce P. Bean. 2003. "Ionic Mechanisms of Burst Firing in Dissociated Purkinje Neurons." *Journal of Neuroscience* **23** (29): 9650–63. https://doi.org/10.1523/JNEUROSCI.23-29-09650.2003.

Travis, J. 1992. "Scoring a Technical Knockout in Mice." *Science* **256** (5062): 1392–94. https://doi.org/10.1126/science.1351316.

Waddington, C. H. 1942. "Canalization of Development and the Inheritance of Acquired Characters." *Nature* 150 (3811): 563–65. https://doi.org/10.1038/150563a0.

Wagner, Andreas. 2005. *Robustness and Evolvability in Living Systems*. Princeton University Press. http://books.google.com/books?hl=en&lr=&id=pRFYAQAAQBAJ&oi=fnd&pg=PP1&dq=%22things+are+unimaginably+complex,+yet+they+have+withstood+a%22+%22cause+key+organismal+functions+to+fail+catastrophically%3F%22+%22etc.+(171,+183,+186,+368,+472,+499,+578)%E2%80%94have+been+used%22+&ots=2SPsKl5UWZ&sig=jVWjFeMKUuX3R-1i05SepxtxP_Y.

Weinberger, Naftali. (2018). "Faithfulness, Coordination and Causal Coincidences." *Erkenntnis* **83**: 113-133. https://doi.org/10.1007/s10670-017-9882-6.

Weisberg, Michael. 2006. "Robustness Analysis." *Philosophy of Science* **73** (5): 730–42. https://doi.org/10.1086/518628.

Whitacre, James M. 2010. "Degeneracy: A Link between Evolvability, Robustness and Complexity in Biological Systems." *Theoretical Biology and Medical Modelling* **7** (1): 6. https://doi.org/10.1186/1742-4682-7-6.

Williams, Terri A., and Lisa M. Nagy. 2001. "Developmental Modularity and the Evolutionary Diversification of Arthropod Limbs." *Journal of Experimental Zoology* **291** (3): 241–57. https://doi.org/10.1002/jez.1101.

Wimsatt, W. C. (1972, January). Complexity and organization. In PSA: Proceedings of the biennial meeting of the Philosophy of Science Association (Vol. 1972, pp. 67-86). D. Reidel Publishing.

Woodward, James. 2003. *Making Things Happen: A Theory of Causal Explanation*. Oxford University Press.

———. 2008. "Invariance, Modularity, and All That: Cartwright on Causation." Nancy Cartwright's Philosophy of Science. June 3, 2008. https://doi.org/10.4324/9780203895467-22.

———. 1997. "Explanation, Invariance, and Intervention." *Philosophy of Science* **64**: S26–41.

———. (forthcoming). "Causation and Mechanisms in Biology." *Minnesota Studies in Philosophy of Science*.

Wrinn, Kerri, and George Uetz. 2007. "Impacts of Leg Loss and Regeneration on Body Condition, Growth, and Development Time in the Wolf Spider Schizocosa Ocreata." *Canadian Journal of Zoology* **85** (August): 823–31. https://doi.org/10.1139/Z07-063.

Zhang, J. and Spirtes, P. (2008). "Detection of Unfaithfulness and Robust Causal Inference." *Minds and Machines*, **18**: 239–71.

Zucker, Robert S. 1999. "Calcium- and Activity-Dependent Synaptic Plasticity." *Current Opinion in Neurobiology* **9** (3): 305–13. https://doi.org/10.1016/S0959-4388(99)80045-2.