Binding Specificity and Causal Selection in Drug Design

Oliver M. Lean*†

Binding specificity is a centrally important concept in molecular biology, yet it has received little philosophical attention. Here I aim to remedy this by analyzing binding specificity as a causal property. I focus on the concept's role in drug design, where it is highly prized and hence directly studied. From a causal perspective, understanding why binding specificity is a valuable property of drugs contributes to an understanding of causal selection—of how and why scientists distinguish between causes, not just causes from noncauses. In particular, the specificity of drugs is precisely what underwrites their value as experimental interventions on biological processes.

1. Introduction. Nothing in molecular biology makes sense except in the light of specificity.[1] Trainees in the field learn from the very beginning that the function of a biomolecule—of proteins in particular—is critically tied to its structure. In turn, that structure-function relationship is understood in terms of the molecule's ability to interact preferentially, or specifically, with certain entities while avoiding others. Specificity of this kind applies to a wide variety of interactions between biomolecules: between enzymes and their substrates, transcription factors and their DNA binding sites, antibodies and the antigens they recognize, and DNA-DNA and DNA-RNA interactions, to name a few. It is because of specific interactions like these that biological

Received August 2018; revised May 2019.

*To contact the author, please write to: Institut supérieur de philosophie, Université catholique de Louvain, 1348 Louvain-la-Neuve, Belgium; e-mail: oliver.lean@uclouvain.be.

†This article was vastly improved thanks to feedback from Lindley Darden, Catherine Kendig, Lauren Ross, Ken Waters, James Woodward, two anonymous referees, and audiences at the 2018 SPSP conference in Ghent and the 2018 PSA in Seattle. This work was completed with the support of the John Templeton Foundation, under the research grant From Biological Practice to Scientific Metaphysics.

1. This is, of course, a variation on a famous statement by Dobzhansky (1973).

Philosophy of Science, 87 (January 2020) pp. 70–90. 0031-8248/2020/8701-004$10.00
Copyright 2020 by the Philosophy of Science Association. All rights reserved.

https://doi.org/10.1086/706093 Published online by Cambridge University Press

processes can be precisely regulated and organized in space and time, giving rise to the complexity and variety of life and its preservation within and across generations. I will refer to this property as *binding specificity*.

For such a centrally important concept, binding specificity has received surprisingly little philosophical attention. While all philosophers of molecular biology are certainly aware of it, it is usually mentioned only in passing, such as in the course of explicating other concepts such as arbitrariness (Stegmann 2004) or information (Sarkar 2005; Griffiths et al. 2015). Perhaps most notably, it is not among the three concepts ("mechanism," "information," and "gene") that have their own sections in the current "Molecular Biology" entry in the *Stanford Encyclopedia* (Tabery, Piotrowska, and Darden 2015). There, specificity in the binding sense is mentioned on two occasions: once during its account of DNA replication and once as a historical term that was later "replaced" by talk of information. As these examples show, binding specificity has almost exclusively been confined to the background in philosophical discussions.

This claim may come as a surprise to some, since "specificity" has been the direct focus of a debate in the philosophy of biology for the last decade (Weber 2006, 2017; Waters 2007; Woodward 2010; Griffiths and Stotz 2013; Griffiths et al. 2015; Stotz and Griffiths 2017). That debate is primarily centered around the issue of gene centrism—the question of whether, and to what extent, we can attribute to genes a privileged causal or explanatory role in development relative to nongenetic factors. The most recent iteration of this debate is whether genes' role in development is more "causally specific," at least with respect to protein synthesis.

However, the specificity concept discussed in that literature—what is sometimes called *fine-grained* causal specificity—is markedly different from the concept of binding specificity that is of interest here. A causal relationship is highly specific in the fine-grained sense to the extent that a large number of interventions on the cause each produce a unique effect (see Woodward 2010). By contrast, a simple on-off light switch—with only two states of the cause and two of the effect—is a very nonspecific relationship in this sense. Yet there are many such switch-like causal relationships in biological systems that nevertheless rely on mechanisms involving highly specific binding interactions. Since switch-like relationships are very nonspecific in the fine-grained sense, this suggests that the specificity of these interactions is not captured by the fine-grained causal specificity concept, even if the two are related in interesting ways, as I will discuss.

I do not wish to claim that philosophers have mistakenly conflated these two notions; in fact, they often take care to distinguish the two (e.g., Griffiths et al. 2015). Nor do I mean to criticise this focus on fine-grained specificity, which is quite reasonable in the context of the long-running debate on gene centrism. Nevertheless, a consequence of this focus is that binding specificity

https://doi.org/10.1086/706093 Published online by Cambridge University Press

has fallen by the wayside in the philosophical literature, and so inductees to the field may form the incorrect impression that fine-grained specificity is the most, even the only, important kind of specificity for biology. The primary purpose of this article is to argue that binding specificity is, or should be, as interesting to philosophers of biology as it is important to biology itself—at least as important as the fine-grained variety. There are certainly complex dependencies between these two concepts that are worth exploring (see Bourrat 2019); however, the aim of this article is not primarily one of comparison. Instead, its main goal is to show that binding specificity is philosophically interesting in its own right and to scratch the surface of a philosophical analysis of the concept that I hope will be continued.

The structure of this article is as follows. Section 2 provides an account of binding specificity as a scientific concept, in particular, of the way it is understood in the practice of drug design. Section 3 then develops a philosophical account of binding specificity as a causal concept of the sort discussed by Woodward (2010). It then connects binding specificity so construed to the issue of *causal selection*—the practice of singling out one or more causes of an outcome as being "the" cause, or of particular relevance or importance. I give a number of reasons why scientists often value high binding specificity and select causes on that basis. Section 4 concludes.

2. Binding Specificity in Drug Design. The idea of specificity—that a chemical substance could act selectively on certain tissues and avoid others—was first formed in the 1870s by Paul Ehrlich, who noticed that certain dyes stained only particular parts of a cell. Ehrlich later realized that this discriminate activity might be put to work not just in staining but also for therapeutic purposes. So was born the science of drug discovery and its ideal of a "magic bullet"—a drug that selectively targets a disease while leaving everything else in the body untouched (Ehrlich 1910, 1911; Strebhardt and Ullrich 2008). Ehrlich's vision of therapeutic magic bullets survives to this day: in the modern molecular picture of drug action, a key criterion for a good drug is the specificity with which it binds to its target, avoiding interactions with unwanted "decoys." While it is not the only relevant criterion, binding specificity is a critical point of evaluation for the discovery and design of biologically active molecules.

While it is part of the conceptual furniture throughout the molecular biological sciences, drug design's direct interest in binding specificity has produced an especially sophisticated understanding of it—of what it is, what contributes to it and, most importantly, how to get it. This makes the field of drug design an ideal place to look in search of a clear understanding of binding specificity.

This will require a considerable narrowing of focus, and so I should note from the beginning what is being left out. First, I will deal exclusively with binding interactions between drugs that are small molecules—known as

https://doi.org/10.1086/706093 Published online by Cambridge University Press

ligands—and a protein. I therefore leave aside cases in which the drug target is not a protein—such as chelating agents that bind to metal poisons. This also excludes many kinds of molecular interaction to which the concept of binding specificity applies, such as DNA-DNA and DNA-RNA interactions. While I will not discuss these, I take the associated specificity concept to be essentially the same, and so I intend what follows to be broadly applicable to those as well. I will also restrict discussion to cases in which the disease etiology (its underlying causal process at the molecular level) is at least partly known and, hence, a potential target for drug intervention has been identified. With that established, the aim is to develop a drug that acts on that chosen target to achieve a desired effect. This leaves out much of interest about how the disease etiology is elucidated in the first place (see Darden et al. 2018).

2.1. Specificity and Affinity. The philosophical analysis of binding specificity that I am undertaking here is a broadly pragmatic one: it aims to understand the role that the concept plays in the achievement of specified scientific aims. Such an analysis should, among other things, locate that concept in the broader context in which it plays such a role. As it happens, there is another key notion at work in drug design—the notion of *affinity*—that is not just related to binding specificity but embedded within it, making it especially important to this analysis. This section is therefore dedicated to an explication of affinity and its relationship with binding specificity.

Affinity refers to the strength of the binding interaction between any two molecules. Consider a binding interaction between a ligand (drug) L and a target protein P. (By definition, a "ligand" for a protein is a small molecule that binds to that protein.) P and L bind reversibly to form a *complex*, PL:

$$\mathrm{L} + \mathrm{P} \leftrightarrows \mathrm{PL}.$$

Mixed together in the right solution, L and P will reach a dynamic equilibrium in which the rate of binding and dissociation are equal. As we increase the concentration of L in the solution, more and more of P will be bound to and occupied by L at this equilibrium, until there is virtually no free P left. The stronger the interaction—the higher its affinity—the more readily the complex PL will form as the concentration of L is increased. A common measure of affinity, the "dissociation constant" K_D, gives the concentration of L at which there is as much bound protein as free protein (Schasfoort et al. 2012; see fig. 1); that is, the value of $[L]$ at which $[PL] = [P]$. (The notation "$[X]$" denotes the *concentration* of X.) Hence, a low K_D value indicates a high affinity. This value is particularly important to drug design because it affects the required dosage: a relatively high-affinity drug will require a relatively low dosage to achieve the same effect.

https://doi.org/10.1086/706093 Published online by Cambridge University Press

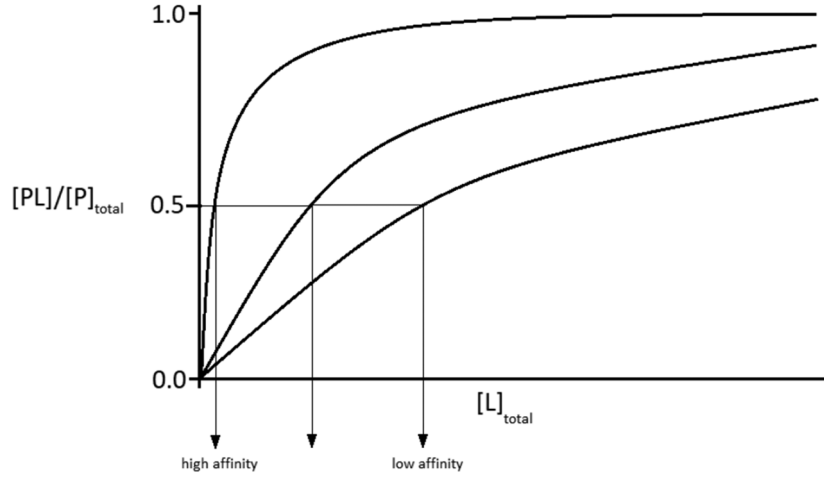


Figure 1. Dissociation curve relating the proportion of bound protein to total ligand concentration.

What makes one pair of molecules bind with higher affinity than another? That is, what are the *difference makers* (Waters 2007) for binding affinity? To answer this, it is important to understand that, thermodynamically speaking, molecules do not generally "like" to be bound together: complex formation produces an ordered and hence low-entropy state, and so most pairings are highly ephemeral. For two molecules to undergo sustained interaction, there is a debt of entropy to be paid, by freeing other molecules or energy into the surrounding environment. Pairs of molecules bind with high affinity when they succeed at paying this debt.

There are several factors that contribute to this thermodynamic book balancing. The first is the "fit" between the binding surfaces of the respective molecules, in terms of both their three-dimensional shape (stereochemistry) and the distribution of electrostatic charge (electrochemistry). Attraction between opposite charges releases energy; the stronger the opposite charges, and the closer the proximity, the more energy is released and hence the more thermodynamically favorable the binding. Another factor affecting affinity is desolvation—the expulsion of bound water molecules in the course of complex formation (Huggins, Sherman, and Tidor 2012). Finally, a less commonly acknowledged contributor to affinity is the flexibility or rigidity of the macromolecules involved: if a molecule is highly flexible, binding interactions restrict its degrees of freedom and make its bound state less favorable. Hence, two molecules with virtually identical binding surfaces may nevertheless bind a target with different affinities if one is more rigid than the other (Eaton, Gold, and Zichi 1995; Carothers, Oestreich, and Szostak 2006).

https://doi.org/10.1086/706093 Published online by Cambridge University Press

With the above account in mind, how does binding affinity relate to binding specificity? The crucial point is that, as we will see below, high affinity is not sufficient to make a binding interaction specific. While affinity refers just to the strength with which a ligand binds its target, specificity refers to "the degree of discrimination expressed by a ligand" for its target (Kangas and Tidor 2000, 9121) or "functional discrimination between molecules competing for a common ligand" (Eaton et al. 1995, 635). In short, a specific drug binds to its target not just strongly but preferentially.

This leads us to a simple way of understanding the relationship between these two concepts: binding specificity can usefully be understood as *relative affinity*. Affinity is comparatively simple: it applies to a dyadic relationship, the binding interaction between a ligand and its target (although this is not symmetrical—the affinity of *P* for *L* is not equal to the affinity of *L* for *P*). Binding specificity, however, is the affinity of a ligand for its target relative to its affinity for decoys. Because of this, quantifying binding specificity requires defining a range of targets and decoys among which relative affinities are being measured. There are different ways of doing this: a canonical treatment of binding specificity by von Hippel and Berg (1986), adapted by Eaton et al. (1995), quantifies it as the ratio of ligands bound to the target versus the sum total of all ligands bound to "background" sites. Others (e.g., Carothers et al. 2006) quantify specificity for a target relative to a single other species. Different measures suit different purposes; in any case, quantifying the specificity of a ligand requires reference to at least one binding partner other than the target in question. A drug that is specific against one background may therefore be nonspecific in another, while the affinity for its target may stay the same.

There is a final complication to binding specificity that will become relevant in what follows, which is that it is often desirable for a drug to act on more than one protein—to a family of related proteins, for instance. Cases like these call for *broad specificity* (Huggins et al. 2012),[2] that is, specificity for some defined range of targets. While specificity for a range of proteins may seem paradoxical, the drug may still discriminate between those targets and another range of decoys. Like the difference between a garden plant and a weed, what makes a given protein a target or a decoy is purely a matter of whether that interaction is desirable for a given purpose. In many cases, then, specificity is not just relative to a given background but defined partly by

2. Huggins et al. (2012) actually use the term "selectivity" instead of "specificity." I retain the latter term for consistency. Pharmacologists often distinguish these two ideas—"selectivity" referring to side effects or lack thereof at the clinical level rather than binding at the molecular level. However, in this case Huggins et al. are evidently attributing "selectivity" to molecular-level binding interactions.

https://doi.org/10.1086/706093 Published online by Cambridge University Press

value judgments that distinguish "good" and "bad" interactions and, hence, "broad" specificity from merely poor specificity.

2.2. How to Design Specific Drugs: Experiments and Heuristics. The previous section presents part of the problem space in the process of drug design. What kinds of strategy do scientists use in navigating this problem space? The range of approaches are, of course, too numerous to list here. I will restrict discussion to a few, which divide into *empirical* and *rational* methods.

Empirical methods are those that reveal candidate drugs primarily through experiment. One example is systematic evolution of ligands by exponential enrichment, or SELEX (Eaton et al. 1995). SELEX involves a selection process on large arrays of *aptamers*—strings of DNA or RNA that, like proteins, can fold three-dimensionally to produce structures that undergo specific binding. The basic principle behind SELEX is to (1) select from these aptamers according to how well they bind to a chosen target, (2) amplify and systematically vary those selected variants, and (3) repeat.

However, there is a practical drawback to this technique: Being an in vitro process, it is not possible to directly select aptamers that bind specifically to their target from the full, messy physiological background of the human body. However, Eaton et al. (1995) describe an experimental heuristic that they find to be effective: by selecting directly for high affinity—which, as we have seen, is not dependent on background—specificity tends to come for free. Apart from past experience, Eaton et al. offer theoretical reasons why the two should tend to coincide: increased affinity for a target comes largely from increasing exactness of fit between the interacting species, and it is likely that changes in the aptamer that increase the fit for its target will decrease the fit for background decoys.

As with all heuristics, however, this has limitations; its reliability depends on the robustness of the relationship between specificity and affinity, and as we have seen the two do not necessarily coincide. Some have offered counterexamples in which this heuristic breaks down: Carothers et al. (2006) find that selecting for high-affinity binding to GTP does not tend to increase specificity against chemically similar GTP analogs. Instead, they suggest that specificity in this case is gained, not from improving intermolecular interactions but from the stability of the intramolecular forces of the aptamer, since a more flexible molecule with more degrees of freedom is better able to deform to accommodate the analogs' chemical differences. Huggins et al. (2012) offer another counterexample, finding that reducing affinity for decoys often requires changes that reduce affinity for the target as well. The general lesson here is this: whether high target affinity begets high specificity is heavily dependent on context, including the ranges of desirable and undesirable targets and the similarities and differences between them.

https://doi.org/10.1086/706093 Published online by Cambridge University Press

2.3. How to Design Specific Drugs: Rational Principles. As well as empirical methods of drug discovery such as SELEX, there are what are called rational approaches to drug design—"rational" in the sense of being driven largely by prediction based on theoretical principles rather than by experimental techniques. Huggins et al. (2012) detail a range of such approaches: by considering (1) the properties contributing to affinity and specificity in general, (2) the details of the structure of the ligand-target complex, and (3) the similarities and differences in those properties between targets and decoys, it is possible to make educated guesses about how to improve the specificity of drug treatments by altering their chemical structures. Those similarities and differences reveal "handles" that are potentially exploitable for attaining selectivity goals by adding, removing, or changing certain chemical groups. These handles include the following factors:

1. *Shape complementarity*: Crystal structures of targets and decoys sometimes reveal a difference in the shape of the active site—often of only a single amino acid residue. Cases like these invite the design of a drug that fits easily into the target site but clashes with the equivalent site on the decoy.
2. *Electrostatic complementarity*: Electrostatic forces change exponentially with distance, which makes them ripe for exploitation for selectivity purposes. This includes both positive design (developing molecules that interact especially well with targets) or negative design (creating clashes with decoys that are tolerated by targets).
3. *Flexibility*: As we have seen, despite the similarities in binding sites, ligands can be designed to favor the target if that target is more deformable than decoys. This can be done by designing a ligand that binds to the target's deformed state rather than the state that it has in common with the more rigid decoy (e.g., Liu and Gray 2006).
4. *Hydrophobic effects*: Sometimes a target and decoy will differ in how strongly they bind to a water molecule when unbound. If the decoy binds the water more strongly than does the target, displacing that water from the decoy will be less favorable, and so the drug's affinity for the decoy relative to the target will be reduced.
5. *Allosteric pockets and noncompetitive binding*: A simple way to inhibit a target and avoid decoys is to target an *allosteric* site (a binding site other than the active site) that the target has while the decoys do not.

The above sets out just some of what is philosophically interesting about specificity and its role in drug design: it reveals problems faced and solutions found by scientists with well-defined practical goals and the role played by concepts such as specificity in working toward achieving them. For one, the real but tenuous link between specificity and affinity affords the use of a

https://doi.org/10.1086/706093 Published online by Cambridge University Press

simple but exception-ridden heuristic—one that works in some cases but breaks down in others. Second, the complexities of specificity profiles and the rational approaches for optimizing them demonstrate the value of understanding the difference makers that are relevant to one's particular context—for example, the ranges of desirable targets and undesirable decoys and their shared and contrasting features. With the above account of binding specificity in hand, we can now investigate the lessons these issues hold for philosophers interested in causation.

3. Binding Specificity and Causation. The previous section outlined the role of the concept of binding specificity in drug design—why it matters, how is defined and quantified, and the methods scientists use in their search for drugs that possess it. Here I explore how these lessons may be relevant to philosophers interested in causation in biology.

In particular, the theoretical importance of specificity to drug design, and the explicit practical motivations for reasoning about it, may inform philosophical discussion about *causal selection*. Causal selection refers to a cluster of philosophical issues surrounding the following observation: while any given event in the world has, strictly speaking, many or even infinitely many causes, we tend to select or single out only one or a few of these as being "the" cause of that event (Hesslow 1988). Philosophers since Mill (1843/1981) have doubted that any principled rationale can be offered for this phenomenon, given how context sensitive and "capricious" it appears to be. However, philosophers in recent times have challenged this widespread pessimism: as elaborated below, Woodward (2010) argues that as well as distinguishing causes from noncauses in biology, causal relationships can themselves differ in a variety of properties. In turn, these properties can be more or less conducive to a variety of practical goals.

My aim in this section is to expand on Woodward's discussion of specificity as a basis for causal selection, using lessons about its use in the field of drug design detailed above. Drug design offers a highly tractable case study for this purpose: as we have seen, the field has explicit practical goals, and it selects drugs based (partly) on binding specificity precisely because it is conducive to those goals. Hence, if we can interpret binding specificity as a causal property, then the various reasons why it is valued in drug design constitute a range of scientific justifications for causal selection. My aim in this section is to develop this line of thought.

As an aside, the reader may have the following objection: in cases of causal selection discussed in the literature so far, there are a large number of causes that jointly produce a given effect (such as a phenotype), and the question is why we tend to single out one or a subset of those causes (such as genes) from among the full complement of factors that interact to produce the outcome. In drug design, in contrast, what scientists are selecting between are alternative

https://doi.org/10.1086/706093 Published online by Cambridge University Press

candidates for playing the same causal role. Because of this, one might object that this does not count as a case of causal selection at all. It is certainly worth keeping in mind that the type of "selection" being done in this context is markedly different. However, I take a central explanandum in discussions of causal selection to be as follows: as well as distinguishing causes from noncauses, scientists also distinguish between causes on the basis of further features in which they differ. Hence, if binding specificity can be understood as a causal property, then the fact that drugs with this property tend to be favored more than those without it still constitutes an instance of the phenomenon that we are aiming to understand.

A final potential objection: my argument follows Woodward in discussing causes and causal selection from an *interventionist* perspective (see Woodward 2003, 2010).[3] It should be acknowledged that this approach has popular rivals. Most significantly, interventionism is often contrasted with *mechanistic* accounts (Machamer, Darden, and Craver 2000; Glennan 2017): on these accounts, a causal understanding of a biological phenomenon centrally concerns discovering its underlying mechanisms. In contrast, my discussion will treat causal relations in terms of the manipulability of an effect by intervention on its causes. For some, this interventionist approach misses much of what is important to the scientific investigation of biological phenomena, especially at the molecular level. I should therefore take some time to place the following discussion in context with this debate.

I am sympathetic to claims that the interventionist approach, especially its use of causal graphs that I will be employing, is not omnipotent as a means of clarifying the full range of explanatorily salient features of molecular processes (Kaiser 2016). Despite this, however, I take the view—which I cannot defend in full here—that mechanistic and interventionist approaches deal with different but equally important dimensions of biological reasoning and practice. Mechanistic considerations certainly play a critical role in drug design's understanding of specificity: as explained in section 2.3, rational approaches centrally depend on a detailed understanding of the mechanisms by which the drug binds to its target and decoys (see Darden et al. 2018). This kind of mechanistic understanding is clearly valuable for understanding and improving drug specificity in some contexts.

Even so, it is evident from what we have seen that binding specificity itself is quantifiable without reference to these mechanistic details. In other words, specificity is multiply realizable: the same end can be achieved through a variety of mechanistic means. Because of this, specificity as a quantity offers a common currency by which to evaluate otherwise different types of binding interaction for a variety of practical purposes. More generally, this shows

3. An alternative account of causal selection to Woodward's can be found in Franklin-Hall (2015). I lack the space for a comparison between these two accounts here.

https://doi.org/10.1086/706093 Published online by Cambridge University Press

that it is possible to distinguish causal relations from each other in ways that abstract from their underlying mechanisms, and in many contexts—although certainly not all—it is useful to do so. Instead, we can differentiate causal relations in terms of the way they respond to various external manipulations. At the same time, I acknowledge that there is much interesting work to be done in exploring how the mechanistic understanding of disease processes may further illuminate the causal, practical, and epistemic dimensions of specificity.

3.1. Binding Specificity as a Causal Property. Here we come to the first key question: Can binding specificity, as elaborated in section 2, be interpreted as a causal property in the interventionist sense discussed by Woodward (2010)? There, Woodward argues that causal relations in biology vary along three different dimensions: *stability*, *proportionality*, and *specificity*. A natural starting point, then, is to see whether the notion of binding specificity in drug design can be interpreted as an instance of this third property of causal specificity.

Importantly, there are in fact two different senses of causal specificity highlighted in that work: the first is the aforementioned fine-grained sense of specificity discussed in section 1—the degree to which a variety of interventions on the cause uniquely specify a range of values of the effect. The second sense is one that "seems prima-facie rather different": the idea "that a causal relationship is specific to the extent that a single (type of) cause produces only a single (type of) effect and to the extent that each single type of effect is produced only by a (type of) single cause" (Woodward 2010, 308). In contrast, a causal relationship is nonspecific in this sense either if the cause produces many other effects or if the effect is also brought about by many other causes. I will henceforth call this *one-one specificity*.

As discussed above, it is difficult to interpret binding specificity as causal specificity in the fine-grained sense, since even binary switch-like causal relationships—which are very nonspecific in the fine-grained sense—rely on highly specific binding interactions. This suggests that the one-one variety may be better suited to a causal analysis of binding specificity. Woodward appears to agree, as evidenced in his discussion of enzyme-substrate interactions: "Enzymes are commonly described as 'very specific' with respect to the substrates on which they act and the reactions they catalyze. This is usually understood to mean that a particular enzyme will often act only [on] a particular substrate or a small set of these (rather than a large class of different substrates) and that it will catalyze just one kind of reaction with this substrate. In other words, the smaller the number of different substrates an enzyme can bind, the greater its 'specificity'" (2010, 309). Since the action of an enzyme on its substrate begins with their binding to form a complex, it is evident that Woodward is interpreting (what I here call) binding specificity as an instance of one-one causal specificity. I proceed in broad agreement with

https://doi.org/10.1086/706093 Published online by Cambridge University Press

this claim; however, the complexities of specificity in drug design as discussed in section 2 call for some elaboration and refinements.

The first step is to specify exactly what causes are standing in such one-one relationships in the case of binding interactions between molecules. For Woodward, cause-effect relationships are (or can be represented as) relationships between *variables*, each with two or more *values*. For example, to represent a causal relationship between a light switch and lightbulb, we need to define a variable for each—say, $X = \{\text{switch on, switch off}\}$ and $Y = \{\text{light on, light off}\}$. Given that, we can represent the causal relationship between the two in directed graph form:

$$X \rightarrow Y$$

The arrow represents the claim "*X* causes *Y*." The meaning of this claim is that there are *ideal interventions* on *X*—surgical, exogenous manipulations of *X*'s value—that would change *Y* (see Woodward [2010] for more detail).

With this in mind, there is work to be done toward interpreting binding specificity as an (interventionist) causal concept. First of all, translating binding specificity into a causal property requires us to define precisely the variables and their values to which the concept of binding specificity applies. Here there seems to be a problem: we saw in the previous section that binding specificity is predicated of the binding of a ligand with a target biomolecule (or between two molecular species more generally). But molecules are apparently *entities*, not variables, and so they are not the sort of thing that can stand in causal relations in the interventionist sense. What, then, are the variables whose causal relation can be described as more or less one-one specific?

Section 2 suggests that we should begin with a causal interpretation of binding affinity, since specificity is in effect relative affinity. As we have seen (fig. 1), affinity is derived from a dissociation curve relating total ligand concentration $[L]_{\text{total}}$ and proportion of occupied target sites $[PL]/[P]_{\text{total}}$. This reflects the fact that varying the ligand concentration causes changes in the proportion of bound target—a clear causal relationship between variables in the interventionist sense. Hence, we can interpret affinity as a property of the causal relation $[L]_{\text{total}} \rightarrow [PL]/[P]_{\text{total}}$. (To simplify, I henceforth use the value $[PL]$ as the effect variable, rather than $[PL]/[P]_{\text{total}}$; the resulting causal relation remains true.)

Some clarification is needed at this point: My aim here is not to clarify the *causes of specificity*. If it were, it would be necessary to treat affinity not as a property of relations between causal variables but as a causal variable in itself: that is, if the affinity K_D of a binding interaction is an effect, what are its causes? The answers to this question include the properties listed in section 2.3: for example, mutations affecting the flexibility of the target protein will sometimes change the affinity of the binding interaction. Instead, my

https://doi.org/10.1086/706093 Published online by Cambridge University Press

aim here is to understand binding specificity as *a causal property*, that is, as a property of causal relations, not as an effect that has causes. My motivation for doing so is to understand how binding specificity relates to other causal properties and to discuss how and why these factors are conducive to scientists' practical and epistemic aims. Again, this is not so say that the concrete mechanistic details of binding interactions are not important tout court; they are simply orthogonal to the particular question at issue here.

Now that we have established affinity as a causal property, the next step is to derive from it a similarly causal version of binding specificity: if the affinity of L for P pertains to the causal relation $[L]_{\text{total}} \rightarrow [PL]$, and if binding specificity amounts to relative affinity, then binding specificity as a causal property should be a comparison between that relation and the equivalent relations with some range of decoys. As we saw, there are different ways of picking out this range, for example, by picking out a single one or including the full range of other proteins to which the ligand can bind in a given physiological environment. As Woodward acknowledges, "a great deal will turn on how this restricted range is specified" (2010, 311). For simplicity I focus on the case of a single target P and a single decoy D. Hence, the causal relations being compared are those between ligand concentration $[L]_{\text{total}}$ and the formation of complexes $[PL]$ and $[DL]$ (see fig. 2).

We can now begin to elaborate on Woodward's proposal that binding specificity is an instance of one-one causal specificity—that it implies a single cause producing a single effect and a single effect arising from a single cause. Take, for example, the claim "ligand L is highly specific for its target P relative to decoy D"—call this claim "Spec" for short. How exactly might we interpret Spec in a way that involves a one-to-one causal relationship?

At this point we can see some complications. First, we cannot take Spec to mean that the ligand binds to the target exclusively. Woodward acknowledges this (2010, 312), while leaving open how exactly we are to interpret the one-one specificity claim instead. One solution may be to impose some threshold affinity below which a decoy interaction can be ignored, which may be chosen differently depending on context. Another possible solution,

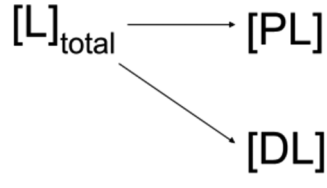


Figure 2. Affinity of L for P is a feature of the causal relation $[L]_{\text{total}} \rightarrow [PL]$. Since specificity is relative affinity, specificity should be a matter of comparison between that and at least one other binding relation—say, with a decoy D.

https://doi.org/10.1086/706093 Published online by Cambridge University Press

somewhat less context sensitive, is analogous to the move from deterministic to probabilistic causation: rather than a function relating causes to unique effects, we can simply consider the *distribution* of ligands between target and decoy, for example, the ratio $[PL]/[DL]$. Sure enough, for a highly specific ligand that distribution will be heavily weighted toward the target, whereas the distribution will be flatter for a relatively nonspecific ligand. This interpretation honors the fact that specificity is a matter of degree, with "ideal" specificity—a strict one-one relationship—being just that: an ideal that is never actually attained.

There is another qualification to add: however preferentially L binds to P, Spec does not imply that only L can bind to P in this way. That is, Spec does not rule out the idea that another ligand L^* can also bind to P with high affinity. So while Spec may rule out a one-to-many relationship between cause and effect, it is compatible with a many-to-one relationship. However, recall that binding interactions are in an important sense symmetrical: neither L nor P is "the" cause of binding. We can just as easily hold $[L]$ fixed and consider the causal relationship $[P] \rightarrow [PL]$. The specificity of that relationship is indeed threatened by the presence of the rival ligand L^*. In short, whether a binding interaction is specific depends on which of the entities we treat as the target. Because of this, we may only require a slight amendment to Woodward's claim, from a conjunction to a disjunction: binding specificity implies that a cause produces only one effect *or* that an effect is only produced by one cause, depending on which of the interacting entities we treat as the cause and which the effect. Interestingly, this difference in perspective may underlie a further distinction that scientists appear to make: between specific *action* and specific *recognition*. This question and its potential significance will be addressed in future work.

Setting these issues aside, section 2 has raised another complexity to binding specificity that a causal interpretation would do well to accommodate: the idea of broad specificity. It is often desirable to design a drug with the right specificity "profile"—one that binds to a range of desirable targets while still avoiding undesirable ones. So far we lack a way of accounting for the difference between drugs with this broad specificity and those that are simply indiscriminate. How might we incorporate broad specificity into the emerging causal account?

I suggest that Woodward's (2010) notion of proportionality provides a solution. Briefly, a causal explanation for some effect is proportional to that effect to the extent that it includes relevant information and excludes irrelevant information about the causal relation in question. When citing smoking as a cause of lung cancer, for example, we generally do not care to specify which brand of cigarettes is smoked. Why do we typically favor the claim that smoking causes cancer over the claim that smoking Pall Malls causes cancer, even though both are strictly true? The difference is that the latter claim

https://doi.org/10.1086/706093 Published online by Cambridge University Press

omits relevant information: it does not negate, say, the false proposition that one could avoid cancer by switching to Marlboros. In general, then, proportionality guides our choice of the grain of the variables being related: their values should be selected to contain all and only the relevant information about which interventions would make a difference to the outcome.[4]

With that in mind, a possible way of accounting for specificity profiles might be to group desirable targets and undesirable decoys into single values, then compare the extent to which the ligand affects each of them taken together (see fig. 3). This follows the principle of proportionality because it groups values whose differences are taken to be irrelevant or at least less relevant: the aim is to bind all targets and avoid all decoys.

There is more to be said about how the proposed causal interpretation of binding specificity will require adjustments and nuances based on context. However, the version presented captures at least some of the important features of binding specificity revealed in section 2 and provides the beginnings of a way to translate what was found there into a case study for how scientists choose between drugs on the basis of their different causal properties. We now move on to an exploration of the purposes guiding that selection process.

3.2. Binding Specificity as a Rationale for Causal Selection. Woodward's discussion of differences between causal relations has an explicitly pragmatic aim: the point is not simply that there exist differences between causes of a given effect in terms of properties like stability, specificity, and proportionality; it is that these differences matter to us because they affect how those causes can be exploited in the course of scientific investigation. Interventionist approaches to causation, especially Woodward's, are based on the notion that causes matter to us largely (or at least partly) because only relations that are causal can be exploited for purposes of manipulation or control. More than that, however, Woodward argues that our distinguishing between causes can also be understood with that in mind. For example, we tend to prefer or emphasise highly stable causal relationships over unstable ones because the former are more reliable targets for control purposes: intervening on them is more likely to change the effect under a range of background circumstances and under a wider range of interventions (Woodward 2010, 315; more below).

Now that we have an interpretation of binding specificity as a causal concept, we can interpret the distinctions scientists make between causes on the

4. Franklin-Hall (2016) critiques Woodward's treatment of proportionality on the basis that it cannot avoid recommending variables of a vacuously high level, i.e., radical disjunctions of possible causes of a given effect. I lack the space to address this criticism in detail. In brief, I think there are in fact very good reasons not to allow disjunctive variables of this kind: variables in causal models should have values that are mutually exclusive, i.e., that cannot be manipulated independently of each other (see Woodward 2017).

https://doi.org/10.1086/706093 Published online by Cambridge University Press

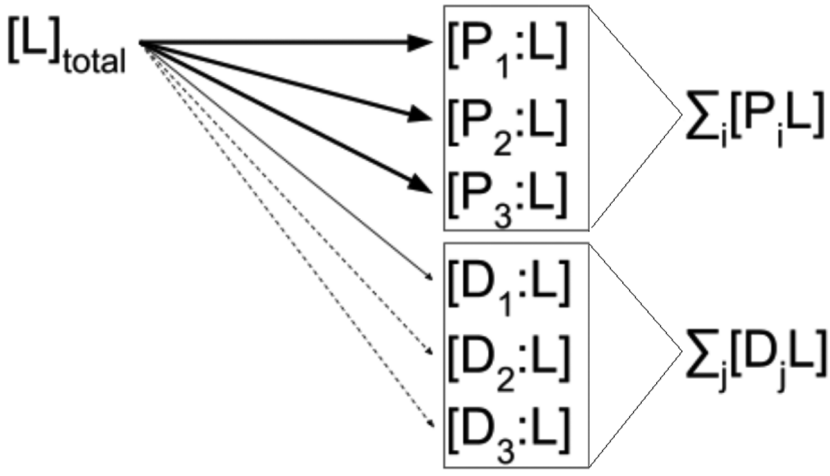


Figure 3. Variables can be grouped through the principle of proportionality to accommodate broad specificity. We can arrange the binding interactions into "targets" and "decoys"—P_1–P_3 and D_1–D_3—on the basis of whether those interactions are desirable. We can then group binding interactions with targets and decoys separately, and measure the specificity of L for one set over the other. (Here I part from the convention of causal graphs by weighting the arrows to represent varying affinities.)

basis of this property, and their reasons for doing so, as an instance of this same phenomenon of causal selection. With that in mind, what is the practical upshot of the causal interpretation of binding specificity as outlined in section 3.1? That is, for what purposes do scientists distinguish between drugs on the basis of their binding specificity?

The first and most obvious point on this matter has run through the entire discussion: drugs with low specificity—those that interact liberally with decoys—are likely to produce a host of unwanted side effects. This point is also made by Woodward: Chemotherapy, for example, is far from ideal because it kills healthy cells as well as cancer cells, and so the search for better cancer treatments is largely the search for treatments that avoid these off-target effects. As well as the negative downstream effects they tend to produce, decoy interactions are also to be avoided because they make some of the drug unavailable for binding to the target, and so a higher dose is needed to achieve the desired outcome.

Another observation Woodward makes is that while one-one and fine-grained specificity are not strictly the same notion, the two are at least interconnected. One reason is that cases of fine-grained specificity will also tend to qualify as one-one specific, in that each value of the cause variable tends to cause one and only one value of the effect variable. In other words, if we

https://doi.org/10.1086/706093 Published online by Cambridge University Press

combine a set of one-one specific causes and consider them as a single variable, intervening on that variable will allow a variety of fine-grained alterations. In the case of drug design, we could interpret this claim in the following way: once we have a range of highly specific drugs at our disposal, we can make a range of fine-grained alterations to physiology by choosing from among that range. To put it the other way round, if you want to exert fine-grained control over something, you need a toolkit of interventions that are each specific in the one-one sense. However, as we saw in section 3.1, binding specificity only really implies that the cause in question produces a single effect, not that the effect can only be produced by that cause. This makes sense in the case of drug design: it does not affect our evaluation of a drug's effectiveness that there are other drugs that have that effect.

The above are benefits to one-one specificity in molecular cases that Woodward has already identified. The first of these has to do with the aforementioned notion of broad specificity: one situation that calls for a drug with a broad specificity profile is when the target protein is owned by a rapidly mutating virus. In cases like these, we want a drug that can potentially target not just the wild type of the protein but any mutant versions that may arise. This leads to an interesting connection when we interpret this in causal terms: in cases like these, the purpose of designing a drug with a broad specificity profile is to ensure that the drug's effect on the virus is highly stable.

To elaborate: in Woodward's sense, the stability of a causal relationship is the extent to which it continues to hold under a range of changes in background conditions (2010, 291). Here a "background condition" has a technical meaning: it is any condition not explicitly represented in the particular causal relationship being discussed (in this case, the relationship between the administering of the drug and the curing of the infection). With that in mind, the form of the protein expressed by the virus qualifies as a "background" condition even though it is the direct target of the drug. Since changes in that condition through mutation can alter whether the drug is effective, then a drug whose specificity profile covers a larger range of possible mutant proteins therefore has a more stable causal relationship with the treatment of the infection: it remains effective under a wider range of virus mutations.

3.3. The Epistemic Value of Binding Specificity. Finally, there is a second dimension to the scientific importance of binding specificity that Woodward does not raise. That is, as well as the various therapeutic benefits to binding specificity—the avoidance of side effects, the maintenance of stable effects, and so on—there is also a more directly epistemological aspect to its value. As we have seen, many drugs (although not all) are developed with the guidance of knowledge about the causal process underlying the disease in question—for example, knowledge of the physiological processes that the disease

https://doi.org/10.1086/706093 Published online by Cambridge University Press

disrupts. This informs many decisions about treatment strategies, including decisions about which protein to target with a drug in the first place. As it happens, molecules with high specificity are not just an outcome of this process but a valuable means to gaining that causal knowledge in the first place. This is summarized neatly by Eaton et al. (1995, 634): "Many ideas about which target is the right one to use to treat a particular disease will be wrong, since our understanding of biology is far from perfect; to go forward from a right idea or to discard a wrong idea requires data that definitively show that a given compound, which inhibits the activity of its target completely and specifically, does or does not affect the course of a disease in an appropriate animal model (or in a human clinical trial). The overall process of drug discovery is thus facilitated by experiments with highly specific compounds."

In other words, in order to know that a given target is appropriate for treating a disease, one has to manipulate that target (e.g., inhibit it) and see what happens. But to know that the observed effects are in fact because of the inhibition of that particular target, one has to be sure that one's experimental intervention did not also affect other things as well; otherwise, the observed effect may be because of those off-target effects instead of the intended one. This is why getting reliable data about the etiology of a disease, and hence discovering the most appropriate drug target in the first place, is aided by the use of compounds with high binding specificity.

This suggests a deeper interpretation of binding specificity in relation to Woodward's causal framework. Recall that on this framework, causal claims are effectively claims about the effects of ideal interventions, which fully and uniquely determine the value of their target variable without affecting anything else in the system. The interventions built into the meaning of causal claims are called ideal because, in reality, it is always at least logically possible that something else that has not been explicitly controlled for is affecting the outcome. This is why no experiment we can actually perform to test causal claims yields categorical certainty: the best one can do is control for as many relevant factors as possible.

This characterization of ideal interventions is strikingly similar to Eaton et al.'s (1995) account of the epistemological value of binding specificity. As we have seen, the ideally specific ligand is one that fully inhibits its target (at a certain concentration) without affecting anything else. Given the probabilistic fuzziness of thermodynamics, however, this can only ever be an ideal: decoy interactions can be minimized to a tolerable level but never eliminated. Nevertheless, the notion of an ideally specific drug serves as a useful guide to be approached but never reached.

This analogy leads to a compelling suggestion for how to connect binding specificity to Woodwardian causal language: *a ligand is specific for its target to the extent that its introduction approximates an ideal intervention on that target*. As well as the therapeutic benefits to these precise manipulations, their

https://doi.org/10.1086/706093 Published online by Cambridge University Press

epistemic value as highlighted by Eaton et al. (1995) neatly exemplifies the more general notion at work in Woodward's analysis of causal claims: the more closely that ideal can be approached, the more confidence we have in the accuracy of those causal models in the first place. We can potentially generalize this to apply beyond the confines of drug design. The same principle applies, for example, to the creation of monoclonal antibodies in fluorescence imaging: the specificity of those fluorescent dyes for particular cell components is critical for avoiding background noise, thus improving the quality of the images and the reliability of the data. The closer one gets to total specificity, and hence to an ideal intervention, the more informative the resulting signal.

Another case worthy of further study is the use of experimental manipulations such as RNA interference: the value of these techniques is in large part the precision with which they target specific RNAs in specific tissues and, hence, the confidence they yield in the causal role of the RNAs being targeted. However, another important source of their value is how easy it is to knock down an RNA as long as you know its sequence. As we have seen, the same is not true of proteins: knowing its primary sequence, or even its tertiary structure, leaves much to be done in terms of designing ligands that target it specifically. Evidently, there is much more to be explored on these issues outside the realm of drug design as well as within it.

4. Conclusion. In this article I have shown that the notion of binding specificity is (1) centrally important to molecular biology and related fields, especially to pharmacology and drug design; (2) understudied by philosophers, particularly in relation to causation in biology; and therefore (3) worthy of more philosophical attention. I described the notion of binding specificity as it is understood and aimed for in drug design, including its relationship with the embedded concept of affinity, how the two are related both conceptually and experimentally, and the various difference makers that contribute to them. This account reveals a number of observations that are ripe for further study, such as how specificity is often sought in ways that are indirect but more practically feasible or how rational design principles can often precede experimentation in achieving it.

I then developed a causal interpretation of binding specificity, elaborating on Woodward's suggestion that it is an instance of the more general notion of one-one causal specificity. I broadly supported this interpretation while proposing some amendments prompted by the study of its involvement in drug design. This includes qualifying the one-one condition in two ways: first, to accommodate the probabilistic nature of molecular interactions and, second, to note that specificity claims often tolerate either many-to-one or one-to-many relations.

I have also shown a number of ways in which binding specificity relates to causal selection, that is, for why scientists often distinguish between drugs on

https://doi.org/10.1086/706093 Published online by Cambridge University Press

the basis of that causal property. In particular, I have pointed out that broad specificity of a drug for its target can be conducive to treatments that are causally stable under mutations in the target virus. I have also highlighted a link between broad specificity and the causal notion of proportionality.

As well as these therapeutic benefits, I have highlighted a critical epistemic dimension to why molecules with high binding specificity are valuable: their use in experimental interventions provides reliable data about the causal structure being investigated in the first place. This point suggests that as well as being viewed as a property of causal relations, alongside stability or proportionality, binding specificity may actually be linked to the notion of ideal interventions that are central to the interventionist framework as a whole.

Much of the discussion in this article has been open-ended: each answer proposed raises further questions and invites further study. As well as being inevitable, this is also deliberate. As discussed in section 1, the amount of philosophical work investigating the role of binding specificity is proportionally tiny relative to the importance of the concept in molecular biology and neighboring disciplines. Hence, as well as its explicit aim of understanding binding specificity as a causal concept, this article has a broader aim of inviting further investigation into this concept and its practical and explanatory importance. It is evident from what I discuss here that binding specificity presents a vast and largely untapped well of resources for philosophers of biology.

REFERENCES

Bourrat, Pierrick. 2019. "Variation of Information as a Measure of One-to-One Causal Specificity." *European Journal for Philosophy of Science* 9:11.

Carothers, James M., Stephanie C. Oestreich, and Jack W. Szostak. 2006. "Aptamers Selected for Higher-Affinity Binding Are Not More Specific for the Target Ligand." *Journal of the American Chemical Society* 128 (24): 7929–37.

Darden, Lindley, Lipika R. Pal, Kunal Kundu, and John Moult. 2018. "The Product Guides the Process: Discovering Disease Mechanisms." In *Building Theories: Heuristics and Hypotheses in the Sciences*, ed. David Danks and Emiliano Ippoliti, 101–17. Cham: Springer.

Dobzhansky, Theodosius. 1973. "Nothing in Biology Makes Sense Except in the Light of Evolution." *American Biology Teacher*, 35 (3): 125–29.

Eaton, Bruce E., Larry Gold, and Dominic A. Zichi. 1995. "Let's Get Specific: The Relationship between Specificity and Affinity." *Chemistry and Biology* 2 (10): 633–38.

Ehrlich, P. 1910. "Die Behandlung der Syphilis mit dem Ehrlichschen Präparat 606." *Deutsche medizinische Wochenschrift*: 36(41): 1889–924.

———. 1911. "Aus Theorie und Praxis der Chemotherapie." *Folia Serologica* 7:697–714.

Franklin-Hall, Laura R. 2015. "Explaining Causal Selection with Explanatory Causal Economy: Biology and Beyond." In *Explanation in Biology*, ed. Christophe Malaterre and Pierre-Alain Braillard, 413–38. New York: Springer.

———. 2016. "High-Level Explanation and the Interventionist's 'Variables Problem.'" *British Journal for the Philosophy of Science* 67:533–77.

Glennan, Stuart. 2017. *The New Mechanical Philosophy*. Oxford: Oxford University Press.

Griffiths, Paul E., Arnaud Pocheville, Brett Calcott, Karola Stotz, Hyunju Kim, and Rob Knight. 2015. "Measuring Causal Specificity." *Philosophy of Science* 82:529–55.

Griffiths, Paul E., and Karola Stotz. 2013. *Genetics and Philosophy: An Introduction*. New York: Cambridge University Press.

https://doi.org/10.1086/706093 Published online by Cambridge University Press

Hesslow, Germund. 1988. "The Problem of Causal Selection." In *Contemporary Science and Natural Explanation: Commonsense Conceptions of Causality*, ed. Denis J. Hilton, 11–32. New York: New York University Press.
Hippel, Peter H. von, and Otto G. Berg. 1986. "On the Specificity of DNA-Protein Interactions." *Proceedings of the National Academy of Sciences of the USA* 83:1608–12.
Huggins, David J., Woody Sherman, and Bruce Tidor. 2012. "Rational Approaches to Improving Selectivity in Drug Design." *Journal of Medicinal Chemistry* 55 (4): 1424–44.
Kaiser, Marie I. 2016. "On the Limits of Causal Modeling: Spatially-Structurally Complex Biological Phenomena." *Philosophy of Science* 83 (5): 921–33.
Kangas, Erik, and Bruce Tidor. 2000. "Electrostatic Specificity in Molecular Ligand Design." *Journal of Chemical Physics* 112 (20): 9120–31.
Liu, Yi, and Nathanael S. Gray. 2006. "Rational Design of Inhibitors That Bind to Inactive Kinase Conformations." *Nature Chemical Biology* 2:358–64.
Machamer, Peter, Lindley Darden, and Carl F. Craver. 2000. "Thinking about Mechanisms." *Philosophy of Science* 67 (1): 1–25.
Mill, John S. 1843/1981. *A System of Logic*. Toronto: University of Toronto Press.
Sarkar, Sahotra. 2005. "How Genes Encode Information for Phenotypic Traits." In *Molecular Models of Life: Philosophical Papers on Molecular Biology*, 261–83. Cambridge, MA: MIT Press.
Schasfoort, Richard B. M., Wim de Lau, Alex van der Kooi, Hans Clevers, and Gerard H. M. Engbers. 2012. "Method for Estimating the Single Molecular Affinity." *Analytical Biochemistry* 421 (2): 794–96.
Stegmann, Ulrich E. 2004. "The Arbitrariness of the Genetic Code." *Biology and Philosophy* 19 (2): 205–22.
Stotz, Karola, and Paul E. Griffiths. 2017. "Biological Information, Causality and Specificity: An Intimate Relationship." In *From Matter to Life: Information and Causality*, ed. Sara I. Walker, Paul C. W. Davies, and George F. R. Ellis, 366–90. Cambridge: Cambridge University Press.
Strebhardt, Klaus, and Axel Ullrich. 2008. "Ehrlich's Magic Bullet Concept: 100 Years of Progress." *Nature Reviews Cancer* 8:473–80.
Tabery, James, Monika Piotrowska, and Lindley Darden. 2015. "Molecular Biology." In *Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta. Stanford, CA: Stanford University. https://plato.stanford.edu/archives/sum2018/entries/molecular-biology/.
Waters, C. Kenneth. 2007. "Causes That Make a Difference." *Journal of Philosophy* 104 (11): 551–79.
Weber, Marcel. 2006. "The Central Dogma as a Thesis of Causal Specificity." *History and Philosophy of the Life Sciences* 28 (4): 595–609.
———. 2017. "Which Kind of Causal Specificity Matters Biologically?" *Philosophy of Science* 84 (3): 574–85.
Woodward, James. 2003. *Making Things Happen: A Theory of Causal Explanation*. Oxford: Oxford University Press.
———. 2010. "Causation in Biology: Stability, Specificity, and the Choice of Levels of Explanation." *Biology and Philosophy* 25 (3): 287–318.
———. 2017. "Response to Franklin-Hall and Weslake on Stability and Proportionality." Unpublished manuscript, PhilSci Archive. http://philsci-archive.pitt.edu/14117/1/Response%20to%20Franklin-Hall%20and%20Weslake.docx.

https://doi.org/10.1086/706093 Published online by Cambridge University Press