When Causal Specificity Does Not Matter (Much): Insights from HIV Treatment

Jacob P. Neal*†

Philosophers of biology generally agree that causal specificity tracks biological importance: more specific causes are more important. I argue that this correlation does not hold in much research aimed at biomedical intervention. Applying James Woodward's analysis of causal specificity to the development and design of HIV treatments, I show that drugs that are less causally specific produce better therapeutic outcomes and are more highly valued. Thus, I conclude that the importance of biological causes does not track their specificity but instead varies depending on the biological practice we consider and the goals of that practice.

1. Introduction. Woodward (2010) identifies a biologically important type of causal specificity. He defines it as the degree to which varying the state of C "exerts a fine-grained kind of control over which state of E is realized" (305). This notion of causal specificity plays an important role within what Schaffner (2016) has called the 'developmentalist challenge' of Developmental Systems Theory. Within this context, supporters of Developmental Systems Theory present the causal parity thesis as a challenge to the standard view of molecular genetics. The parity thesis claims a fundamental symmetry between the roles of DNA and nongenetic factors in organismal development (Oyama 1985; Griffiths and Gray 2004). Opponents of this view typically think there are principled reasons for maintaining a privileged causal role for DNA. Waters (2007), for instance, invokes causal specificity as a justification for privileging DNA in explanations of organismal variation since it is often a causally specific difference maker. Griffiths and Stotz (2013) respond that other developmental factors are also causally specific difference makers.

* To contact the author, please write to: Department of History and Philosophy of Science, University of Pittsburgh, 1101 Cathedral of Learning, 4200 Fifth Avenue, Pittsburgh, PA, 15260; e-mail: jacobpneal@gmail.com.

† I would like to thank Jim Woodward, Mike Dietrich, Ken Schaffner, and David Colaço as well as the audience at PSA 2018 for helpful comments on earlier drafts of this article.

Philosophy of Science, 86 (December 2019) pp. 836–846. 0031-8248/2019/8605-0002$10.00
Copyright 2019 by the Philosophy of Science Association. All rights reserved.

https://doi.org/10.1086/705510 Published online by Cambridge University Press

What is notable is that philosophers on both sides of the developmentalist challenge have implicitly accepted the notion that causal specificity tracks biological importance. I aim to show that this presumed importance of specific rather than nonspecific causes often does not hold in biomedical research that aims at therapeutic intervention. Using Woodward's analysis of causal specificity, I show that within the context of interventions on HIV infection, therapeutic agents that are less rather than more causally specific produce better therapeutic outcomes and are thus more highly valued by researchers and clinicians. Because the primary philosophical debates on causal specificity have arisen within the context of scientific explanation, philosophers have overlooked the importance of nonspecific causes in biomedical interventions.

In what follows, I first present Woodward's account of causal specificity (sec. 2). I then turn to the therapeutic context of HIV drug development and design (sec. 3). Using Woodward's three ideals of causal specificity, I assess whether HIV treatments that more closely approximate these ideals are more highly valued by biomedical researchers and clinicians. Ultimately, I conclude that less rather than more causally specific agents often lead to better therapeutic outcomes (sec. 4). Shifting from the context of explanation to intervention shows that there is no necessary correlation between the specificity of a cause and its importance. What counts as a more or less important biological cause varies depending on which aspects of biological practice we consider.

2. Causal Specificity. Woodward (2010) proposes an account of causal specificity relevant in biological contexts. This type of causal specificity (INF) attempts to capture the intuitive idea that causes that afford 'fine-grained influence' over their effects are more specific than those that do not (302). For a cause C to be specific, on this view, it must be able to assume multiple states, and when it does, it must bring about different states of the effect E. The canonical example used to distinguish relatively specific from relatively nonspecific causes, in this sense, is an analog radio with both a tuning dial and an on/off switch. Both features of the radio are causally necessary in order to listen to a radio station, but we can nevertheless distinguish how causally specific each is in determining the actual station we hear. There is an asymmetry between the causes of the radio station we hear that tracks causal specificity: the tuning dial is a relatively specific cause, whereas the on/off switch is relatively nonspecific.

Woodward (2010) develops a more precise definition of this INF-type causal specificity. He builds on a minimalist conception of causation, under which C causes E if and only if there are certain background conditions B such that if an ideal intervention were to change the value of C, then the value of E would also change. The causal relata, in this case, are construed as variables, C and E, each of which can assume a number of possible

https://doi.org/10.1086/705510 Published online by Cambridge University Press

states, and the background conditions are those conditions not explicitly represented in the causal relation, which must remain stable during an ideal intervention on C (Woodward 2003).

In order to assess the extent to which C is a more rather than less specific cause of E, the causal relation must be or approximate a function F that maps C to E. For the ideal case of causal specificity, F will (1) be a *function* that is both (2) *surjective* and (3) *injective*, or in other words, F will be a *bijective* function. Each state of C will correspond to a unique state of E and vice versa. Hence, any intervention on C will allow for exquisitely fine-grained control over the state of E. While this ideal is unlikely to be satisfied in biological cases, it provides a means to assess the relative specificity of causes. Since causal specificity comes in degrees, Woodward defines it in relation to this ideal. He proposes that C will be a more specific cause of E to the extent that F satisfies the following conditions and approximates the corresponding ideals (in italics):

1. Few states of C correspond to more than one state of E. (F approximates a *function*.)
2. Most states of E have at least one corresponding state in C. (F is nearly *surjective*.)
3. Many states of C correspond to unique states of E. (F is nearly *injective*.)

In other words, C will become a more specific cause of E as the mapping function F moves closer toward these three ideals.

We can illustrate this presentation of causal specificity diagrammatically. Consider the causal mapping meant to illustrate the ideal bijective function that satisfies all three of Woodward's conditions (fig. 1*A*). The arrows in the diagram represent the causal relation that maps specific values of C to specific values of E.[1] If an ideal intervention were performed on C to set it to state C_1, for instance, then the causal arrow shows that this intervention would lead the effect variable to take state E_1. Intervening on C thus gives us fine-grained control over which state of E is realized.

Actual biological cases will likely deviate from this bijective ideal of causal specificity in one or more ways, and we can rank the specificity of different causes on the basis of the extent to which they deviate from the three ideal conditions. First, let us consider where the mapping relations between C and E fail to be functions. In cases of this sort, there will be at least one cause state that corresponds to more than one effect state (fig. 1*B*). This case fits the radio example when C represents the on/off switch. The on/off switch can take on exactly two values, C_1 and C_2. Intervening on C to put it in the 'off'

1. Unlike a directed graph, the nodes in these diagrams are not variables but rather the individual states or values of variables.

https://doi.org/10.1086/705510 Published online by Cambridge University Press

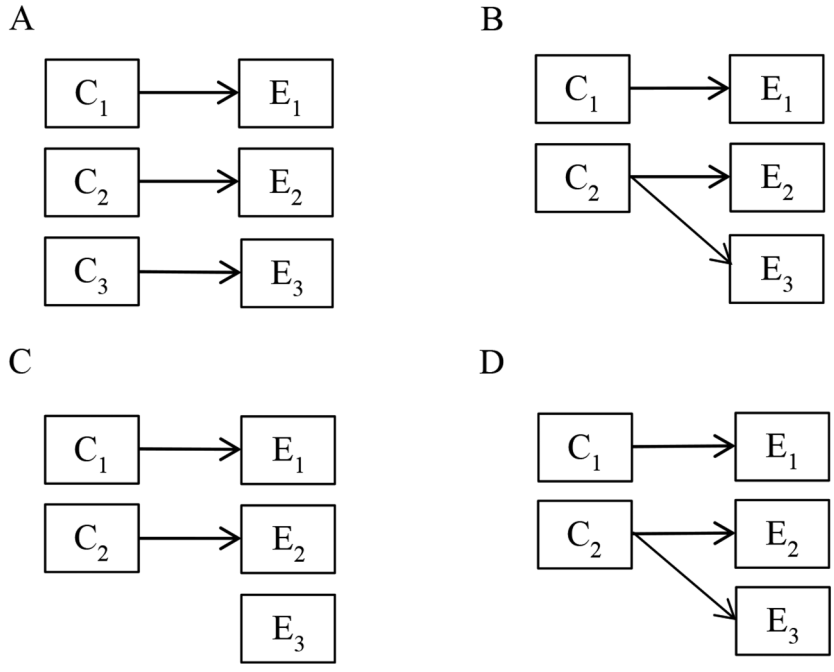


Figure 1. Causal mappings of bijective ideal of causal specificity and deviations: *A*, bijective ideal; *B*, violation of functionality; *C*, violation of surjection; *D*, violation of injection.

position, C_1, yields a determinate result, namely, E_1 or silence. However, an intervention that puts C in the 'on' position, C_2, does not fully determine the value of E. Sometimes when we turn the radio on, we will hear NPR programming (E_2), sometimes easy listening (E_3), and so on, depending on the position of the tuner dial. While the on/off switch is a cause of our hearing these stations, it is a relatively nonspecific cause with little influence over which radio station we actually hear. C will be a less rather than more specific cause of E to the extent that it has more states of C that map to multiple states of E. The more F deviates from the ideal of functionality, the less ability an intervention on C has to influence the state of E.

Second, actual biological cases might fail to be surjective. The ideal of surjection ensures that every state of E corresponds to at least one state of C. If the mapping function that relates C to E is not surjective, it means that at least one state of E is not accessible by intervention on C (fig. 1*C*). If we consider C to be the radio dial and E to be the station we hear on our speakers, then C_1 and C_2 represent discrete frequencies that correspond to

https://doi.org/10.1086/705510 Published online by Cambridge University Press

their respective stations. Tuning the dial to either of these stations will determine whether we hear NPR or the easy listening station. In contrast, E_3 represents an effect state that is inaccessible by modulating C, namely, silence. As the diagram shows, the tuner is specific in that it will allow us to intervene to choose one of the two stations, but it is unable to switch the radio off to bring about silence. This follows from condition 2. The more states of E that are not the image under F of any states of C, the less specific the causal relation between C and E.

Third, a biological case of interest will deviate from the ideal case of causal influence insofar as it fails to be injective. The injective, or 1-to-1, function requires that each cause state corresponds to a unique effect state. To the extent that F is not injective, a given state of E will correspond to more than one state of C, resulting in a causal relation that is less specific (fig. 1*D*). According to Woodward, of the three deviations from the ideal case, those captured in conditions 1 and 2 are the ones that most affect causal specificity. That is, the closer the mapping relation between C and E is to a surjective function, the more control we will have over the state of E by intervening on C. This last deviation, where F fails to be injective, is less important to the notion of causal specificity because this causal redundancy will not hinder our ability to control which state of E is realized.

By considering the extent to which causal relations deviate from the ideal conditions expressed in 1–3 while also giving additional weight to deviations from 1 and 2, we can distinguish relatively nonspecific causes from relatively specific ones. Woodward's formulation thus offers a principled way to compare the specificity of different causes.

3. Causal Specificity and HIV Antiretroviral Development. The recent philosophical interest in causal specificity has tended to focus on explanations of normal biological function. Although providing explanations of biological phenomena is an important goal of biological research, it is far from the only one. Here, I shift the focus from explanation to intervention. Using the Woodwardian framework, I assess the role of causal specificity in biomedical research, specifically HIV drug design. My analysis of this type of therapeutic intervention suggests a different picture of the correlation between specificity and importance. In this case, developing therapeutic interventions that are more causally specific and exert fine-grained control over their effects is not the goal. Instead, I argue that although biomedical researchers consider the relative specificity of interventions, they often privilege the less specific rather than more specific therapeutic interventions.

3.1. Violations of Functionality: One Cause → Many Effects (Condition 1). Most antiretrovirals (ARVs) that target HIV violate the ideal of causal specificity expressed in condition 1: the mapping relation between

https://doi.org/10.1086/705510 Published online by Cambridge University Press

C and E is not a function because a single cause state has multiple effect states. This empirical fact prevents the fine-tuning of a single causal variable to yield unique effects in different circumstances. Biomedical researchers, however, do not view this lack of fine-grained causal specificity to be problematic but instead consider it a virtue.

To understand why, we must consider certain features of HIV replication and the aim of therapeutic intervention.[2] HIV is a retrovirus that consists of a viral capsid, which contains an RNA genome and reverse transcriptase (RT). Like all retroviruses, HIV hijacks the cellular machinery of the host cell to express its viral proteins. The virus replicates very rapidly but with relatively low fidelity. Hence, any HIV infection will consist of a population of virions with slightly different genotypes, which gives rise to variability in viral proteins. To be effective, HIV treatments must be able to slow viral replication across a range of HIV variants. Researchers engaged in ARV design and development do not want their drugs to be too specific. Instead, they want to have individual drugs capable of targeting many different viral variants. The scientific literature on HIV is replete with examples showing that, in this context, relatively less specific ARVs are more powerful and better therapeutic agents than more specific ones. In fact, within the history of HIV drug development, many second- and third-generation ARVs were developed with the express purpose of decreasing the specificity of an earlier drug.

Consider, for example, nucleoside reverse transcriptase inhibitors (NRTIs), the first class of drugs designed to target HIV RT. These ARVs are nucleoside analogues that bind at the active site of the RT during transcription and result in chain termination. Viral mutations, such as M184V, that increase steric hindrance at the active site prevent certain NRTIs from binding (Yin et al. 2006). This mutation adversely affects the efficacy of compounds, such as lamivudine, that have particularly rigid structures. These rigid compounds are unable to adapt to the change in shape in the active site caused by the viral mutation. Tenofovir, a more recently approved NRTI, has the ability to bind both wild type and M184V mutant RTs. X-ray crystal structures of tenofovir bound to RT show that this compound can assume two different conformations. It is speculated that this greater flexibility of the compound enables it to be an effective inhibitor of both the wild type and mutant RTs. Drug design in this case has proceeded from more specific to less specific compounds: first generation NRTIs such as lamivudine have a causal influence over fewer viral variants than second-generation NRTIs such as tenofovir. This example of the development of NRTIs suggests that within the context

2. The information about HIV replication and treatment is from Wlodawer (2002), Yin, Das, and Mitsuya (2006), and Mehellou and De Clercq (2010).

https://doi.org/10.1086/705510 Published online by Cambridge University Press

of HIV intervention, therapeutic agents that are less rather than more specific are more highly valued.

Woodward's notion of INF-type causal specificity applies to this empirical example from HIV drug development research. According to Woodward, condition 1 is underwritten by the intuition that a single cause that leads to multiple effects exerts less influence over its effects than a cause with a single effect, since the effects cannot be modulated separately. Assessing the relative deviation from the ideal of condition 1, we are justified in calling one ARV more causally specific relative to another. Consider the causal mappings for the two NRTIs just discussed (fig. 2).

The diagram represents the causal relation between each NRTI and four different HIV variants: wild type (wt), L74V, M184V, and Q151M. The cause state is the appropriate therapeutic dose of the stated ARV, and the effect states are the specific viral populations. An arrow indicates the presence of a particular causal relation between the two states, namely, the inhibition of the HIV RT variant by the ARV. As we can see from the diagram, tenofovir affects all four viral populations, while lamivudine only affects two. Tenofovir thus deviates further from the ideal of functionality than does lamivudine, since a single cause state yields four different effect states instead of only two. Of these two therapeutic agents, tenofovir is less causally specific than lamivudine, other things being equal. Tenofovir exhibits less fine-grained control over the effect on individual viral populations, since its presence leads to a number of different effect states.

Following the current literature in philosophy of biology, one might expect the more specific ARV, lamivudine, to be more highly valued or more important, in some sense, than the less causally specific drug, tenofovir. After all, the former exerts more fine-grained causal influence over its effects than does the latter, since it more closely approximates the ideal of functionality. But in fact, all things being equal, tenofovir is the better HIV treatment, precisely because it leads to multiple desired effects. Biomedical researchers and clinicians who aim to control HIV infection have no reason to modulate the levels of wt HIV independent of the various RT mutants.

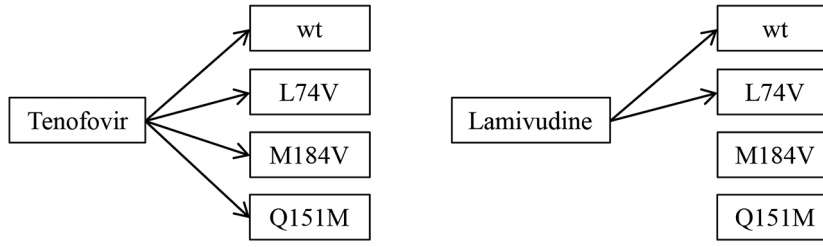


Figure 2. Causal mapping for antiretrovirals and HIV reverse transcriptase variants.

https://doi.org/10.1086/705510 Published online by Cambridge University Press

Thus, the relative reduction in specificity from lamivudine to tenofovir enables the latter to cause multiple effects, all of which are desirable therapeutic outcomes.

Biomedical and medicinal chemistry research privileges ARVs that are less rather than more specific for therapeutic and pragmatic reasons. From a therapeutic standpoint, the clinician attempting to control an HIV infection within an individual does not know all the exact HIV variants present. Thus, if two ARVs are available, one that suppresses viral replication for wt HIV as well as a large number of mutants and one that is only efficacious against wt HIV, the former ARV is the better treatment. Another therapeutic consideration involves toxicity. If it were possible to design ARVs that approximated the ideal of functionality, then the effective treatment of any HIV infection would likely require many more ARVs than are currently necessary. The greater combination would likely lead to greater risk of cytotoxicity and deleterious—or even fatal—side effects. From a pragmatic standpoint, developing individual ARVs that have multiple desirable effects is more cost effective. A drug, such as tenofovir, will be a good candidate ARV treatment for a larger population of infected individuals than lamivudine, since it is efficacious against a larger number of HIV variants. With regard to the ideal of functionality, HIV researchers tend to privilege less rather than more causally specific therapeutic agents.

3.2. Violations of Surjection: Effects without Associated Causes (Condition 2). Of the three axes of causal specificity expressed in Woodward's three conditions, only the violation of the ideal of surjection is not valued by those who wish to intervene to control HIV infection. For the causal mapping to be surjective, every effect state must have a corresponding cause state. When the causal relation fails this criterion, then there are certain effect states that are not accessible by tuning the cause state. All current HIV treatments violate this ideal to varying extents and thus fail to be maximally causally specific. Along this axis of causal specificity, however, ARVs that deviate more from the ideal (i.e., ARVs that exert less causal influence over their effects) are worse therapeutic agents.

To see why this is so, consider again the causal relation between lamivudine and the four possible effect states depicted in figure 2. The causal arrows linking this drug to wt HIV and L74V show that the presence of this drug inhibits the replication of these two viral variants. No causal arrows link this drug to the other two viral variants, representing the fact that lamivudine is not causally related to these two HIV mutants. In other words, these two effect states are inaccessible to the cause state. No matter how we fine-tune the cause variable, by increasing the concentration of lamivudine, for instance, we are unable to bring about the inhibition of the M184V or Q151M mutants, and this is detrimental to causal specificity.

https://doi.org/10.1086/705510 Published online by Cambridge University Press

With regard to therapeutic intervention into HIV infection, the inability to control mutant viral populations is very much a concern. The deviation from the ideal of surjection and the corresponding loss of causal specificity, in this case, requires biomedical researchers and clinicians to develop additional ARVs that exert causal influence over the HIV mutants not inhibited by other ARVs. Loss of causal specificity of this sort hinders our ability to effectively intervene to suppress viral replication. Hence, from the perspective of therapeutic interventions, a more powerful ARV will have fewer viral variants causally inaccessible to it. With regard to the ideal of surjection, biomedical researchers privilege therapeutic interventions that are more rather than less causally specific.

3.3. Violations of Injection: Many Causes → One Effect (Condition 3). Effective HIV treatments deviate from the third ideal of causal specificity, since the causal mapping violates the ideal of injection. In an injective mapping, only one cause state is associated with each effect state. But, in the case of HIV treatment, many different cause states, namely, different ARVs, intervene to bring about a single effect, the inhibition of HIV replication. This deviation from injection suggests that current HIV treatments are less rather than more specific causes, with respect to condition 3. Far from being a shortcoming, however, this deviation is an important feature of HIV drug design and treatment.

Biomedical and clinical research has shown that effective HIV treatment typically requires at least three different ARVs from two different drug classes (DHHS 2014). In other words, three causes are therapeutically necessary to bring about the desired effect of indefinite suppression of viral replication. This relative lack of causal specificity, however, is the least detrimental to causal specificity. As Woodward notes, it will be "less damaging" to our ability to control E by fine-tuning C "if several different states of C lead to the same state of E" than if the causal relation deviates from the ideals of functionality or surjection (2010, 305). Woodward's claim is that despite this casual redundancy, interventions that set C to a specific state will still be able to pick out specific states of E. Along this axis of causal specificity, a cause exerts a greater influence over its effect if fewer different states of C are mapped onto the same state of E.

When we consider HIV treatments, we see many ARVs that aim to bring about the same effect state. Figure 3 shows the high degree of causal redundancy in this domain. On the left, we see a subset of NRTIs currently approved for the treatment of HIV infection. Each one of these ARVs binds the viral RT active site to prevent transcription of viral RNA into DNA. In this case, tuning the causal state to any one of these ARVs will bring about the same effect state: inhibition of viral RT. According to the ideal of injection, this causal structure is less specific than if there were only one route to

https://doi.org/10.1086/705510 Published online by Cambridge University Press

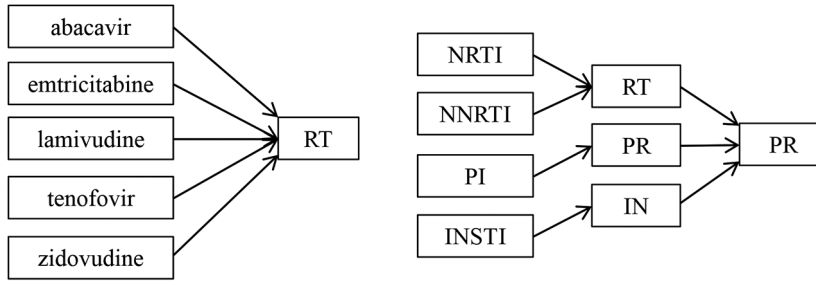


Figure 3. Many NRTIs inhibit RT (*left*), and four different HIV drug classes inhibit different viral proteins to inhibit HIV replication (*right*). Drug class abbreviations: NRTI (nucleoside reverse transcriptase inhibitor), NNRTI (nonnucleoside reverse transcriptase inhibitor), PI (protease inhibitor), and INSTI (integrase strand transfer inhibitor). Viral enzyme abbreviations: RT (reverse transcriptase), PR (protease), and IN (integrase).

RT inhibition. Importantly, this causal mapping only highlights a small portion of the entire web of causal relations shown on the right. The ultimate effect of all these therapeutic interventions is the inhibition of HIV replication. However, as the diagram shows, the different drug classes have intermediate effects based on the particular viral enzyme they target. Each one of these drug classes includes multiple individual ARVs. Thus, the causal redundancy of HIV treatments is much greater than actually depicted in the diagram.

This causal redundancy of HIV treatments is inversely correlated with causal specificity, according to the ideal of injection. The continual development of novel ARVs contributes to this redundancy, as does the fact that effective treatment requires at least three distinct therapeutic agents to maintain viral suppression. From the perspective of therapeutic intervention, this deviation from the ideal of injection in the causal mapping between ARVs and HIV inhibition is a desirable feature. It is precisely this relative lack of causal specificity—this causal redundancy—that enables us to successfully intervene to control HIV infection.

3.4. Bijection Is Not the Ideal in HIV Intervention. A maximally specific cause is one in which the function that maps the cause to its effect is bijective. According to Woodward, bijection is an ideal of specificity that real biological causes can only more or less approximate. Whereas causes that more closely approximate this ideal of specificity play an important role in biological explanations, we have seen that within the context of HIV treatment and drug development, researchers tend to value therapeutic agents that are less causally specific. HIV researchers do not aim for the ideal of bijection;

https://doi.org/10.1086/705510 Published online by Cambridge University Press

they do not privilege ARVs that more closely hew to the ideals of specificity expressed in Woodward's three conditions. In this case, deviation from the ideal is considered a virtue.

4. Conclusion. To date, causal specificity has typically been addressed within the context of scientific explanation, and philosophers of biology are generally in agreement that the importance of a biological cause correlates with its degree of specificity. I have sought to show that researchers engaged in therapeutic intervention often value those causes that are less specific. Whereas biologists focused on offering explanations may privilege more specific causes, researchers intent on intervening on human diseases, such as HIV, often privilege less specific causes. Thus, the importance of biological causes does not track their specificity but instead is highly relative to the particular biological practice we consider and its goals.

REFERENCES

DHHS (Department of Health and Human Services). 2014. "Guidelines for the Use of Antiretroviral Agents in HIV-1-Infected Adults and Adolescents." DHHS, Washington, DC.

Griffiths, P., and R. D. Gray. 2004. "The Developmental Systems Perspective: Organism-Environment Systems as Units of Selection." In *Phenotypic Integration: Studying the Ecology and Evolution of Complex Phenotypes*, ed. M. Pigliucci and K. Preston. Oxford: Oxford University Press.

Griffiths, P., and K. Stotz. 2013. *Genetics and Philosophy: An Introduction*. Cambridge: Cambridge University Press.

Mehellou, Y., and E. De Clercq. 2010. "Twenty-Six Years of Anti-HIV Drug Discovery." *Journal of Medicinal Chemistry* 53 (2): 521–38.

Oyama, S. 1985. *The Ontogeny of Information: Developmental Systems and Evolution*. Durham, NC: Duke University Press.

Schaffner, K. F. 2016. *Behaving: What's Genetic, What's Not, and Why Should We Care?* Oxford: Oxford University Press.

Waters, C. K. 2007. "Causes That Make a Difference." *Journal of Philosophy* 104 (11): 551–79.

Wlodawer, A. 2002. "Rational Approach to AIDS Drug Design." *Annual Review of Medicine* 53:595–614.

Woodward, J. 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.

Yin, P. D., D. Das, and H. Mitsuya. 2006. "Overcoming HIV Drug Resistance through Rational Drug Design." *Cellular and Molecular Life Sciences* 63 (15): 1706–24.

https://doi.org/10.1086/705510 Published online by Cambridge University Press