

# Observation Selection Effects and Discrimination-Conduciveness

William Roche\* and Elliott Sober†

\* Department of Philosophy, Texas Christian University, Fort Worth, TX, USA,  
w.roche@tcu.edu

† Department of Philosophy, University of Wisconsin at Madison, Madison, WI, USA,  
ersober@wisc.edu

*Abstract:* We conceptualize *observation selection effects* (OSEs) by considering how a shift from one process of observation to another affects discrimination-conduciveness, by which we mean the degree to which possible observations discriminate between hypotheses, given the observation process at work. OSEs in this sense come in degrees and are causal, where the cause is the shift in process, and the effect is a change in degree of discrimination-conduciveness. We contrast our understanding of OSEs with others that have appeared in the literature. After describing conditions of adequacy that an acceptable measure of degree of discrimination-conduciveness must satisfy, we use those conditions of adequacy to evaluate several possible measures. We also discuss how the effect of shifting from one observation process to another might be measured. We apply our framework to several examples, including the ravens paradox and the phenomenon of publication bias.

*Keywords:* evidence, fine-tuning argument, likelihoods, measure-sensitivity, observation selection effects, publication bias, ravens paradox.

## 1 Introduction

It is widely recognized that the process used to make observations often has a significant effect on how hypotheses should be evaluated in light of those observations. Arthur Stanley Eddington (1939, Ch. II) provides a classic example.<sup>1</sup> You're at a lake and are interested in the size of the fish in it. You know, from testimony, that at least some of the fish in the lake are big (i.e., at least 10 inches long), but beyond that you're in the dark. You devise a plan of attack: get a net and use it to draw a sample of fish from the lake. You carry out your plan and observe:

O<sub>100</sub>:      100% of the fish in the net are big.

---

<sup>1</sup> Here we introduce some minor modifications to Eddington's example.

You think that  $O_{100}$  is helpful evidence, since it discriminates between at least the following two hypotheses of interest:

$H_{100}$ : 100% of the fish in the lake are big.

$H_{50}$ : 50% of the fish in the lake are big.

In particular,  $O_{100}$  seems to you to discriminate between these two hypotheses because it favors  $H_{100}$  over  $H_{50}$ . But then you see that these first impressions are wrong because you notice that the holes in your net are so big that it can't hold fish that are medium-sized (i.e., at least 5 inches long but less than 10 inches long) or small (i.e., less than 5 inches long). You realize, furthermore, that had you used a net with small holes, the situation would have been different.<sup>2</sup> Then  $O_{100}$  would discriminate between  $H_{100}$  and  $H_{50}$ , since the observation really would favor the first hypothesis over the second.

Your using one net rather than the other makes an *epistemic* difference. Given your epistemic goal of wanting to test the two hypotheses against each other, you are better off using one net rather than the other. This is a paradigm case of an *observation selection effect* (an OSE).

The net with big holes has a deficiency, but this is only because you are considering hypotheses about the *size* of the fish in the lake. Had you been interested in the *color* of the *big* fish in the lake, the net with big holes would have worked just fine. Your “interests” can be taken out of the picture by putting the point like this:

**Relativity:** Whether an observational process  $p$  induces an OSE is relative to a set of competing hypotheses  $\Gamma_{\text{hyp}} = \{H_1, H_2, \dots, H_n\}$ .

In the initial fish example, the OSE involved is *extreme*. A net with big holes is totally useless, whereas a net with small holes works perfectly, given that you observe  $O_{100}$  and are considering hypotheses  $H_{100}$  and  $H_{50}$ . However, there are plenty of observation processes that fall in between. To illustrate the existence of non-extreme cases, we move to a slightly more complicated example. Instead of focusing on just two nets (one with big holes, the other with small), let's add a third – a net with medium holes. It retains fish that are medium and large, but not fish that are small. And we'll now consider three possible observations. Each involves catching a single fish, which can be big, medium, or small. Call these  $O_b$ ,  $O_m$ , and  $O_s$ ; they comprise a three-membered set of possible observations  $\Gamma_{\text{obs}} = \{O_b, O_m, O_s\}$ . And, finally, in this new example, we'll set  $H_{100}$  and  $H_{50}$  aside and consider two new hypotheses, namely:

$H_{50-50-0}$ : 50% of the fish in the lake are big, 50% are medium, and 0% are small.

$H_{5-5-90}$ : 5% of the fish in the lake are big, 5% are medium, and 90% are small.

---

<sup>2</sup> In saying that the net has “small” holes, we mean that it holds onto fish that are small, medium, and large, once they swim into the net.

There's a clear sense in which the net with small holes is better than the net with medium holes, which in turn is better than the net with large holes. If you use the net with small holes, all three possible observations can occur and each discriminates between the two hypotheses. But use a net with medium holes, and only two of the possible observations can occur. Each of these favors  $H_{50-50-0}$  over  $H_{5-5-90}$ , but the third possible observation,  $O_s$ , cannot occur. The net with medium-size holes therefore has a defect, but it is better than the net with large holes, since the latter allows only one observation, and that observation fails to discriminate between the two hypotheses. This leads to the following conclusion:

**Degree:** An observation process  $p$  has a degree of discrimination-conduciveness, by which we mean the degree to which possible observations discriminate between hypotheses, given that  $p$  is the process at work.

This concept of discrimination-conduciveness has the following characteristic:

**Three-Place Relation:** Discrimination-conduciveness is a relation among an observation process  $p$ , a set of hypotheses  $\Gamma_{hyp}$ , and a set of possible observations  $\Gamma_{obs}$ .

Discrimination-conduciveness comes in degrees, denoted by “**DDC**( $\Gamma_{hyp}, \Gamma_{obs} | p$ ),” which allows us to state a formal point about the character of OSEs:

**Contrastive:** The OSE that an observational process  $p$  induces on hypotheses  $\Gamma_{hyp}$  and possible observations  $\Gamma_{obs}$  should be understood by comparing **DDC**( $\Gamma_{hyp}, \Gamma_{obs} | p$ ) and **DDC**( $\Gamma_{hyp}, \Gamma_{obs} | p^*$ ), for some  $p^* \neq p$ .

OSE is a causal concept, and its causal character is captured by the idea that shifting from one observation process to another can make a difference in the degree to which possible observations discriminate between hypotheses. Here we are using a familiar fact about causal claims. To say that  $C$  caused  $E$  is often shorthand for saying that it was  $C$  rather than  $C^*$  that caused  $E$  rather than  $E^*$ .<sup>3</sup> What was the effect of using a net with medium holes in the fishing example just discussed? This question is under-specified. Do you mean the effect of using a net with medium holes rather than a net with large holes? Or do you mean the effect of using a net with medium holes rather than a net with small holes? If the former, using the net with medium holes improved discrimination-conduciveness. If the latter, using that net reduced discrimination-conduciveness.

When there is an OSE, the presence of one observation process rather than another makes a difference in discrimination-conduciveness. However, observations can fail to discriminate between competing hypotheses even though the use of one observation process rather than another makes no difference. Suppose hypotheses  $H$  and  $H^*$  are *observationally equivalent* to each other. You observe that  $O$ , and then notice the inevitable:  $O$  fails to discriminate between  $H$

---

<sup>3</sup> For discussion of “contrastive causation,” and for references, see Schaffer (2005).

and  $H^*$ . However, this is *not* an effect, strictly speaking, of whatever observation process led to  $O$ .<sup>4</sup> For, if  $H$  and  $H^*$  are observationally equivalent to each other, no alternative observation process could have done any better. Even so, the OSE concept and the concept of observational equivalence both address cases in which  $\mathbf{DDC}(\Gamma_{\text{hyp}}, \Gamma_{\text{obs}} | p)$  takes the minimum possible value.

Because discrimination-conduciveness comes in degrees, it will be important to figure out how discrimination-conduciveness should be measured. We need to be able to say *how much* the shift from observation process  $p$  to observation process  $p^*$  affects the ability of observations to discriminate between hypotheses. This quantitative question is not only theoretically important; it also may have important practical applications. For example, nets with small holes are more discrimination-conducive than nets with medium holes, which in turn are more discrimination-conducive than nets with big holes. If the first type of net is more expensive than the second, and the second is more expensive than the third, is it worth your while to buy a more expensive net? It will help answer this question if you can characterize how discrimination-conducive each type of net is.

For the reader's convenience, we state definitions in the accompanying table of the three main abbreviations we use in what follows. Each represents a concept that stands in need of clarification. Later items on the list build on earlier ones.

<b>Abbreviations</b>	
<b><math>d(H_1, H_2, O   P)</math></b>	the degree to which an observation proposition $O$ discriminates between hypotheses $H_1$ and $H_2$ , given that $P$ is a true appropriate description of the observer's observation process $p$
<b><math>\mathbf{DDC}(\Gamma_{\text{hyp}}, \Gamma_{\text{obs}}   p)</math></b>	the degree to which the members of a set $\Gamma_{\text{obs}}$ of observation propositions discriminate between the members of a set $\Gamma_{\text{hyp}}$ of hypotheses, given that the observer's observation process is $p$
<b><math>\mathbf{OSE}(\Gamma_{\text{hyp}}, \Gamma_{\text{obs}}   p_1 \text{ rather than } p_2)</math></b>	the degree to which the observer's using observation process $p_1$ rather than $p_2$ affects the degree to which the members of the observation

---

<sup>4</sup> We will often write as though observation processes yield *propositions* as outputs as opposed to *perceptual beliefs* with propositional contents. But, strictly speaking, they don't do that; they yield perceptual beliefs, not propositions, as outputs.

	set $\Gamma_{\text{obs}}$ discriminate between the members of the hypothesis set $\Gamma_{\text{hyp}}$
--	--

We will proceed as follows. In Section 2, we describe three alternative conceptions of OSEs and compare them to the one we endorse. In Section 3, we clarify the idea of hypothesis discrimination, and propose an adequacy condition on measures of degree of discrimination-conduciveness. We call our proposed condition “**Minimality**.” In Section 4, we consider three candidate measures of degree of discrimination-conduciveness. Two are inspired by conceptions of OSEs described in Section 2, and the third is inspired by yet another conception in the literature. We argue that each of these three candidate measures fails to meet **Minimality** and so is inadequate.<sup>5</sup> In Section 5, we describe a schema for measuring degree of discrimination-conduciveness, and show that any instance of it meets **Minimality**. We also provide two *prima facie* plausible measures, and compare them to several alternatives. In Section 6, we turn to the related issue of how degree of OSE should be measured. In Section 7, we put our general framework to work, applying it to an old chestnut in Bayesian confirmation theory – the ravens paradox. In Section 8, we do the same but for a newer topic – publication bias.<sup>6</sup> In Section 9, we offer some concluding comments.

## 2 Alternative Conceptions of OSEs

We use the term “OSE” to denote cases in which the presence of one observation process rather than another makes a difference in the degree to which observations discriminate between competing hypotheses. However, the term has sometimes been used with different meanings.

For example, Nick Bostrom (2002) uses “OSE” to refer to *a bias due to an observer’s mere existence*. He writes:

In these examples, a selection effect is introduced by the fact that the instrument you use to collect data (a fishing net, a mail survey, preserved trading records) samples only from a proper subset of the target domain. Analogously, there are selection effects that arise not from the limitations of some measuring device but from the fact that all observations require

---

<sup>5</sup> Since, in part, the measures in question are merely *inspired by*, as opposed to *taken from*, various writings in which “OSE” is used, our objections to them aren’t meant to tell against anything in the writings in question.

<sup>6</sup> These are just two of many potential applications of our general framework. For other potential applications, see, for example, Roush (2003) on Kant’s “Copernican revolution,” Ćirković, Sandberg, and Bostrom (2010) on estimating the risk of human extinction, Titelbaum (2010) on “no-lose” epistemologies, and Dawid (1976) on data-bases and medical diagnosis.

the existence of an appropriately positioned observer. Our data is filtered not only by limitations in our instrumentation but also by the precondition that somebody be there to “have” the data yielded by the instruments (and to build the instruments in the first place). *The biases that occur due to that precondition—we shall call them observation selection effects—are the subject matter of this book.* (Bostrom 2002, p. 2, emphasis added)

Another use of “OSE” that differs from ours comes from Darren Bradley (2011), who uses the term to refer to *an effect of a sampling procedure on the inferences that can be drawn on the basis of samples obtained via that procedure.* He writes:

Whenever a sample is drawn from a population, some particular method must be used. This method is the selection procedure. *The effect this has on the inference is the (observation) selection effect.* Eddington's ([1939]) classic example involves fishing with a net. If we catch a sample of fish from a lake, and all the fish in the sample are bigger than six inches, this appears to confirm the hypothesis that all the fish in the lake are bigger than six inches. But if we then find out that the net used cannot catch anything smaller than six inches due to the size of its holes, the hypothesis is no longer confirmed. *So the inference depends on the method of obtaining the sample, i.e. on the selection procedure.* (Bradley 2011, p. 324, emphasis added)

A third construal is presented by Manson (2003, 2009) and White (2003), who think that OSEs arise precisely when the observation process ensures that some possible observation cannot occur.

We see no substantive disagreement among Bostrom, Bradley, Manson, and White and ourselves if each is simply stipulating what he means by “OSE.” They have their usages and we have ours. However, even if stipulation is the name of the game,<sup>7</sup> it is important to be clear on how our OSE concept differs from these others.

When you fish with a net that has big holes, the OSE derives from the net used, not from the mere existence of an observer. For us, this example is a paradigm case of an OSE; according to Bostrom’s stipulation, it is nothing of the kind.

To see how our picture differs from Bradley’s, suppose you enter a room and want to find out whether the walls are red or white. You carefully scan each wall in its entirety and note that the walls look red. You take this datum to favor the hypothesis that the walls are red over the

---

<sup>7</sup> An alternative to stipulation is *explication* (see Carnap 1950, Maher 2007, Olsson 2015, and Schupbach forthcoming); here you revise a concept already in use with the goal of making it clearer and more precise. Explications can be undertaken for concepts that are part of a widely used natural language, but they also have their place for concepts used by narrower populations – e.g. groups of scientists or philosophers. We think it makes sense to explicate the OSE concept, but won’t defend the claim that our account of OSEs is a good explication.

hypothesis that they are white. But then you learn that the walls are lit by red lights in such a way that the walls would look red whether they were red or white. Call this case “RED LIGHTS.” It does not involve a literal sampling of objects from a population of objects (as was true in the fishing examples), so it does not involve an OSE in Bradley’s sense.<sup>8</sup> However, it is an OSE in our sense, since the fact that there are red lights (rather than normal lighting) diminishes the ability of your sense impressions (or propositions describing them) to discriminate between competing hypotheses.

Turning now to Manson and White, we note that our three fish stories align well with their usage. If the fish in the lake come in three categories – big, medium, and small – and you’re going to catch just one fish, then there are three possible observations if you use a net with small holes, two possible observations if you use a net with medium holes, and just one possible observation if you use a net with large holes. Increasingly severe OSEs reflect more and more possible observations receiving a probability of zero.

Even so, there are other examples of OSEs in our sense in which shifting from one observation process to another leaves unchanged the number of possible observations that can occur. Suppose you’re a doctor, and Joe is your patient. You want to test him for tuberculosis, so you use your tuberculosis test kit *k*. There are two hypotheses about Joe (he has tuberculosis, or he does not) and two possible observational outcomes (positive and negative). Suppose this test kit has small error probabilities. This means that if Joe has tuberculosis, he’ll almost certainly have a positive test result, and if he doesn’t have tuberculosis, he’ll almost certainly have a negative test result. Now let’s compare test kit *k* with an alternative test kit *k\**. When the latter is used on a given patient, a randomizing device inside the kit flips a fair coin. If the coin comes up heads, *k\** says that the patient has tuberculosis. If the coin doesn’t come up heads, *k\** says that the patient does not have tuberculosis. Call this case “TUBERCULOSIS.” Shifting from test kit *k* to test kit *k\** induces an OSE in our sense, but neither of the possible observations has its probability plunge to zero.<sup>9</sup>

---

<sup>8</sup> We don’t deny that there are *models* of this example according to which your experience is the result of sampling from a population of possible experiences (see Bradley 2011). However, these are mere models and shouldn’t be taken literally.

<sup>9</sup> Here’s another example that leads to the same conclusion. People with normal binocular vision can detect the presence of objects in a visual field of about 180 degrees when their heads are stationary. Strokes and other injuries can cause peripheral blindness so that the chance of detecting objects in the periphery (i.e., outside of, say, 150 degrees) goes to zero. This would induce an OSE both in our sense and in the sense of ruling out possible observations. But now imagine a less extreme case where a stroke or other injury merely lowers the probability of detecting objects in the periphery to some value close to but greater than zero. This would induce an OSE in our sense but not in the sense of ruling out possible observations.

Although our usage of “OSE” isn’t universal, it isn’t entirely idiosyncratic, either. In fact, it is inspired by the meaning that Eddington (1939) assigns to “selective effect” (and “selection”). He doesn’t restrict selective effects in his sense to effects of a sampling procedure, as is clear from the following passages:

It seems appropriate to call the philosophical outlook that we have here reached selective subjectivism. “*Selective*” is to be interpreted *broadly*. *I do not wish to assert that the influence of the procedure of observing on the knowledge obtained is confined to simple selection, like passing through a net.* (Eddington 1939, p. 26, emphasis added)

In introducing subjective selection (p. 16), I attributed it to “the sensory and intellectual equipment” used in obtaining observational knowledge. The inclusion of intellectual equipment may have seemed surprising. It is easy to see that our *sensory equipment* has a *selective effect*—that the nature and extent of our knowledge of the external world must be largely conditioned by its *lines of communication with consciousness, provided by our sense organs*. ... (Eddington 1939, p. 114, emphasis added)

By “subjective” he seems to mean something like “dependent on an observation process.”<sup>10</sup> Eddington (1939, Ch. II) also notes that selective effects are relative to hypothesis sets. Our usage of “OSE” is identical to his usage of “selective effect” on each of these two fronts.

If you think that nets with big holes can induce OSEs, that there is an OSE in the case of RED LIGHTS, and that the same is true in TUBERCULOSIS, then you’ll be inclined to conclude that the stipulated meanings that Bostrom, Bradley, Manson, and White introduce for their OSE concepts are too narrow. But perhaps you have no preconceptions concerning how this terminology should be used. In that case, you perhaps will grant that OSEs in Bostrom’s, Bradley’s, Manson’s, and White’s senses are epistemically interesting. It’s crucial to note, however, that observation processes have effects on hypothesis discrimination additional to the ones they identify. OSEs in *our* sense cover effects of exactly that sort, and are worthy of careful consideration in and of themselves. We now turn to the task of fleshing out the idea that OSEs are matters of degree and that they involve shifts from one observation process to another that affect discrimination-conduciveness.

---

<sup>10</sup> See, for example, Eddington (1939, p. 66). This way of understanding “subjective” is compatible with what we say in Section 3 about “true” descriptions of observation processes and about “objective” probabilities.



### 3 Minimality

In describing our three examples of net fishing, we used the terms “favoring” and “discrimination,” but we did not explain what we mean by these terms. It now is time for us to put some cards on the table.

When you fish with a net that has big holes (FISH-B),  $O_{100}$  fails to discriminate between  $H_{100}$  and  $H_{50}$  in that

$$(3.1) \quad \Pr(O_{100} \mid P_B \ \& \ H_{100}) = \Pr(O_{100} \mid P_B \ \& \ H_{50}).$$

Here “ $P_B$ ” is a true appropriate description of your observation process in FISH-B.<sup>11</sup> In contrast, when you fish with a net that has small holes (FISH-S),  $O_{100}$  discriminates between  $H_{100}$  and  $H_{50}$  in that

$$(3.2) \quad \Pr(O_{100} \mid P_S \ \& \ H_{100}) > \Pr(O_{100} \mid P_S \ \& \ H_{50}),$$

where “ $P_S$ ” is a true appropriate description of your observation process in FISH-S. There’s a likelihood equality in FISH-B and a likelihood inequality in FISH-S.

We are assuming here that hypothesis discrimination is to be understood as follows:

$$(3.3) \quad \text{For any hypotheses } H_1 \text{ and } H_2, \text{ observation } O, \text{ and observation process } p, \text{ where } P \text{ is a true appropriate description of } p, O \text{ discriminates between } H_1 \text{ and } H_2 \text{ given } P \text{ precisely when } \Pr(O \mid P \ \& \ H_1) \neq \Pr(O \mid P \ \& \ H_2).^{12}$$

This condition is inspired by the law of likelihood:

$$(3.4) \quad \text{For any hypotheses } H_1 \text{ and } H_2, \text{ observation } O, \text{ and observation process } p, \text{ where } P \text{ is a true appropriate description of } p, O \text{ favors } H_1 \text{ over } H_2 \text{ given } P \text{ precisely when } \Pr(O \mid P \ \& \ H_1) > \Pr(O \mid P \ \& \ H_2).^{13}$$

---

<sup>11</sup> We explain below what we mean by “appropriate.”

<sup>12</sup> This is distinct from the following:

For any hypotheses  $H_1$  and  $H_2$ , and observation  $O$ ,  $O$  discriminates between  $H_1$  and  $H_2$  precisely when  $\Pr(O \mid H_1) \neq \Pr(O \mid H_2)$ .

Here there’s no explicit mention of an observation process, though the probability function may take that relevant fact into account. But then again, it may not. It’s clear from cases like FISH-B and FISH-S that the role of observation processes should be made explicit. This can also be done by subscripting the probability function, like this:  $\Pr_p(-)$ .

<sup>13</sup> This is our preferred formulation of the law of likelihood. Hacking (1965) uses a simpler formulation. Tricky issues sometimes arise when the law is used in a given case. See Weisberg

Propositions (3.3) and (3.4) together entail that if  $O$  favors  $H_1$  over  $H_2$  given  $P$ , then  $O$  discriminates between  $H_1$  and  $H_2$  given  $P$ , but not conversely. Note that a key difference between discrimination and favoring is that the former but not the latter is symmetric.  $O$  discriminates between  $H_1$  and  $H_2$  given  $P$  only if  $O$  discriminates between  $H_2$  and  $H_1$  given  $P$ . However, favoring is *asymmetric* in that  $O$  favors  $H_1$  over  $H_2$  given  $P$  only if  $O$  *does not* favor  $H_2$  over  $H_1$  given  $P$ .<sup>14,15</sup>

The reader may wonder why we use (3.3) to characterize discrimination rather than say that  $O$  discriminates between  $H_1$  and  $H_2$  given  $P$  precisely when  $\Pr(H_1 | P \ \& \ O) \neq \Pr(H_2 | P \ \& \ O)$ . The reason is that the posterior probabilities in this last inequality typically depend on the prior probabilities of the hypotheses mentioned, whereas the likelihoods used in (3.3) typically do not. For example, a tuberculosis test kit has error probabilities that do not depend on how common or rare the disease is.

When is  $P$  an “appropriate” description of  $p$ ? This is a delicate question, but there are clear examples of *inappropriate* descriptions. In FISH-S, if the hypothesis set is  $\Gamma_{\text{hyp}} = \{H_{100}, H_{50}\}$ , and the observation set is  $\Gamma_{\text{obs}} = \{O_{100}, O_{50}\}$ , each observation discriminates between the two hypotheses, and this marks a difference between FISH-S and FISH-B. But suppose you use a net with small holes, learn  $O_{100}$ , and then describe the process of observation like this: “sampling fish from the lake with a net with small holes and catching 10 fish each of which is big.” Given this description of the process, the two hypotheses confer the same probability on the observation. The lesson here is that the process *outcome* (the observation yielded) shouldn’t be “packed into” the process *description*, on pain of entailing the nihilistic thesis that observations never discriminate between hypotheses.<sup>16,17</sup>

We require that  $P$  be a *true* description of  $p$  because we are interested in the discrimination-conduciveness of your *actual* observation process, and not the discrimination-conduciveness of what you *believe* your process to be. If, for example, your net has big holes but you believe

---

(2005) and Kotzen (2012) for discussion. We believe that each issue can be adequately resolved, but this isn’t the place to work out the details. See Sober (2009, 2018) for relevant discussion.

<sup>14</sup> The term “favoring” is sometimes used with a meaning that differs from the one used in (3.4). See Sober (2008, p. 36) for discussion.

<sup>15</sup> Discrimination is also distinct from “confirmation” in the sense of increase in probability. We return to this point in Section 5.3.

<sup>16</sup> There are other, and more subtle, caveats concerning when a process description is appropriate. See Sober (2009, 2018) for discussion. See also Eddington (1939, Ch. II, p. 20).

<sup>17</sup> A similar point has been made about the appropriate description of a system’s state at a given time in the context of defining determinism (see Berofsky 1971 and Earman 1986).

otherwise, the fact remains that the members of  $\Gamma_{\text{obs}}$  fail to discriminate between the members of  $\Gamma_{\text{hyp}}$  (where the hypotheses concern the size of the fish in the lake), given your actual observation process.<sup>18</sup> So P should be a true description of p.<sup>19</sup>

We mean for the likelihoods that describe p's degree of discrimination-conduciveness to be understood as *objective* probabilities in that they're independent of the observer's (and everyone else's) actual credences. For example, even if, in FISH-B, your credence in  $O_{100}$  given P and  $H_{100}$  is greater than your credence in  $O_{100}$  given P and  $H_{50}$ , this wouldn't change the fact that your observation process is not at all discrimination-conducive.<sup>20</sup>

How is degree of discrimination to be measured? We use " $\mathbf{d}(H_1, H_2, O | P)$ " to denote the degree to which, given P, O discriminates between hypotheses  $H_1$  and  $H_2$ . We assume that any adequate measure of degree of discrimination should meet each of the following conditions:

(3.5) For any hypotheses  $H_1$  and  $H_2$ , observation O, and observation process p, where P is a true appropriate description of p,  $\mathbf{d}(H_1, H_2, O | P) = \mathbf{d}(H_2, H_1, O | P)$ .

(3.6) For any hypotheses  $H_1$  and  $H_2$ , observation O, and observation process p, where P is a true appropriate description of p,  $\mathbf{d}(H_1, H_2, O | P)$  is minimal (i.e., takes the minimum possible value) precisely when  $\Pr(O | P \ \& \ H_1) = \Pr(O | P \ \& \ H_2)$ .

At this point, we don't assume any particular measure meeting these conditions.<sup>21</sup>

It's important to distinguish between cases of "fragile non-discrimination" and cases of "robust non-discrimination." The former are cases in which

---

<sup>18</sup> Compare: there can be cases where an observer's perceptual belief-forming processes are unreliable, and yet her information suggests otherwise.

<sup>19</sup> The issues discussed in this and the prior paragraph concern the "right-hand" sides (the conditioning sides) of the likelihoods in (3.3). There are also issues concerning the "left-hand" sides. For example, there's the issue of whether, as per the Requirement of Total Evidence, O should be the logically strongest true description of what you observed. See Barrett and Sober (forthcoming) and Epstein (2017) for a recent exchange on this issue.

<sup>20</sup> There are several interpretations of objective probability. These include propensity interpretations, frequency interpretations, and rational-credence interpretations. For helpful discussion, see Hájek (2012). We do not assume any particular interpretation. Indeed, we do not even assume that a reductive interpretation is needed; see Sober (2000, Ch. 3, sec. 3.2) on the "no-theory theory of probability."

<sup>21</sup> We will return to this issue in Section 5.2.

$$(3.7) \quad \Pr(O_\alpha | P \ \& \ H_j) = \Pr(O_\alpha | P \ \& \ H_k) \text{ for all } j \text{ and } k \text{ where } O_\alpha \text{ is the actual observation yielded by } p, \text{ but } \Pr(O_i | P \ \& \ H_j) \neq \Pr(O_i | P \ \& \ H_k) \text{ for some } j \text{ and } k \text{ where } i \neq \alpha.$$

The latter are cases where:

$$(3.8) \quad \Pr(O_i | P \ \& \ H_j) = \Pr(O_i | P \ \& \ H_k) \text{ for all } i, j, \text{ and } k.$$

Fragile non-discrimination is an unfortunate fact of experimental life. Sometimes perfectly good observation processes yield a non-discriminating observation, as when random sampling from an urn yields 50% green balls and the two hypotheses are “52% of the balls in the urn are green” and “48% of the balls in the urn are green.” Robust non-discrimination is a horse of a different color. Here the actual observation yielded is non-discriminating, and so too is every other member of the observation set at issue.

We now can state our first adequacy condition on measures of degree of discrimination-conduciveness:

**Minimality:** For any hypothesis set  $\Gamma_{\text{hyp}} = \{H_1, H_2, \dots, H_n\}$ , observation set  $\Gamma_{\text{obs}} = \{O_1, O_2, \dots, O_m\}$ , and observation process  $p$ , where  $P$  is a true appropriate description of  $p$ ,  $\mathbf{DDC}(\Gamma_{\text{hyp}}, \Gamma_{\text{obs}} | p)$  is minimal precisely when  $\mathbf{d}(H_i, H_j, O_k | P)$  is minimal for all  $i, j$ , and  $k$ .

In other words, the minimum possible degree of discrimination-conduciveness arises precisely when there’s robust non-discrimination. Notice that “ $\mathbf{d}(-)$ ” denotes the degree of discrimination that a single observation provides whereas “ $\mathbf{DDC}(-)$ ” denotes the degree of discrimination provided by a set of possible observations. In addition  $\mathbf{d}(-)$  concerns two hypotheses, whereas  $\mathbf{DDC}(-)$  concerns sets of hypotheses that may be larger.

We aren’t assuming that hypothesis sets or observation sets must be partitions (i.e., sets of mutually exclusive and jointly exhaustive propositions), though we are assuming that they are sets of mutually exclusive propositions.<sup>22</sup> However, if  $\Gamma_{\text{hyp}}$  is a partition, then it follows from **Minimality** that  $\mathbf{DDC}(\Gamma_{\text{hyp}}, \Gamma_{\text{obs}} | p)$  is minimal precisely when:

$$(3.9) \quad \Pr(O_i | P \ \& \ H_j) = \Pr(O_i | P) \text{ for all } i \text{ and } j.^{23}$$

---

<sup>22</sup> It’s straightforward to turn non-partitions into partitions. Take  $\Gamma_{\text{hyp}} = \{H_1, H_2, \dots, H_n\}$ , and suppose it’s not a partition. Now add to  $\Gamma_{\text{hyp}}$  the “catch-all” hypothesis  $\sim(H_1 \vee H_2 \vee \dots \vee H_n)$ . The result is a partition:  $\Gamma^*_{\text{hyp}} = \{H_1, H_2, \dots, H_n, \sim(H_1 \vee H_2 \vee \dots \vee H_n)\}$ . However, this procedure for constructing a partition often runs into the problem that likelihoods for catch-all hypotheses are unavailable (perhaps even undefined). For this reason, we want to allow for cases where the hypothesis sets of interest aren’t partitions.

<sup>23</sup> This is shown in Appendix A.

This means that  $p$  “takes over” and completely shuts out each member of  $\Gamma_{\text{hyp}}$ ;  $P$  “screens off”  $\Gamma_{\text{hyp}}$  from  $\Gamma_{\text{obs}}$  in that, given  $P$ , no member of  $\Gamma_{\text{hyp}}$  has any impact on the probability of any member of  $\Gamma_{\text{obs}}$ .<sup>24</sup> It doesn’t follow from **Minimality**, though, that  $\text{DDC}(\Gamma_{\text{hyp}}, \Gamma_{\text{obs}} | p)$  is minimal precisely when (3.9) holds. In fact, there are cases where  $\Gamma_{\text{hyp}}$  is *not* a partition,  $\mathbf{d}(H_i, H_j, O_k | P)$  is minimal for all  $i, j$ , and  $k$ , and (3.9) is false.<sup>25</sup>

We considered a case in Section 1 where the hypotheses of interest are observationally equivalent to each other. Any such case is a case of robust non-discrimination and so by **Minimality** is a case where  $\text{DDC}(\Gamma_{\text{hyp}}, \Gamma_{\text{obs}} | p)$  is minimal. This is the right result. The issue here is degree of *discrimination-conduciveness* as opposed to degree of OSE. When two hypotheses are observationally equivalent, no observation can discriminate between them, even though the situation would be the same if some alternative observation process were used.

#### 4 How *Not* to Measure Degree of Discrimination-Conduciveness

Here we construct three candidate measures of degree of discrimination-conduciveness and address the question of whether they meet **Minimality**. All are inspired by the extant literature on OSE concepts. We say “inspired by” rather than “found in” because the authors whose work we use here do not provide a measure; they treat OSEs as an on/off phenomenon. We argue that each of these candidate measures fails to meet **Minimality** and thus is inadequate. We then offer a brief diagnosis and suggest a way forward.

##### 4.1 $\text{DDC}_{\text{ROO}}^*$ , $\text{DDC}_{\text{BSP}}^*$ , and $\text{DDC}_{\text{CLV}}^*$

We noted in Section 2 that our usage of the expression “OSE” differs from usages that others have employed. Here we want to examine whether those alternative usages, or usages inspired by them, can be fashioned into an adequate measure of degree of discrimination-conduciveness.

First there’s the suggestion that OSEs occur precisely when  $p$  *rules out at last one* member of  $\Gamma_{\text{obs}}$ :

$$\text{ROO}(\Gamma_{\text{hyp}}, \Gamma_{\text{obs}} | p): \Pr(O_i | P) = 0 \text{ for some } i.$$

---

<sup>24</sup> There are different ways of talking about screening-off. Our claim could instead be put like this:  $P$  screens off *each member* of  $\Gamma_{\text{hyp}}$  from *each member* of  $\Gamma_{\text{obs}}$  in that, given  $P$ , no member of  $\Gamma_{\text{hyp}}$  has any impact on the probability of any member of  $\Gamma_{\text{obs}}$ . This is just a terminological matter.

<sup>25</sup> This is shown in Appendix B.

Second there's the suggestion that OSEs occur precisely when there is a *biased sampling process*:

**BSP**( $\Gamma_{\text{hyp}}, \Gamma_{\text{obs}} \mid p$ ):  $p$  is a biased (non-random) sampling process with respect to the population at issue in the hypotheses in  $\Gamma_{\text{hyp}}$ .

There's a third proposal in the literature to the effect that OSEs occur precisely when a "simple" likelihood value *changes* when  $p$  is taken into account:

**CLV**( $\Gamma_{\text{hyp}}, \Gamma_{\text{obs}} \mid p$ ):  $\Pr(O_i \mid P \ \& \ H_j) \neq \Pr(O_i \mid H_j)$  for some  $i$  and  $j$ .<sup>26</sup>

By "simple" we mean "not conditioned on  $P$ ." Can any of **ROO**, **BSP**, and **CLV** be used to develop an adequate measure of degree of discrimination-conduciveness?<sup>27</sup>

An initial difficulty is that none of **ROO**, **BSP**, and **CLV** is a matter of degree. Each either holds in a given case or does not; there is no middle ground. This difficulty can be remedied by constructing the following variants:

**ROO\***( $\Gamma_{\text{hyp}}, \Gamma_{\text{obs}} \mid p$ ): the percentage of  $O_i$  in  $\Gamma_{\text{obs}}$  such that  $\Pr(O_i \mid P) = 0$ .

**BSP\***( $\Gamma_{\text{hyp}}, \Gamma_{\text{obs}} \mid p$ ): the extent to which  $p$  is a biased sampling process with respect to the population at issue in the hypotheses in  $\Gamma_{\text{hyp}}$ .

**CLV\***( $\Gamma_{\text{hyp}}, \Gamma_{\text{obs}} \mid p$ ): the average degree of change from  $\Pr(O_i \mid H_j)$  to  $\Pr(O_i \mid P \ \& \ H_j)$  for all  $i$  and  $j$ .

It's not immediately clear how exactly to measure **BSP\*** or **CLV\***. But clearly they come in degrees, and so does **ROO\***.

This is a start, but a patch is needed. As stated, each of **ROO\***, **BSP\***, and **CLV\*** is inadequate when understood as a measure of degree of discrimination-conduciveness. Take **ROO\*** for example, and suppose that:

$$(4.1.1) \quad \Pr(O_i \mid P) = 0 \text{ for all } i.$$

It follows that **ROO\***( $\Gamma_{\text{hyp}}, \Gamma_{\text{obs}} \mid p$ ) is maximal at 1. Given (4.1.1), though, it follows that:

$$(4.1.2) \quad \Pr(O_i \mid P \ \& \ H_j) = 0 = \Pr(O_i \mid P \ \& \ H_k) \text{ for all } i, j, \text{ and } k.$$

Hence by **Minimality** it follows that **DDC**( $\Gamma_{\text{hyp}}, \Gamma_{\text{obs}} \mid p$ ) is minimal and thus not maximal.

There's an easy fix here. Consider:

---

<sup>26</sup> See, e.g., Sober (2003, 2009) and Roberts (2012).

<sup>27</sup> We don't see a way to turn Bostrom's OSE concept into a matter of degree, which is why we don't try to do so here.

$$\mathbf{DDC}_{\mathbf{ROO}^*}(\Gamma_{\text{hyp}}, \Gamma_{\text{obs}} \mid p): 1 - \mathbf{ROO}^*(\Gamma_{\text{hyp}}, \Gamma_{\text{obs}} \mid p).$$

$$\mathbf{DDC}_{\mathbf{BSP}^*}(\Gamma_{\text{hyp}}, \Gamma_{\text{obs}} \mid p): 1 - \mathbf{BSP}^*(\Gamma_{\text{hyp}}, \Gamma_{\text{obs}} \mid p).$$

$$\mathbf{DDC}_{\mathbf{CLV}^*}(\Gamma_{\text{hyp}}, \Gamma_{\text{obs}} \mid p): 1 - \mathbf{CLV}^*(\Gamma_{\text{hyp}}, \Gamma_{\text{obs}} \mid p).$$

Here  $\mathbf{DDC}_{\mathbf{ROO}^*}(\Gamma_{\text{hyp}}, \Gamma_{\text{obs}} \mid p)$  is a *decreasing* function of  $\mathbf{ROO}^*(\Gamma_{\text{hyp}}, \Gamma_{\text{obs}} \mid p)$ ,  $\mathbf{DDC}_{\mathbf{BSP}^*}(\Gamma_{\text{hyp}}, \Gamma_{\text{obs}} \mid p)$  is a *decreasing* function of  $\mathbf{BSP}^*(\Gamma_{\text{hyp}}, \Gamma_{\text{obs}} \mid p)$ , and  $\mathbf{DDC}_{\mathbf{CLV}^*}(\Gamma_{\text{hyp}}, \Gamma_{\text{obs}} \mid p)$  is a *decreasing* function of  $\mathbf{CLV}^*(\Gamma_{\text{hyp}}, \Gamma_{\text{obs}} \mid p)$ . So, for example, when (4.1.1) holds,  $\mathbf{DDC}_{\mathbf{ROO}^*}(\Gamma_{\text{hyp}}, \Gamma_{\text{obs}} \mid p)$  is minimal at 0, which is just as it should be.

Each of  $\mathbf{DDC}_{\mathbf{ROO}^*}$ ,  $\mathbf{DDC}_{\mathbf{BSP}^*}$ , and  $\mathbf{DDC}_{\mathbf{CLV}^*}$  has some *prima facie* plausibility as a measure of degree of discrimination-conduciveness. This is because lower degrees of discrimination-conduciveness often come with higher values of  $\mathbf{ROO}^*$ , higher values of  $\mathbf{BSP}^*$ , and higher values of  $\mathbf{CLV}^*$ . However, do  $\mathbf{DDC}_{\mathbf{ROO}^*}$ ,  $\mathbf{DDC}_{\mathbf{BSP}^*}$ , and  $\mathbf{DDC}_{\mathbf{CLV}^*}$  satisfy the requirement of **Minimality**?

#### 4.2 $\mathbf{DDC}_{\mathbf{ROO}^*}$ and Minimality

Here we draw on the tuberculosis example from Section 2. Let  $\Gamma_{\text{hyp}} = \{H_t, H_{\sim t}\}$  and  $\Gamma_{\text{obs}} = \{O_t, O_{\sim t}\}$ , where:

- $H_t$ : Joe has tuberculosis.
- $H_{\sim t}$ : Joes does not have tuberculosis.
- $O_t$ :  $k^*$  says that Joe has tuberculosis.
- $O_{\sim t}$ :  $k^*$  says that Joe does not have tuberculosis.

Recall that  $k^*$  isn't much of a test kit. When used on a given patient, a randomizing device inside of  $k^*$  flips a fair coin. If the coin comes up heads,  $k^*$  says that the patient has tuberculosis. If the coin doesn't come up heads,  $k^*$  says that the patient doesn't have tuberculosis. Suppose you use  $k^*$  on Joe, and it tells you that he has tuberculosis. It follows that neither  $O_t$  nor  $O_{\sim t}$  discriminates between  $H_t$  and  $H_{\sim t}$ :

$$(4.2.1) \quad \Pr(O_t \mid P \ \& \ H_t) = \frac{1}{2} = \Pr(O_t \mid P \ \& \ H_{\sim t})$$

$$(4.2.2) \quad \Pr(O_{\sim t} \mid P \ \& \ H_t) = \frac{1}{2} = \Pr(O_{\sim t} \mid P \ \& \ H_{\sim t})$$

But it also follows that neither  $O_t$  nor  $O_{\sim t}$  is ruled out by  $p$ :

$$(4.2.3) \quad \Pr(O_t \mid P) = \frac{1}{2} = \Pr(O_{\sim t} \mid P)$$

Given (4.2.3),  $\mathbf{DDC}_{\mathbf{ROO}^*}(\Gamma_{\text{hyp}}, \Gamma_{\text{obs}} \mid p)$  is *maximal*. But given (4.2.1) and (4.2.2), it follows by **Minimality** that  $\mathbf{DDC}(\Gamma_{\text{hyp}}, \Gamma_{\text{obs}} \mid p)$  is *minimal*.  $\mathbf{DDC}_{\mathbf{ROO}^*}$  therefore fails to meet **Minimality**.

### 4.3 DDC<sub>BSP\*</sub> and Minimality

Now let's return to RED LIGHTS from Section 2. You're about to enter a room. Let  $\Gamma_{\text{hyp}} = \{H_r, H_w\}$  and  $\Gamma_{\text{obs}} = \{O_r, O_w\}$ , where:

- $H_r$ : The walls in the room are red.
- $H_w$ : The walls in the room are white.
- $O_r$ : The walls in the room look red to you.
- $O_w$ : The walls in the room look white to you.

You carefully scan each wall in its entirety and come to learn that  $O_r$ . But, unbeknownst to you, the walls in the room are lit by red lights in such a way that the walls would look red whether they were red or white. It follows (given natural assumptions) that this is a case of robust non-discrimination:

$$(4.3.1) \quad \Pr(O_r | P \ \& \ H_r) = 1 = \Pr(O_r | P \ \& \ H_w)$$

$$(4.3.2) \quad \Pr(O_w | P \ \& \ H_r) = 0 = \Pr(O_w | P \ \& \ H_w)$$

Hence by **Minimality** it follows that  $\mathbf{DDC}(\Gamma_{\text{hyp}}, \Gamma_{\text{obs}} | p)$  is minimal. However, your observation process isn't a sampling process and thus it isn't a *biased* sampling process. There's no literal population from which you've extracted member objects. Given this,  $\mathbf{DDC}_{\text{BSP}^*}(\Gamma_{\text{hyp}}, \Gamma_{\text{obs}} | p)$  is either maximal or undefined. Thus  $\mathbf{DDC}_{\text{BSP}^*}$ , like  $\mathbf{DDC}_{\text{ROO}^*}$ , fails to meet **Minimality**.

### 4.4 DDC<sub>CLV\*</sub> and Minimality

Imagine a variant of TUBERCULOSIS where it's a matter of chance which test kit you will use on Joe. You are going to flip a fair coin. If it comes up heads, you will use  $k$  (the good test kit) on Joe, where  $k$  is such that:

$$(4.4.1) \quad \Pr(O_t | P_k \ \& \ H_t) = 0.99$$

$$(4.4.2) \quad \Pr(O_{\sim t} | P_k \ \& \ H_{\sim t}) = 0.95$$

If, instead, the coin doesn't come up heads, you will use the totally useless kit  $k^*$  (the bad test kit, as described in Section 2) on Joe. It follows that:

$$\begin{aligned} (4.4.3) \quad \Pr(O_t | H_t) &= \Pr(P_k | H_t) \Pr(O_t | P_k \ \& \ H_t) + \Pr(P_{k^*} | H_t) \Pr(O_t | P_{k^*} \ \& \ H_t) \\ &= (0.5)(0.99) + (0.5)(0.5) \\ &= 0.745 \end{aligned}$$



$$\begin{aligned}
(4.4.4) \quad \Pr(O_t | H_{\sim t}) &= \Pr(P_k | H_{\sim t})\Pr(O_t | P_k \& H_{\sim t}) + \Pr(P_{k^*} | H_{\sim t})\Pr(O_t | P_{k^*} \& H_{\sim t}) \\
&= (0.5)(0.05) + (0.5)(0.5) \\
&= 0.275
\end{aligned}$$

Suppose you flip the coin and it comes up heads. You then use  $k^*$  on Joe, and it says that he has tuberculosis. Call this case “TUBERCULOSIS-FLIP.”  $\mathbf{DDC}_{CLV^*}(\Gamma_{hyp}, \Gamma_{obs} | p_{k^*})$  is given by the average of the following:

$$(4.4.5) \quad \text{the degree of change from } \Pr(O_t | H_t) = 0.745 \text{ to } \Pr(O_t | P_{k^*} \& H_t) = 0.5$$

$$(4.4.6) \quad \text{the degree of change from } \Pr(O_{\sim t} | H_t) = 0.255 \text{ to } \Pr(O_{\sim t} | P_{k^*} \& H_t) = 0.5$$

$$(4.4.7) \quad \text{the degree of change from } \Pr(O_t | H_{\sim t}) = 0.275 \text{ to } \Pr(O_t | P_{k^*} \& H_{\sim t}) = 0.5$$

$$(4.4.8) \quad \text{the degree of change from } \Pr(O_{\sim t} | H_{\sim t}) = 0.725 \text{ to } \Pr(O_{\sim t} | P_{k^*} \& H_{\sim t}) = 0.5$$

We noted in Section 4.1 that it’s not immediately clear how to measure  $CLV^*$ . It’s clear, though, that this is not a case where  $CLV^*$  is maximal.<sup>28</sup> Hence  $\mathbf{DDC}_{CLV^*}(\Gamma_{hyp}, \Gamma_{obs} | p_{k^*})$  isn’t minimal. Now consider just the second probabilities in (4.4.5)-(4.4.8). Since the second probabilities in (4.4.5) and (4.4.7) are equal to each other, and the same is true of the second probabilities in (4.4.6) and (4.4.8), it follows by **Minimality** that  $\mathbf{DDC}(\Gamma_{hyp}, \Gamma_{obs} | p_{k^*})$  is minimal. This means that  $\mathbf{DDC}_{CLV^*}$ , like the first two candidate measures, fails to meet **Minimality**.<sup>29</sup>

## 4.5 Diagnosis

$\mathbf{DDC}_{ROO^*}$ ,  $\mathbf{DDC}_{BSP^*}$ , and  $\mathbf{DDC}_{CLV^*}$  sometimes violate **Minimality** because each measures something that isn’t fully determined by the various degrees to which, given  $P$ , the members of  $\Gamma_{obs}$  discriminate between the members of  $\Gamma_{hyp}$ . For example,  $\mathbf{DDC}_{ROO^*}$  measures the percentage of ruled-out observations, but the percentage of ruled-out observations isn’t fully determined by

---

<sup>28</sup> If TUBERCULOSIS-FLIP were changed so that, say,  $\Pr(O_t | P_k \& H_t) = 0.999$  and  $\Pr(O_{\sim t} | P_k \& H_{\sim t}) = 0.995$ , then each of the four degrees of change at issue would be greater, and so the average degree of change would be greater. It follows that TUBERCULOSIS-FLIP isn’t a case where  $CLV^*$  is maximal.

<sup>29</sup> It might be insisted that  $\mathbf{DDC}_{CLV^*}$  should be understood so that  $\Pr(O_i | H_j)$  is elliptical for

$$\Pr(O_i | P^* \& H_j),$$

where  $P^*$  is what the observer takes to be a true appropriate description of  $p$ . But this wouldn’t help. To see why, return to RED LIGHTS, and suppose you’re aware of the red lights so that  $P^* = P$ . It follows by  $\mathbf{DDC}_{CLV^*}$  on this new way of understanding it that the degree of discrimination-conduciveness is *maximal*. This isn’t the right result.

the various degrees to which, given  $P$ , the members of  $\Gamma_{\text{obs}}$  discriminate between the members of  $\Gamma_{\text{hyp}}$ . This follows from the fact that some cases of robust non-discrimination, for example, FISH-B and TUBERCULOSIS-FLIP, differ from each other in terms of the percentage of ruled out observations. What's needed, we suggest, is a measure of degree of discrimination-conduciveness that directly measures something that's fully determined by the various degrees to which, given  $P$ , the members of  $\Gamma_{\text{obs}}$  discriminate between the members of  $\Gamma_{\text{hyp}}$ .

## 5 How to Measure Degree of Discrimination-Conduciveness

This section is divided into four subsections. In Section 5.1, we develop a schema for measuring degree of discrimination-conduciveness, and argue that any instance of it meets **Minimality**. In Section 5.2, we describe two *prima facie* plausible instances of our schema. In Section 5.3, we consider some alternative measures of discrimination-conduciveness, and argue that they are inferior to the two measures from Section 5.2. In Section 5.4, we give a brief summary of our main findings.

### 5.1 Average Degree of Discrimination

Consider:

$$\mathbf{DDC}_{\text{ADD}}(\Gamma_{\text{hyp}}, \Gamma_{\text{obs}} | p): \frac{\sum_{i < j} \mathbf{d}(H_i, H_j, O_k | P)}{\binom{m}{i} \sum_{i=1}^{i=n-1} (i)}$$

Here  $\Gamma_{\text{hyp}} = \{H_1, H_2, \dots, H_n\}$  is some hypothesis set,  $\Gamma_{\text{obs}} = \{O_1, O_2, \dots, O_m\}$  is some observation set,  $p$  is some observation process, where  $P$  is a true appropriate description of  $p$ , and  $\mathbf{d}(H_i, H_j, O_k | P)$  is some measure of degree of discrimination.  $\mathbf{DDC}_{\text{ADD}}(\Gamma_{\text{hyp}}, \Gamma_{\text{obs}} | p)$  is the average degree to which, given  $P$ , the members of  $\Gamma_{\text{obs}}$  discriminate between the members of  $\Gamma_{\text{hyp}}$ . The subscript in “ $\mathbf{DDC}_{\text{ADD}}$ ” is short for *Average Degree of Discrimination*.

An illustration is in order. Let  $\Gamma_{\text{hyp}} = \{H_1, H_2, H_3\}$  and  $\Gamma_{\text{obs}} = \{O_1, O_2, O_3\}$ . Then:

$$\mathbf{DDC}_{\text{ADD}}(\Gamma_{\text{hyp}}, \Gamma_{\text{obs}} | p) = \frac{\mathbf{d}(H_1, H_2, O_1 | P) + \mathbf{d}(H_1, H_3, O_1 | P) + \mathbf{d}(H_2, H_3, O_1 | P) + \mathbf{d}(H_1, H_2, O_2 | P) + \mathbf{d}(H_1, H_3, O_2 | P) + \mathbf{d}(H_2, H_3, O_2 | P) + \mathbf{d}(H_1, H_2, O_3 | P) + \mathbf{d}(H_1, H_3, O_3 | P) + \mathbf{d}(H_2, H_3, O_3 | P)}{9}$$

Notice that the sum on the right includes  $\mathbf{d}(H_1, H_2, O_1 | P)$  but not  $\mathbf{d}(H_2, H_1, O_1 | P)$ . This is because degree of discrimination is symmetric (as noted in Section 3), and so it would be double counting to include both.

$\mathbf{DDC}_{\text{ADD}}$  isn't itself a measure of degree of discrimination-conduciveness. Rather, it's just a *schema* for measuring degree of discrimination-conduciveness. There are different ways of measuring degree of discrimination  $\mathbf{d}(-)$ , and different such measures lead to different instances of  $\mathbf{DDC}_{\text{ADD}}$ .

We assume here and throughout that any instance of  $\mathbf{DDC}_{\text{ADD}}$  is such that the underlying measure of degree of discrimination meets both (3.5) and (3.6) from Section 3. The former is the symmetry condition that  $\mathbf{d}(H_1, H_2, O | P) = \mathbf{d}(H_2, H_1, O | P)$ ; the latter is the condition that  $\mathbf{d}(H_1, H_2, O | P)$  is minimal precisely when  $\Pr(O | P \ \& \ H_1) = \Pr(O | P \ \& \ H_2)$ .

It's easy to see that every instance of  $\mathbf{DDC}_{\text{ADD}}$  (so understood) meets **Minimality**. Let  $\mathbf{DDC}_{\text{ADD-X}}$  be some arbitrary instance of  $\mathbf{DDC}_{\text{ADD}}$ , where degree of discrimination is measured by  $\mathbf{d}_X$ . First, suppose that  $\mathbf{d}_X(H_i, H_j, O_k | P)$  is minimal for all  $i, j$ , and  $k$ . Then since  $\mathbf{DDC}_{\text{ADD-X}}(\Gamma_{\text{hyp}}, \Gamma_{\text{obs}} | p)$  is an average of degrees of discrimination where each degree is minimal, it follows that  $\mathbf{DDC}_{\text{ADD-X}}(\Gamma_{\text{hyp}}, \Gamma_{\text{obs}} | p)$  itself is minimal. Second, suppose that  $\mathbf{DDC}_{\text{ADD-X}}(\Gamma_{\text{hyp}}, \Gamma_{\text{obs}} | p)$  is minimal. Since  $\mathbf{DDC}_{\text{ADD-X}}(\Gamma_{\text{hyp}}, \Gamma_{\text{obs}} | p)$  is an average of degrees of discrimination, it is minimal only if each degree of discrimination is minimal. Thus  $\mathbf{d}_X(H_i, H_j, O_k | P)$  is minimal for all  $i, j$ , and  $k$ . In conformity with **Minimality**,  $\mathbf{DDC}_{\text{ADD-X}}(\Gamma_{\text{hyp}}, \Gamma_{\text{obs}} | p)$  is minimal precisely when  $\mathbf{d}_X(H_i, H_j, O_k | P)$  is minimal for all  $i, j$ , and  $k$ .

Now let's return to our diagnosis of why  $\mathbf{DDC}_{\text{ROO}^*}$ ,  $\mathbf{DDC}_{\text{BSP}^*}$ , and  $\mathbf{DDC}_{\text{CLV}^*}$  sometimes violate **Minimality**, and our suggestion that what's needed is a measure of degree of discrimination-conduciveness that directly measures something that's fully determined by the various degrees to which, given  $P$ , the members of  $\Gamma_{\text{obs}}$  discriminate between the members of  $\Gamma_{\text{hyp}}$ . Every instance of  $\mathbf{DDC}_{\text{ADD}}$  is a measure of exactly that sort. This is why no instance of  $\mathbf{DDC}_{\text{ADD}}$  has trouble with **Minimality**.

It's hardly surprising, on reflection, that every instance of  $\mathbf{DDC}_{\text{ADD}}$  meets **Minimality**. We don't mean to suggest otherwise. The important point is that, whether surprising or not, every instance of  $\mathbf{DDC}_{\text{ADD}}$  satisfies this condition of adequacy.

It won't suffice, though, to rest content with  $\mathbf{DDC}_{\text{ADD}}$ . There's still the issue of how to measure degree of discrimination. We turn to that issue now.

## 5.2 $\mathbf{DDC}_{\text{ADD-AD}}$ and $\mathbf{DDC}_{\text{ADD-SD}}$

There are numerous measures of degree of discrimination in logical space. We want to focus on these:

$$\mathbf{d}_{\text{AD}}(H_1, H_2, O | P) = \left| \Pr(O | P \ \& \ H_1) - \Pr(O | P \ \& \ H_2) \right|$$

$$\mathbf{d}_{\text{SD}}(H_1, H_2, O | P) = \left[ \Pr(O | P \ \& \ H_1) - \Pr(O | P \ \& \ H_2) \right]^2$$

The subscript in “ $\mathbf{d}_{AD}$ ” is short for “Absolute Difference.” The subscript in “ $\mathbf{d}_{SD}$ ” is short for “Squared Difference.”

Do  $\mathbf{d}_{AD}$  and  $\mathbf{d}_{SD}$  meet (3.5) and (3.6)? They meet (3.5) since absolute differences and squared differences are order-invariant. And they meet (3.6) because absolute differences and squared differences are minimal at zero precisely when the two values in question are equal to each other.

Each of  $\mathbf{d}_{AD}$  and  $\mathbf{d}_{SD}$  can be used to flesh out  $\mathbf{DDC}_{ADD}$ . If  $\mathbf{d}_{AD}$  is substituted for  $\mathbf{d}$  in  $\mathbf{DDC}_{ADD}$ , the result is:

$$\mathbf{DDC}_{ADD-AD}(\Gamma_{hyp}, \Gamma_{obs} | p) : \frac{\sum_{i < j} \mathbf{d}_{AD}(H_i, H_j, O_k | P)}{\binom{m}{i} \sum_{i=1}^{i=n-1} (i)} = \frac{\sum_{i < j} |\Pr(O_k | P \& H_i) - \Pr(O_k | P \& H_j)|}{\binom{m}{i} \sum_{i=1}^{i=n-1} (i)}$$

If, instead,  $\mathbf{d}_{SD}$  is substituted for  $\mathbf{d}$  in  $\mathbf{DDC}_{ADD}$ , the result is:

$$\mathbf{DDC}_{ADD-SD}(\Gamma_{hyp}, \Gamma_{obs} | p) : \frac{\sum_{i < j} \mathbf{d}_{SD}(H_i, H_j, O_k | P)}{\binom{m}{i} \sum_{i=1}^{i=n-1} (i)} = \frac{\sum_{i < j} [\Pr(O_k | P \& H_i) - \Pr(O_k | P \& H_j)]^2}{\binom{m}{i} \sum_{i=1}^{i=n-1} (i)}$$

Each of these measures has a range of  $[0, 1]$ .<sup>30</sup> Since every instance of  $\mathbf{DDC}_{ADD}$  meets **Minimality**, it follows immediately that  $\mathbf{DDC}_{ADD-AD}$  and  $\mathbf{DDC}_{ADD-SD}$  both meet **Minimality**.

---

<sup>30</sup> If  $\Pr(O_k | P \& H_i) = \Pr(O_k | P \& H_j)$  for all  $i, j$ , and  $k$ , then  $\mathbf{DDC}_{ADD-AD}(\Gamma_{hyp}, \Gamma_{obs} | p)$  and  $\mathbf{DDC}_{ADD-SD}(\Gamma_{hyp}, \Gamma_{obs} | p)$  are both equal to 0. If  $\Pr(O_k | P \& H_i) \neq \Pr(O_k | P \& H_j)$  for some  $i, j$ , and  $k$ , then  $\mathbf{DDC}_{ADD-AD}(\Gamma_{hyp}, \Gamma_{obs} | p)$  and  $\mathbf{DDC}_{ADD-SD}(\Gamma_{hyp}, \Gamma_{obs} | p)$  are both greater than 0. If each of  $\Gamma_{hyp}$  and  $\Gamma_{obs}$  has exactly two members and each member of  $\Gamma_{hyp}$  confers a probability of 1 on a different member of  $\Gamma_{obs}$ , then both  $\mathbf{DDC}_{ADD-AD}(\Gamma_{hyp}, \Gamma_{obs} | p)$  and  $\mathbf{DDC}_{ADD-SD}(\Gamma_{hyp}, \Gamma_{obs} | p)$  equal 1. If either of  $\Gamma_{hyp}$  and  $\Gamma_{obs}$  has more than two members, then each of  $\mathbf{DDC}_{ADD-AD}(\Gamma_{hyp}, \Gamma_{obs} | p)$  and  $\mathbf{DDC}_{ADD-SD}(\Gamma_{hyp}, \Gamma_{obs} | p)$  is less than 1. If, for example,  $\Gamma_{hyp} = \{H_1, H_2\}$ ,  $\Gamma_{obs} = \{O_1, O_2, O_3\}$ , and  $H_1$  and  $H_2$  maximally disagree on  $O_1$  and  $O_2$  in that  $\Pr(O_1 | H_1) = 1 > 0 = \Pr(O_1 | H_2)$  and  $\Pr(O_2 | H_1) = 0 < 1 = \Pr(O_2 | H_2)$ , then  $H_1$  and  $H_2$  maximally agree on  $O_3$  in that  $\Pr(O_3 | P \& H_1) = 0 = \Pr(O_3 | P \& H_2)$  and thus both  $\mathbf{DDC}_{ADD-AD}(\Gamma_{hyp}, \Gamma_{obs} | p)$  and  $\mathbf{DDC}_{ADD-SD}(\Gamma_{hyp}, \Gamma_{obs} | p)$  are less than 1.

Now consider:

**Dominance:** For any hypothesis set  $\Gamma_{\text{hyp}} = \{H, \sim H\}$ , observation set  $\Gamma_{\text{obs}} = \{O, \sim O\}$ , and observation processes  $p$  and  $p^*$ , where  $P$  is a true appropriate description of  $p$ , and  $P^*$  is a true appropriate description of  $p^*$ , if  $\Pr(O | P \& H) > \Pr(O | P^* \& H) > \Pr(O | P^* \& \sim H) > \Pr(O | P \& \sim H)$ , then  $\mathbf{DDC}(\Gamma_{\text{hyp}}, \Gamma_{\text{obs}} | p) > \mathbf{DDC}(\Gamma_{\text{hyp}}, \Gamma_{\text{obs}} | p^*)$ .

If  $\Pr(O | P \& H) > \Pr(O | P^* \& H) > \Pr(O | P^* \& \sim H) > \Pr(O | P \& \sim H)$ , then  $p$  dominates  $p^*$  with respect to  $\Gamma_{\text{hyp}}$  and  $\Gamma_{\text{obs}}$  in that regardless of which member of  $\Gamma_{\text{obs}}$  is true, the degree to which it discriminates between  $H$  and  $\sim H$  given  $P$  is greater than the degree to which it discriminates between  $H$  and  $\sim H$  given  $P^*$ . **Dominance** says that if  $p$  dominates  $p^*$  with respect to  $\Gamma_{\text{hyp}}$  and  $\Gamma_{\text{obs}}$ , then  $p$ 's degree of discrimination-conduciveness with respect to  $\Gamma_{\text{hyp}}$  and  $\Gamma_{\text{obs}}$  is greater than  $p^*$ 's degree of discrimination-conduciveness with respect to  $\Gamma_{\text{hyp}}$  and  $\Gamma_{\text{obs}}$ .

It turns out that  $\mathbf{DDC}_{\text{ADD-AD}}$  and  $\mathbf{DDC}_{\text{ADD-SD}}$  both meet **Dominance** (in addition to meeting **Minimality**). Take some case where:

$$(5.2.1) \quad \Pr(O | P \& H) > \Pr(O | P^* \& H) > \Pr(O | P^* \& \sim H) > \Pr(O | P \& \sim H).$$

It follows that:

$$(5.2.2) \quad \Pr(\sim O | P \& \sim H) > \Pr(\sim O | P^* \& \sim H) > \Pr(\sim O | P^* \& H) > \Pr(\sim O | P \& H)$$

Given (5.2.1) and (5.2.2), it follows that:

$$(5.2.3) \quad \mathbf{d}_{\text{AD}}(H, \sim H, O | P) > \mathbf{d}_{\text{AD}}(H, \sim H, O | P^*)$$

$$(5.2.4) \quad \mathbf{d}_{\text{AD}}(H, \sim H, \sim O | P) > \mathbf{d}_{\text{AD}}(H, \sim H, \sim O | P^*)$$

$$(5.2.5) \quad \mathbf{d}_{\text{SD}}(H, \sim H, O | P) > \mathbf{d}_{\text{SD}}(H, \sim H, O | P^*)$$

$$(5.2.6) \quad \mathbf{d}_{\text{SD}}(H, \sim H, \sim O | P) > \mathbf{d}_{\text{SD}}(H, \sim H, \sim O | P^*)$$

Hence, as per **Dominance**,  $\mathbf{DDC}_{\text{ADD-AD}}(\Gamma_{\text{hyp}}, \Gamma_{\text{obs}} | p)$  is greater than  $\mathbf{DDC}_{\text{ADD-AD}}(\Gamma_{\text{hyp}}, \Gamma_{\text{obs}} | p^*)$ , and  $\mathbf{DDC}_{\text{ADD-SD}}(\Gamma_{\text{hyp}}, \Gamma_{\text{obs}} | p)$  is greater than  $\mathbf{DDC}_{\text{ADD-SD}}(\Gamma_{\text{hyp}}, \Gamma_{\text{obs}} | p^*)$ .

There's more.  $\mathbf{DDC}_{\text{ADD-AD}}$  and  $\mathbf{DDC}_{\text{ADD-SD}}$  are ordinally equivalent if the hypothesis sets of interest have exactly two members and the observation sets at issue are two-membered partitions:

$$(5.2.7) \quad \text{For any hypothesis sets } \Gamma_{\text{hyp}} = \{H_1, H_2\} \text{ and } \Gamma^*_{\text{hyp}} = \{H^*_1, H^*_2\}, \text{ observation sets } \Gamma_{\text{obs}} = \{O, \sim O\} \text{ and } \Gamma^*_{\text{obs}} = \{O^*, \sim O^*\}, \text{ and observation processes } p \text{ and } p^*, \text{ where } P \text{ is a true appropriate description of } p, \text{ and } P^* \text{ is a true appropriate description of } p^*, \mathbf{DDC}_{\text{ADD-AD}}(\Gamma_{\text{hyp}}, \Gamma_{\text{obs}} | p) > / = / < \mathbf{DDC}_{\text{ADD-AD}}(\Gamma^*_{\text{hyp}}, \Gamma^*_{\text{obs}} | p^*)$$

$p^*$ ) if and only if  $\mathbf{DDC}_{\text{ADD-SD}}(\Gamma_{\text{hyp}}, \Gamma_{\text{obs}} | p) > / = / < \mathbf{DDC}_{\text{ADD-SD}}(\Gamma_{\text{hyp}}^*, \Gamma_{\text{obs}}^* | p^*)$ .<sup>31</sup>

This is significant, since in many contexts the crucial questions concern *relative* degrees of discrimination-conduciveness as opposed to *absolute* degrees of discrimination-conduciveness.

However, ordinal equivalence fails in the general case where the hypothesis sets and the observation sets can have more than two members and don't need to be partitions. There are cases, for instance, where  $\Gamma_{\text{hyp}} = \{H_1, H_2\}$ ,  $\Gamma_{\text{obs}} = \{O_1, O_2, O_3\}$ ,  $\Gamma_{\text{obs}}^* = \{O^*_1, O^*_2, O^*_3\}$ , and  $\mathbf{DDC}_{\text{ADD-AD}}(\Gamma_{\text{hyp}}, \Gamma_{\text{obs}} | p)$  is less than  $\mathbf{DDC}_{\text{ADD-AD}}(\Gamma_{\text{hyp}}, \Gamma_{\text{obs}}^* | p)$ , whereas  $\mathbf{DDC}_{\text{ADD-SD}}(\Gamma_{\text{hyp}}, \Gamma_{\text{obs}} | p)$  is greater than  $\mathbf{DDC}_{\text{ADD-SD}}(\Gamma_{\text{hyp}}, \Gamma_{\text{obs}}^* | p)$ .<sup>32</sup>

We see no clear reason at this point for preferring one of our two target measures over the other. But that's okay for our purposes. It isn't essential here that we defend some particular measure of degree of discrimination-conduciveness as superior to all rival measures. We return to this issue in Section 9.

### 5.3 Other Measures of Degree of Discrimination-Conduciveness

There are numerous measures of degree of discrimination in addition to  $\mathbf{d}_{\text{AD}}$  and  $\mathbf{d}_{\text{SD}}$ . Here is one:

$$\mathbf{d}_{\mathbf{R}}(H_1, H_2, O | P) = \frac{\Pr(O | P \ \& \ H_1)}{\Pr(O | P \ \& \ H_2)}$$

The subscript in “ $\mathbf{d}_{\mathbf{R}}$ ” is short for “Ratio.” This measure is clearly inadequate. If, say,  $\Pr(O | P \ \& \ H_1) > \Pr(O | P \ \& \ H_2)$ , then  $\mathbf{d}_{\mathbf{R}}(H_1, H_2, O | P) > \mathbf{d}_{\mathbf{R}}(H_2, H_1, O | P)$ . Hence it fails to meet (3.5). If, instead,  $\Pr(O | P \ \& \ H_1) = \Pr(O | P \ \& \ H_2) > 0$ , then  $\mathbf{d}_{\mathbf{R}}(H_1, H_2, O | P) = 1$ , which is greater than its minimal value of 0. It thus fails to meet (3.6).<sup>33</sup>

---

<sup>31</sup> This is shown in Appendix C.

<sup>32</sup> This is shown in Appendix D.

<sup>33</sup> It doesn't follow that no ratio-based measure of degree of discrimination meets both (3.5) and (3.6). Consider:

$$\mathbf{d}_{\mathbf{R}^*}(H_1, H_2, O | P) = \left| \log \frac{\Pr(O | P \ \& \ H_1)}{\Pr(O | P \ \& \ H_2)} \right|$$

This result has implications for how degree of discrimination-conduciveness should be understood. Consider:

$$\mathbf{DDC}_{\text{ADD-R}}(\Gamma_{\text{hyp}}, \Gamma_{\text{obs}} | p) : \frac{\sum_{i < j} \mathbf{d}_{\mathbf{R}}(H_i, H_j, O_k | P)}{\binom{m}{i} \sum_{i=1}^{i=n-1} (i)} = \frac{\sum_{i < j} \frac{\Pr(O_k | P \& H_i)}{\Pr(O_k | P \& H_j)}}{\binom{m}{i} \sum_{i=1}^{i=n-1} (i)}$$

This resembles  $\mathbf{DDC}_{\text{ADD-AD}}$  and  $\mathbf{DDC}_{\text{ADD-SD}}$  except that the underlying measure of degree of discrimination is  $\mathbf{d}_{\mathbf{R}}$  as opposed to  $\mathbf{d}_{\text{AD}}$  or  $\mathbf{d}_{\text{SD}}$ . But  $\mathbf{DDC}_{\text{ADD-R}}$ , unlike  $\mathbf{DDC}_{\text{ADD-AD}}$  and  $\mathbf{DDC}_{\text{ADD-SD}}$ , fails to meet **Minimality**. Any case of robust non-discrimination where none of the probabilities equals 0 is a case where  $\mathbf{DDC}_{\text{ADD-R}}(\Gamma_{\text{hyp}}, \Gamma_{\text{obs}} | p)$  equals 1. However,  $\mathbf{DDC}_{\text{ADD-R}}(\Gamma_{\text{hyp}}, \Gamma_{\text{obs}} | p)$  can be less than 1.<sup>34</sup>

Now consider the following variant of  $\mathbf{d}_{\text{AD}}$ :

$$\mathbf{d}_{\text{AD}^*}(H_1, H_2, O | P) = \frac{|\Pr(O | P \& H_1) - \Pr(O | P \& H_2)|}{\Pr(O | P)}$$

This measure, like  $\mathbf{d}_{\text{AD}}$  and unlike  $\mathbf{d}_{\mathbf{R}}$ , meets both (3.5) and (3.6).<sup>35</sup> It is thus at least minimally adequate.

If  $\mathbf{d}_{\text{AD}^*}$  is plugged in for  $\mathbf{d}$  in  $\mathbf{DDC}_{\text{ADD}}$ , then the result is:

---

This is a ratio-based measure. But, unlike  $\mathbf{d}_{\mathbf{R}}$ , it meets both (3.5) and (3.6). The key here is that

$$\log \frac{\Pr(O | P \& H_1)}{\Pr(O | P \& H_2)} \text{ equals } \log \Pr(O | P \& H_1) \text{ minus } \log \Pr(O | P \& H_2).$$

<sup>34</sup> Let  $\Gamma_{\text{hyp}} = \{H, \sim H\}$  and  $\Gamma_{\text{obs}} = \{O, \sim O\}$ , and suppose that  $\Pr(O | P \& H) = 0$  and  $\Pr(O | P \& \sim H) = 0.01$ . Then  $\mathbf{DDC}_{\text{ADD-R}}(\Gamma_{\text{hyp}}, \Gamma_{\text{obs}} | p)$  equals  $\frac{0}{0.01} + \frac{1}{0.99}$  which is approximately 0.505.

<sup>35</sup> We're assuming here, for the sake of argument, that  $\Pr(O | P) > 0$ . If this assumption is dropped, then  $\mathbf{d}_{\text{AD}^*}$  fails to meet (3.6). For, there are cases where  $\Pr(O | P \& H) = 0 = \Pr(O | P \& \sim H)$  and thus  $\Pr(O | P) = 0$ . (3.6) implies that  $\mathbf{d}_{\text{AD}^*}(H, \sim H, O | P)$  is minimal in such cases. But  $\mathbf{d}_{\text{AD}^*}(H, \sim H, O | P)$  is undefined and thus fails to be minimal.

$$\mathbf{DDC}_{\text{ADD-AD}^*}(\Gamma_{\text{hyp}}, \Gamma_{\text{obs}} | \mathbf{p}) : \frac{\sum_{i < j} \mathbf{d}_{\text{AD}^*}(H_i, H_j, O_k | \mathbf{P})}{(m) \sum_{i=1}^{i=n-1} (i)} = \frac{\sum_{i < j} \frac{|\Pr(O_k | \mathbf{P} \& H_i) - \Pr(O_k | \mathbf{P} \& H_j)|}{\Pr(O_k | \mathbf{P})}}{(m) \sum_{i=1}^{i=n-1} (i)}$$

This is an instance of  $\mathbf{DDC}_{\text{ADD}}$ . So, since every instance of  $\mathbf{DDC}_{\text{ADD}}$  meets **Minimality**, it follows that it too meets **Minimality**.<sup>36</sup> Why, then, do we prefer  $\mathbf{DDC}_{\text{ADD-AD}}$  and  $\mathbf{DDC}_{\text{ADD-SD}}$  over  $\mathbf{DDC}_{\text{ADD-AD}^*}$ ?

Our preference here is based in part on the following:

**Likelihoods:** For any hypothesis set  $\Gamma_{\text{hyp}} = \{H_1, H_2, \dots, H_n\}$ , observation set  $\Gamma_{\text{obs}} = \{O_1, O_2, \dots, O_m\}$ , and observation process  $\mathbf{p}$ , where  $\mathbf{P}$  is a true appropriate description of  $\mathbf{p}$ ,  $\mathbf{DDC}(\Gamma_{\text{hyp}}, \Gamma_{\text{obs}} | \mathbf{p})$  is fully determined by  $\Pr(O_i | \mathbf{P} \& H_j)$  for all  $i$  and  $j$ .

Both  $\mathbf{DDC}_{\text{ADD-AD}}$  and  $\mathbf{DDC}_{\text{ADD-SD}}$  meet **Likelihoods**, but  $\mathbf{DDC}_{\text{ADD-AD}^*}$  does not.<sup>37</sup> Given this, and given that we want a measure of degree of discrimination-conduciveness on which **Likelihoods** holds, we prefer  $\mathbf{DDC}_{\text{ADD-AD}}$  and  $\mathbf{DDC}_{\text{ADD-SD}}$  over  $\mathbf{DDC}_{\text{ADD-AD}^*}$ .

To explain why we want a measure of degree of discrimination-conduciveness on which **Likelihoods** holds, we return to TUBERCULOSIS, where  $k$  (the good test kit) is such that

$$(5.3.1) \quad \Pr(k \text{ says that } S \text{ has tuberculosis} | \mathbf{P} \& S \text{ has tuberculosis}) = 0.99$$

$$(5.3.2) \quad \Pr(k \text{ says that } S \text{ has tuberculosis} | \mathbf{P} \& S \text{ does not have tuberculosis}) = 0.05$$

Here  $S$  is a random member of the population. It follows both by  $\mathbf{DDC}_{\text{ADD-AD}}$  and  $\mathbf{DDC}_{\text{ADD-SD}}$  that  $\mathbf{DDC}(\Gamma_{\text{hyp}}, \Gamma_{\text{obs}} | \mathbf{p})$  is high. This is because it follows both by  $\mathbf{d}_{\text{AD}}$  and by  $\mathbf{d}_{\text{SD}}$  that each member of  $\Gamma_{\text{obs}}$  discriminates between the two members of  $\Gamma_{\text{hyp}}$  to a high degree:

$$(5.3.3) \quad \mathbf{d}_{\text{AD}}(H_t, H_{\sim t}, O_t | \mathbf{P}) = \mathbf{d}_{\text{AD}}(H_t, H_{\sim t}, O_{\sim t} | \mathbf{P}) = 0.94$$

$$(5.3.4) \quad \mathbf{d}_{\text{SD}}(H_t, H_{\sim t}, O_t | \mathbf{P}) = \mathbf{d}_{\text{SD}}(H_t, H_{\sim t}, O_{\sim t} | \mathbf{P}) \approx 0.884$$

But things are different with  $\mathbf{DDC}_{\text{ADD-AD}^*}$ , since whether  $\mathbf{d}_{\text{AD}^*}(H_t, H_{\sim t}, O_t | \mathbf{P})$  and  $\mathbf{d}_{\text{AD}^*}(H_t, H_{\sim t}, \sim O_t | \mathbf{P})$  are high depends on  $\Pr(O_t | \mathbf{P})$  and thus depends indirectly on  $\Pr(H_t | \mathbf{P})$  along with the

---

<sup>36</sup> This claim and the prior claim that  $\mathbf{DDC}_{\text{ADD-AD}^*}$  is an instance of  $\mathbf{DDC}_{\text{ADD}}$  are based on the assumption noted in footnote 35. Neither claim holds if this assumption is dropped.

<sup>37</sup> This is shown in Appendix E.



base-rate of tuberculosis in the population as given by  $\Pr(S \text{ has tuberculosis} \mid P)$ . If, say,  $\Pr(S \text{ has tuberculosis} \mid P)$  equals 0.1 and so  $\Pr(H_t \mid P)$  equals 0.1, then:

$$(5.3.5) \quad \mathbf{d}_{AD^*}(H_t, H_{\sim t}, O_t \mid P) \approx 6.528 > 1.098 \approx \mathbf{d}_{AD^*}(H_t, H_{\sim t}, O_{\sim t} \mid P)$$

$$(5.3.6) \quad \mathbf{DDC}_{ADD-AD^*}(\Gamma_{\text{hyp}}, \Gamma_{\text{obs}} \mid p) \approx 3.813$$

If, instead,  $\Pr(S \text{ has tuberculosis} \mid P)$  equals 0.5 and thus  $\Pr(H_t \mid P)$  equals 0.5, then:

$$(5.3.7) \quad \mathbf{d}_{AD^*}(H_t, H_{\sim t}, O_t \mid P) \approx 1.810 < 1.958 \approx \mathbf{d}_{AD^*}(H_t, H_{\sim t}, O_{\sim t} \mid P)$$

$$(5.3.8) \quad \mathbf{DDC}_{ADD-AD^*}(\Gamma_{\text{hyp}}, \Gamma_{\text{obs}} \mid p) \approx 1.883$$

This kind of sensitivity to priors seems out of place in the context of degree of discrimination (and degree of discrimination-conduciveness). As mentioned earlier (Section 3), a tuberculosis test kit's ability to discriminate between "S has tuberculosis" and "S does not have tuberculosis" does not depend on whether tuberculosis is common or rare. Of course the *success rate* of a tuberculosis test kit depends on the frequency of tuberculosis in the population in question. However, these relevant features are extrinsic to the test kit.

It's important to note that *discrimination* is distinct from *confirmation*, where the latter is understood so that, given P, O confirms H precisely when  $\Pr(H \mid P \ \& \ O)$  is greater than  $\Pr(H \mid P)$ . This difference is evident from the fact that there are cases where, given P, O discriminates between  $H_1$  and  $H_2$ , but confirms neither of them, because  $\Pr(H_1 \mid P \ \& \ O)$  is less than  $\Pr(H_1 \mid P)$ , and  $\Pr(H_2 \mid P \ \& \ O)$  is less than  $\Pr(H_2 \mid P)$ .<sup>38</sup> We leave it open that sensitivity to priors is a virtue, not a vice, in the context of degree of confirmation (and degree of "confirmation-conduciveness").

This is not the end of the story on alternatives to  $\mathbf{DDC}_{ADD-AD}$  and  $\mathbf{DDC}_{ADD-SD}$ , but we want to move forward, using the supposition that  $\mathbf{DDC}_{ADD-AD}$  and  $\mathbf{DDC}_{ADD-SD}$  are preferable to their alternatives.

---

<sup>38</sup> Consider a case where a card is randomly drawn from a standard well-shuffled deck of cards. Let  $H_1$  be the proposition that the card drawn is a Diamond,  $H_2$  be the proposition that the card drawn is a Heart, and O be the proposition that the card drawn is a Club, a Spade, the Ace of Diamonds, the Ace of Hearts, or the King of Hearts. Let P be a true appropriate description of your observation process. It follows that:

$$\Pr(O \mid P \ \& \ H_1) = 1/13 \neq 2/13 = \Pr(O \mid P \ \& \ H_2)$$

$$\Pr(H_1 \mid P \ \& \ O) = 1/29 < 1/4 = \Pr(H_1 \mid P)$$

$$\Pr(H_2 \mid P \ \& \ O) = 2/29 < 1/4 = \Pr(H_2 \mid P)$$

Hence, given P, O discriminates between  $H_1$  and  $H_2$ , but confirms neither of them.

## 5.4 Section Summary

We proposed a framework for measuring degree of discrimination-conduciveness, and set out two *prima facie* plausible instances of it, the absolute difference measure ( $\mathbf{DDC}_{\text{ADD-AD}}$ ) and the squared difference measure ( $\mathbf{DDC}_{\text{ADD-SD}}$ ). Each satisfies **Minimality**, **Dominance**, and **Likelihoods**. Given this, and given that we now see no clear reason for choosing between them, we keep both on the table and leave for the future the investigation of whether either is preferable to the other (and also whether either is preferable to alternative measures that obey **Minimality**, **Dominance**, and **Likelihoods**).

It will help, though, for ease of presentation in what follows, to focus on just one of them. Henceforth when we talk about “ $\mathbf{DDC}_{\text{ADD}}$ ” we’ll mean  $\mathbf{DDC}_{\text{ADD-AD}}$ . None of our main points below hinges on this choice between the two measures, this despite the fact that  $\mathbf{DDC}_{\text{ADD-AD}}$  and  $\mathbf{DDC}_{\text{ADD-SD}}$  are not ordinarily equivalent.

We turn now to the related issue of how to measure degree of OSE.

## 6 How to Measure Degree of OSE

Recall from Section 1 that there’s a distinction between degree of discrimination-conduciveness and degree of OSE. The latter, but not the former, is causal. OSEs arise when changing the observation process changes the degree of discrimination-conduciveness. Discrimination-conduciveness is to the size of a balloon as OSEs are to the amount of air you put into the balloon. We have constructed a framework for measuring degree of discrimination-conduciveness. The task now is to use it as a basis for constructing a measure of degree of OSE.

A natural idea is to measure degree of OSE in a particular case by taking the difference between the two degrees of discrimination-conduciveness in question:

$$\mathbf{OSE}_D(\Gamma_{\text{hyp}}, \Gamma_{\text{obs}}, p_1 \text{ rather than } p_2): \mathbf{DDC}_{\text{ADD}}(\Gamma_{\text{hyp}}, \Gamma_{\text{obs}} | p_1) - \mathbf{DDC}_{\text{ADD}}(\Gamma_{\text{hyp}}, \Gamma_{\text{obs}} | p_2)$$

If, for instance, you found yourself in FISH-B, and the alternative observation process on hand was the one in FISH-S, then the degree of OSE would be given by:

$$\mathbf{OSE}_D(\Gamma_{\text{hyp}}, \Gamma_{\text{obs}}, p_B \text{ rather than } p_S): \mathbf{DDC}_{\text{ADD}}(\Gamma_{\text{hyp}}, \Gamma_{\text{obs}} | p_B) - \mathbf{DDC}_{\text{ADD}}(\Gamma_{\text{hyp}}, \Gamma_{\text{obs}} | p_S)$$

If, instead, you found yourself in FISH-M, not FISH-B, and the alternative observation process on hand was the one in FISH-S, then the degree of OSE would be given by:

$$\mathbf{OSE}_D(\Gamma_{\text{hyp}}, \Gamma_{\text{obs}}, p_M \text{ rather than } p_S): \mathbf{DDC}_{\text{ADD}}(\Gamma_{\text{hyp}}, \Gamma_{\text{obs}} | p_M) - \mathbf{DDC}_{\text{ADD}}(\Gamma_{\text{hyp}}, \Gamma_{\text{obs}} | p_S)$$

Given that  $\mathbf{DDC}_{\text{ADD}}(\Gamma_{\text{hyp}}, \Gamma_{\text{obs}} | p_B)$  is minimal and thus is less than  $\mathbf{DDC}_{\text{ADD}}(\Gamma_{\text{hyp}}, \Gamma_{\text{obs}} | p_M)$ , it follows that though both  $\mathbf{OSE}_D(\Gamma_{\text{hyp}}, \Gamma_{\text{obs}} | p_B \text{ rather than } p_S)$  and  $\mathbf{OSE}_D(\Gamma_{\text{hyp}}, \Gamma_{\text{obs}} | p_M \text{ rather than } p_S)$  are negative, the former is less than the latter. This is just as it should be. The negative effect of using  $p_B$  rather than  $p_S$  is greater than the negative effect of using  $p_M$  rather than  $p_S$ .

Another natural idea is to measure degree of OSE in a particular case by taking the ratio of the two degrees of discrimination-conduciveness in question:

$$\mathbf{OSE}_R(\Gamma_{\text{hyp}}, \Gamma_{\text{obs}} | p_1 \text{ rather than } p_2) : \frac{\mathbf{DDC}_{\text{ADD}}(\Gamma_{\text{hyp}}, \Gamma_{\text{obs}} | p_1)}{\mathbf{DDC}_{\text{ADD}}(\Gamma_{\text{hyp}}, \Gamma_{\text{obs}} | p_2)}$$

The right-hand side here is greater than 1 precisely when the right-hand side of  $\mathbf{OSE}_D$  is positive. However,  $\mathbf{OSE}_D$  and  $\mathbf{OSE}_R$  are not ordinally equivalent.<sup>39</sup>

We see no clear basis for choosing between  $\mathbf{OSE}_D$  and  $\mathbf{OSE}_R$ . But, as with the fact that  $\mathbf{DDC}_{\text{ADD-AD}}$  and  $\mathbf{DDC}_{\text{ADD-SD}}$  aren't ordinally equivalent, this is okay for our purposes. We return to this issue in Section 9.

## 7 The Ravens Paradox

The ravens paradox can be understood in terms of the following three theses:

Nicod's Condition: For any object  $a$ , and predicates  $F$  and  $G$ ,  $(\forall x)(Fx \supset Gx)$  is confirmed by  $Fa \ \& \ Ga$ .

Equivalence Condition: For any propositions  $E$ ,  $H$ , and  $H^*$ , if (i)  $H$  and  $H^*$  are logically equivalent to each other and (ii)  $E$  confirms  $H$ , then  $E$  confirms  $H^*$ .

Ravens Condition:  $(\forall x)(Rx \supset Bx)$  is not confirmed by  $\sim Ba \ \& \ \sim Ra$ , where “ $Rx$ ” means “ $x$  is a raven” and “ $Bx$ ” means “ $x$  is black.”

The alleged paradox is that each of these three theses is *prima facie* plausible in isolation, but they are inconsistent as a set.

A standard Bayesian response is to reject RC in favor of one or both of the following:

$$(7.1) \quad \mathbf{c}((\forall x)(Rx \supset Bx), \sim Ba \ \& \ \sim Ra) = \varepsilon > n$$

$$(7.2) \quad \mathbf{c}((\forall x)(Rx \supset Bx), Ra \ \& \ Ba) \gg \mathbf{c}((\forall x)(Rx \supset Bx), \sim Ba \ \& \ \sim Ra)$$

Here  $n$  is the neutral point between confirmation and disconfirmation (where the evidence neither increases nor decreases the probability of the hypothesis), and  $\varepsilon$  is some value very close to  $n$ . (7.1) says that the degree to which  $(\forall x)(Rx \supset Bx)$  is confirmed by  $\sim Ba \ \& \ \sim Ra$  is negligible.

---

<sup>39</sup> This is shown in Appendix F.

(7.2) says that the degree to which  $(\forall x)(Rx \supset Bx)$  is confirmed by  $Ra$  &  $Ba$  is much greater than the degree to which it is confirmed by  $\sim Ba$  &  $\sim Ra$ .<sup>40</sup>

Our preferred measures of degree of discrimination-conduciveness and degree of OSE don't bear directly on (7.1) or (7.2), and don't indicate what to do in light of the fact that the three raven theses are jointly inconsistent. However, there is a connection.

Suppose you want to test the following hypotheses against each other:

H<sub>100</sub>: 100% of all ravens are black.

H<sub>50</sub>: 50% of all ravens are black.

Suppose, further, you want to do this by drawing a random sample and determining whether it includes any non-black ravens. Should your sample be drawn from the population of ravens? Should it instead be drawn from the population of non-black things? Should it instead be drawn from some third population?

Our framework for measuring degree of discrimination-conduciveness can help here. Let  $p$  be the process of sampling at random from the class of ravens, and  $p^*$  be the process of sampling at random from the class of non-black things. Suppose that if  $p$  is used, the observation set is  $\Gamma_{\text{obs}} = \{O_1, O_2\}$ , where:

O<sub>1</sub>:  $Ba$

O<sub>2</sub>:  $\sim Ba$

and suppose that if  $p^*$  is used, the observation set is  $\Gamma^*_{\text{obs}} = \{O^*_1, O^*_2\}$ , where:

O<sup>\*</sup><sub>1</sub>:  $Ra$

O<sup>\*</sup><sub>2</sub>:  $\sim Ra$

It follows that:

$$(7.3) \quad \Pr(O_1 \mid P \ \& \ H_{100}) = 1 > 1/2 = \Pr(O_1 \mid P \ \& \ H_{50})$$

$$(7.4) \quad \Pr(O_2 \mid P \ \& \ H_{100}) = 0 < 1/2 = \Pr(O_2 \mid P \ \& \ H_{50})$$

Given that the class of non-black things is much larger than the class of ravens, it further follows that:

$$(7.5) \quad \Pr(O^*_1 \mid P^* \ \& \ H_{100}) = 0 \approx \Pr(O^*_1 \mid P^* \ \& \ H_{50})$$

$$(7.6) \quad \Pr(O^*_2 \mid P^* \ \& \ H_{100}) = 1 \approx \Pr(O^*_2 \mid P^* \ \& \ H_{50})$$

---

<sup>40</sup> See Fitelson and Hawthorne (2010) for helpful discussion of Bayesian defenses of (7.1) and (7.2).

But then:

$$(7.7) \quad \mathbf{d}(H_{100}, H_{50}, O_1 | P) > \mathbf{d}(H_{100}, H_{50}, O^*_1 | P^*)$$

$$(7.8) \quad \mathbf{d}(H_{100}, H_{50}, O_1 | P) > \mathbf{d}(H_{100}, H_{50}, O^*_2 | P^*)$$

$$(7.9) \quad \mathbf{d}(H_{100}, H_{50}, O_2 | P) > \mathbf{d}(H_{100}, H_{50}, O^*_1 | P^*)$$

$$(7.10) \quad \mathbf{d}(H_{100}, H_{50}, O_2 | P) > \mathbf{d}(H_{100}, H_{50}, O^*_2 | P^*)$$

Hence:

$$(7.11) \quad \mathbf{DDC}_{ADD}(\Gamma_{hyp}, \Gamma_{obs} | p) > \mathbf{DDC}_{ADD}(\Gamma_{hyp}, \Gamma^*_{obs} | p^*)$$

This captures the intuitive idea that  $p$  is a better observational process than  $p^*$  in the context of testing  $H_{100}$  and  $H_{50}$  against each other.<sup>41</sup>

Sampling at random from the ravens and sampling at random from the non-black things are the two obvious observation processes to consider, but there are other possible processes, and they also can be evaluated by using our framework.<sup>42</sup> It is interesting that the ravens paradox provides an example in which it makes sense to compare different observational processes and different observation sets relative to a single set of competing hypotheses.<sup>43</sup>

Eddington's fishing example may seem worlds away from the ravens paradox, but they are connected by the fact that an observation process can diminish (or enhance!) the extent to which observations discriminate between competing hypotheses.

---

<sup>41</sup> See Forster (1994) for a similar treatment of the ravens paradox, which is formulated in terms of predictive accuracy.

<sup>42</sup> In his discussion of the ravens paradox, Royall (1997, pp. 177-179) considers several sampling schemes, but he doesn't consider sampling at random from the non-black things, nor does he consider possible but nonactual observations.

<sup>43</sup> Unsurprisingly, our treatment of the ravens paradox carries over straightforwardly to the grue paradox. Let  $H_1$  be the hypothesis that all emeralds are green and  $H_2$  be the hypothesis that all emeralds are grue (where an object is grue at time  $t$  precisely when it is green and  $t < 2050$  or it is blue and  $t \geq 2050$ ). You are going to sample 100 emeralds. In Situation 1, you do your sampling before 2050. In Situation 2, you do your sampling at or after 2050. In Situation 1, your observation process (described by proposition  $P$ ) takes over in that  $\Pr(O_{100} | P \ \& \ H_1) = 1 = \Pr(O_{100} | P \ \& \ H_2)$  where  $O_{100}$  is the claim that 100% of the emeralds sampled are green, and  $\Pr(O_i | P \ \& \ H_1) = 0 = \Pr(O_i | P \ \& \ H_2)$  for each  $i < 100$ . In Situation 2, your observation process (described by propositions  $P^*$ ) does not take over, since  $\Pr(O_{100} | P^* \ \& \ H_1) = 1 > 0 = \Pr(O_{100} | P^* \ \& \ H_2)$ .

## 8 Publication Bias

The ravens paradox is an old standby in philosophy of science, but the topic of this section is something more contemporary. Various sciences, including medicine and psychology, have recently been overtaken by a “replication crisis,” wherein a large proportion of results reported in refereed journals fail to be replicated.<sup>44</sup> One possible explanation that has been discussed is publication bias. One type of publication bias occurs when researchers obtain negative results but don't try to publish them, and then they “try and try again” until they come up with a positive result, which they then submit to a journal. Another type of publication bias occurs when journal editors are disinclined to accept papers that report negative results though they are happy to do so for submitted papers that report positive findings.<sup>45</sup>

To illustrate how our account of OSEs applies to publication bias, suppose that journal *Positive-and-Negative* publishes every well-designed and properly run study submitted to it regardless of the results, and that journal *Positive-Only* publishes every well-designed and properly run study submitted to it in which the results are positive but none in which the results are negative. Consider two scenarios:

*Scenario 1:* Smith reads *Positive-and-Negative* (and no other journal), and observes that the results are positive in every study published there on the effectiveness of drug D.

*Scenario 2:* Smith reads *Positive-Only* (and no other journal), and observes that the results are positive in every study published there on the effectiveness of drug D.

It seems clear here that:

- (8.1)  $\Pr(\text{the results are positive in every published study that Smith has read on the effectiveness of drug D} \mid P_{S1} \ \& \ D \text{ is more effective than a placebo}) \gg \Pr(\text{the results are positive in every published study that Smith has read on the effectiveness of drug D} \mid P_{S1} \ \& \ D \text{ is not more effective than a placebo})$
- (8.2)  $\Pr(\text{the results are positive in every published study that Smith has read on the effectiveness of drug D} \mid P_{S2} \ \& \ D \text{ is more effective than a placebo}) = \Pr(\text{the results are positive in every published study that Smith has read on the effectiveness of drug D} \mid P_{S2} \ \& \ D \text{ is not more effective than a placebo})$

Here  $P_{S1}$  and  $P_{S2}$  describe Smith's observation processes  $p_{s1}$  and  $p_{s2}$  (but not the observational outcomes!) in *Scenario 1* and *Scenario 2*, respectively. If Smith shifts from  $p_{s1}$  to  $p_{s2}$ , a regrettable OSE has occurred; if Smith shifts in the opposite direction, an OSE has also occurred, but this time it is all to the good.

---

<sup>44</sup> See Bird (forthcoming) for discussion and references.

<sup>45</sup> For discussion of, and references on, publication bias in medical science, see Stegenga (2018).

This example does not require us to quantify how much of an OSE occurs in the shift, but the example can be modified to make that pertinent. All that is needed is the supposition that journals differ in their degrees of publication bias where these are characterized by providing the values of the probabilities in (8.1) and (8.2).

In the fishing examples discussed at the start of this paper, we spoke of “your” using a fishing net and “your” observing the fish caught in the net. The experimenter and the observer of the outcome are one and the same person. However, in the case just described, Smith reads the studies, but she is not the scientist who carried out the trials. This allows for the possibility that the scientists carrying out their studies did flawless work, and yet others are entitled to look with jaundiced eye on these same studies when they read them.

## 9 Concluding Comments

In this paper, we conceptualized OSEs by considering how a shift from one process of observation to another affects discrimination-conduciveness – the degree to which possible observations discriminate between hypotheses, given the observation process at work. This OSE concept is causal. The cause is the shift in process. The effect is a change in degree of discrimination-conduciveness. We described conditions of adequacy that an acceptable measure of degree of discrimination-conduciveness must satisfy, and used those conditions of adequacy to evaluate several possible measures. We then defined two measures of how much of an OSE the shift from one process to another induces.

We were driven to consider choices of measure by the fact that discrimination-conduciveness and OSEs are matters of degree. If they are matters of degree, one is obliged to say degrees of *what*. Although it turned out that each choice of measure faces a problem of measure sensitivity, our goal was not to solve that pair of problems. The situation here is rather similar to the one that Bayesians face when they talk about confirmation. They all agree that O confirms H precisely when  $\Pr(H | O) > \Pr(H)$ , but they disagree about how degree of confirmation should be understood. The disagreement is substantive, since several measures fail to be ordinarily equivalent (Fitelson 1999, Brössel 2013). But even if no uniquely correct measure of degree of confirmation can be defended, Bayesian confirmation theory still has its uses.

In the theory of confirmation, rarely is it important to be able to compare  $\mathbf{c}(H_1, O_1 | P)$  with  $\mathbf{c}(H_2, O_2 | P)$ , where  $\mathbf{c}(-)$  is a measure of degree of confirmation, and  $H_1$  and  $H_2$  are on completely different subject matters, as are  $O_1$  and  $O_2$ . More often, the interesting questions concern whether  $\mathbf{c}(H_1, O | P) > \mathbf{c}(H_2, O | P)$ , and whether  $\mathbf{c}(H, O_1 | P) > \mathbf{c}(H, O_2 | P)$ . Similar points apply to degrees of discrimination conduciveness. Rarely is it important to be able to decide whether  $\mathbf{DDC}(\Gamma_{\text{hyp}}, \Gamma_{\text{obs}} | p) > \mathbf{DDC}(\Gamma^*_{\text{hyp}}, \Gamma^*_{\text{obs}} | p^*)$ . More often, the interesting questions concern whether  $\mathbf{DDC}(\Gamma_{\text{hyp}}, \Gamma_{\text{obs}} | p) > \mathbf{DDC}(\Gamma_{\text{hyp}}, \Gamma_{\text{obs}} | p^*)$ . However, the problem of measure-sensitivity arises even in this restricted context.

Perhaps, in many cases of interest, the ordering of processes in terms of their discrimination-conduciveness is the same regardless of which of several “reasonable” measures you choose. In cases where this unanimity does not arise, it is interesting that different choices of measure lead to different answers. That is a discovery worth making, not a cause for despair. On the other hand, maybe a single best measure of discrimination-conduciveness can be found. There are questions here that merit further investigation.

A *locus classicus* for discussion of OSEs is the fine-tuning argument; since we have not discussed that argument to this point, a few comments are in order. Some have argued that an OSE completely vitiates this argument; they contend that the observation that a physical constant has a value that falls in a very narrow window of life-permitting values fails to discriminate between the competing hypotheses, since the observation was made by beings who are themselves alive (see, e.g., Sober 2009, 2018). Critics of that negative assessment respond that a proper understanding of the epistemology of OSEs reveals that no such vitiation arises (see, e.g., Leslie 1989, Monton 2006, Collins 2009, and Kotzen 2012 for discussion). Much of this disagreement concerns how an “appropriate” description of the process of observation should be formulated, a topic we mentioned earlier but did not explore. The present paper, therefore, does not by itself resolve the controversy about the fine-tuning argument. We hope, however, that it does contribute to this discussion, in that it makes precise how an observation process’s effect on the ability of observations to discriminate between hypotheses should be understood.

One such contribution that we see derives from the fact that OSEs, understood in terms of discrimination-conduciveness, depend on the space of possible observations considered. The fine-tuning argument is usually formulated by considering a dichotomous property – our universe either permits life to exist, or it does not. The question is then asked what the probability is that our universe is life-permitting according to this or that cosmological or theistic hypothesis. However, recent work on the physics of fine-tuning suggests that this dichotomous characterization of the problem may be too crude. Instead of saying whether or not our universe is life-permitting, we might describe how life-friendly our universe is. The possibilities range from life’s being impossible to life’s being inevitable, with lots of other possibilities in between.<sup>46</sup> Suppose our universe is such that life is possible, but is very improbable. When living observers determine how life-friendly our universe is, that fact about the observers isn’t enough to settle what degree of life-friendliness our universe must manifest. This means that even if a massive OSE arises when we consider the dichotomous property, it doesn’t follow that a massive shut-down also occurs when more fine-grained characteristics are considered (Sober 2018).

Since our OSE concept applies to net-fishing, RED LIGHTS, TUBERCULOSIS, the ravens paradox, publication bias, stroke-impaired visual acuity, and the fine-tuning argument, the worry may arise that the account we have given is too broad to be of much use. We disagree. We think

---

<sup>46</sup> See Lewis and Barnes (2016, Ch. 3) for helpful discussion.



our account *unifies* these examples, and *explains* why OSEs often (but not always) impose epistemic costs.

Our OSE concept, though broad, is not a vacuous catch-all that encompasses all forms of bad research practice. This can be seen by noting that research has at least five distinct (but related) stages:

- (9.1) formulating a question or problem
- (9.2) formulating competing hypotheses
- (9.3) choosing a method of gathering observations
- (9.4) making observations
- (9.5) interpreting how the observations bear on the competing hypotheses, given the observation process used

Our OSE framework comes into play at the stage of choosing a method of gathering observations, but flawed research practices also can be found at the other stages. An obvious example is the use of a fallacious mode of inference at stage five (e.g., the base-rate fallacy).<sup>47</sup> Our OSE framework is a tool, not for evaluating research practices as a whole, but for evaluating a limited though central part of that variegated totality.

**Acknowledgments:** We thank the two anonymous referees for this journal for helpful comments. We also thank the participants at *Texas Epistemology Xtravaganza 2019* for helpful discussion.

## Appendix A

We aim to show:

- (A.1) If  $\Gamma_{\text{hyp}}$  is a partition, then it follows from **Minimality** that  $\text{DDC}(\Gamma_{\text{hyp}}, \Gamma_{\text{obs}} \mid p)$  is minimal precisely when  $\Pr(O_i \mid P \ \& \ H_j) = \Pr(O_i \mid P)$  for all  $i$  and  $j$ .

Take some  $\Gamma_{\text{hyp}}$ ,  $\Gamma_{\text{obs}}$ , and  $p$  such that  $\Gamma_{\text{hyp}}$  is a partition. Suppose, first, that:

---

<sup>47</sup> Additional examples include: (a) ignoring a relevant alternative hypothesis, (b) failing to take proper account of imprecision in an observation instrument, (c) giving a spurious burden of proof argument, (d) making a facile application of Ockham's razor, (e) using an indefensible prior probability, and (f) using a rejectionist significance test.

$$(A.2) \quad \Pr(O_i | P \& H_j) = \Pr(O_i | P) \text{ for all } i \text{ and } j.$$

It follows from (A.2) that:

$$(A.3) \quad \begin{aligned} \Pr(O_1 | P \& H_1) &= \Pr(O_1 | P) = \Pr(O_1 | P \& H_2) = \dots = \Pr(O_1 | P \& H_n). \\ \Pr(O_2 | P \& H_1) &= \Pr(O_2 | P) = \Pr(O_2 | P \& H_2) = \dots = \Pr(O_2 | P \& H_n). \\ &\dots \\ \Pr(O_m | P \& H_1) &= \Pr(O_m | P) = \Pr(O_m | P \& H_2) = \dots = \Pr(O_m | P \& H_n). \end{aligned}$$

Hence:

$$(A.4) \quad \Pr(O_i | P \& H_j) = \Pr(O_i | P \& H_k) \text{ for all } i, j, \text{ and } k.$$

Hence by **Minimality** it follows that  $\mathbf{DDC}(\Gamma_{\text{hyp}}, \Gamma_{\text{obs}} | p)$  is minimal. Suppose, second, that by **Minimality** it follows that  $\mathbf{DDC}(\Gamma_{\text{hyp}}, \Gamma_{\text{obs}} | p)$  is minimal. Then:

$$(A.5) \quad \Pr(O_i | P \& H_j) = \Pr(O_i | P \& H_k) \text{ for all } i, j, \text{ and } k.$$

The law of total probability implies that:

$$(A.6) \quad \Pr(O_i | P) = \Pr(H_1 | P)\Pr(O_i | P \& H_1) + \Pr(H_2 | P)\Pr(O_i | P \& H_2) + \dots + \Pr(H_n | P)\Pr(O_i | P \& H_n).$$

(A.5) and (A.6) together imply:

$$(A.7) \quad \Pr(O_i | P) = \Pr(O_i | P \& H_1)[\Pr(H_1 | P) + \Pr(H_2 | P) + \dots + \Pr(H_n | P)].$$

Hence, since  $\Gamma_{\text{hyp}}$  is a partition, it follows that the sum on the right-hand side of (A.7) equals 1 and thus:

$$(A.8) \quad \Pr(O_i | P) = \Pr(O_i | P \& H_1).$$

Given (A.5), it follows that:

$$(A.9) \quad \Pr(O_i | P \& H_j) = \Pr(O_i | P) \text{ for all } i \text{ and } j.$$

Hence (A.1). QED.

## Appendix B

The aim is to show:

$$(B.1) \quad \text{There are cases where } \Gamma_{\text{hyp}} \text{ is not a partition, } \mathbf{d}(H_i, H_j, O_k | P) \text{ is minimal for all } i, j, \text{ and } k, \text{ and it's not the case that } \Pr(O_i | P \& H_j) = \Pr(O_i | P) \text{ for all } i \text{ and } j.$$

Let  $\Gamma_{\text{hyp}} = \{H_1, H_2\}$ ,  $\Gamma^*_{\text{hyp}} = \{H_1, H_2, H_3\}$ , and  $\Gamma_{\text{obs}} = \{O_1, O_2, O_3\}$ . Suppose that  $\Gamma^*_{\text{hyp}}$  is a partition but  $\Gamma_{\text{hyp}}$  isn't. Take the following (partially displayed) probability distribution:

O <sub>1</sub>	O <sub>2</sub>	O <sub>3</sub>	H <sub>1</sub>	H <sub>2</sub>	H <sub>3</sub>	Pr(-   P)
T	F	F	T	F	F	$\frac{1}{6}$
T	F	F	F	T	F	$\frac{1}{6}$
T	F	F	F	F	T	$\frac{1}{3}$
F	T	F	F	F	T	$\frac{1}{3}$

It follows that:

$$(B.2) \quad \Pr(O_1 | P \ \& \ H_1) = 1 = \Pr(O_1 | P \ \& \ H_2)$$

$$(B.3) \quad \Pr(O_2 | P \ \& \ H_1) = 0 = \Pr(O_2 | P \ \& \ H_2)$$

$$(B.4) \quad \Pr(O_3 | P \ \& \ H_1) = 0 = \Pr(O_3 | P \ \& \ H_2)$$

$$(B.5) \quad \Pr(O_1 | P \ \& \ H_1) = 1 > 2/3 = \Pr(O_1 | P)$$

Hence (B.1). QED

## Appendix C

The aim is to show:

$$(C.1) \quad \text{For any hypothesis sets } \Gamma_{\text{hyp}} = \{H_1, H_2\} \text{ and } \Gamma^*_{\text{hyp}} = \{H^*_1, H^*_2\}, \text{ observation sets } \Gamma_{\text{obs}} = \{O, \sim O\} \text{ and } \Gamma^*_{\text{obs}} = \{O^*, \sim O^*\}, \text{ and observation processes } p \text{ and } p^*, \text{ where } P \text{ is a true appropriate description of } p, \text{ and where } P^* \text{ is a true appropriate description of } p^*, \mathbf{DDC}_{\text{ADD-AD}}(\Gamma_{\text{hyp}}, \Gamma_{\text{obs}} | p) > / = / < \mathbf{DDC}_{\text{ADD-AD}}(\Gamma^*_{\text{hyp}}, \Gamma^*_{\text{obs}} | p^*) \text{ if and only if } \mathbf{DDC}_{\text{ADD-SD}}(\Gamma_{\text{hyp}}, \Gamma_{\text{obs}} | p) > / = / < \mathbf{DDC}_{\text{ADD-SD}}(\Gamma^*_{\text{hyp}}, \Gamma^*_{\text{obs}} | p^*).$$

Take some hypothesis sets  $\Gamma_{\text{hyp}} = \{H_1, H_2\}$  and  $\Gamma^*_{\text{hyp}} = \{H^*_1, H^*_2\}$ , observation sets  $\Gamma_{\text{obs}} = \{O, \sim O\}$  and  $\Gamma^*_{\text{obs}} = \{O^*, \sim O^*\}$ , and observation processes  $p$  and  $p^*$ , where  $P$  is a true appropriate description of  $p$ , and where  $P^*$  is a true appropriate description of  $p^*$ . It follows that:

$$\begin{aligned}
\text{(C.2)} \quad & \mathbf{DDC}_{\text{ADD-AD}}(\Gamma_{\text{hyp}}, \Gamma_{\text{obs}} | p) > / = / < \mathbf{DDC}_{\text{ADD-AD}}(\Gamma_{\text{hyp}}^*, \Gamma_{\text{obs}}^* | p^*) \text{ iff} \\
& \left[ \frac{\left| \Pr(O | P \& H_1) - \Pr(O | P \& H_2) \right| + \left| \Pr(\sim O | P \& H_1) - \Pr(\sim O | P \& H_2) \right|}{2} \right] > / = / < \\
& \left[ \frac{\left| \Pr(O^* | P^* \& H_1^*) - \Pr(O^* | P^* \& H_2^*) \right| + \left| \Pr(\sim O^* | P^* \& H_1^*) - \Pr(\sim O^* | P^* \& H_2^*) \right|}{2} \right] \text{ iff} \\
& \left| \Pr(O | P \& H_1) - \Pr(O | P \& H_2) \right| > / = / < \\
& \left| \Pr(O^* | P^* \& H_1^*) - \Pr(O^* | P^* \& H_2^*) \right|
\end{aligned}$$

$$\begin{aligned}
\text{(C.3)} \quad & \mathbf{DDC}_{\text{ADD-SD}}(\Gamma_{\text{hyp}}, \Gamma_{\text{obs}} | p) > / = / < \mathbf{DDC}_{\text{ADD-SD}}(\Gamma_{\text{hyp}}^*, \Gamma_{\text{obs}}^* | p^*) \text{ iff} \\
& \left[ \frac{\left( \Pr(O | P \& H_1) - \Pr(O | P \& H_2) \right)^2 + \left( \Pr(\sim O | P \& H_1) - \Pr(\sim O | P \& H_2) \right)^2}{2} \right] > / = / < \\
& \left[ \frac{\left( \Pr(O^* | P^* \& H_1^*) - \Pr(O^* | P^* \& H_2^*) \right)^2 + \left( \Pr(\sim O^* | P^* \& H_1^*) - \Pr(\sim O^* | P^* \& H_2^*) \right)^2}{2} \right] \text{ iff} \\
& \left( \Pr(O | P \& H_1) - \Pr(O | P \& H_2) \right)^2 > / = / < \\
& \left( \Pr(O^* | P^* \& H_1^*) - \Pr(O^* | P^* \& H_2^*) \right)^2
\end{aligned}$$

But:

$$\begin{aligned}
\text{(C.4)} \quad & \left| \Pr(O | P \& H_1) - \Pr(O | P \& H_2) \right| > / = / < \\
& \left| \Pr(O^* | P^* \& H_1^*) - \Pr(O^* | P^* \& H_2^*) \right| \text{ iff} \\
& \left( \Pr(O | P \& H_1) - \Pr(O | P \& H_2) \right)^2 > / = / < \\
& \left( \Pr(O^* | P^* \& H_1^*) - \Pr(O^* | P^* \& H_2^*) \right)^2
\end{aligned}$$

Hence:

$$\begin{aligned}
\text{(C.5)} \quad & \mathbf{DDC}_{\text{ADD-AD}}(\Gamma_{\text{hyp}}, \Gamma_{\text{obs}} | p) > / = / < \mathbf{DDC}_{\text{ADD-AD}}(\Gamma_{\text{hyp}}^*, \Gamma_{\text{obs}}^* | p^*) \text{ if and only if} \\
& \mathbf{DDC}_{\text{ADD-SD}}(\Gamma_{\text{hyp}}, \Gamma_{\text{obs}} | p) > / = / < \mathbf{DDC}_{\text{ADD-SD}}(\Gamma_{\text{hyp}}^*, \Gamma_{\text{obs}}^* | p^*).
\end{aligned}$$

Hence (C.1). QED

## Appendix D

The aim is to show:

- (D.1) There are cases where  $\Gamma_{\text{hyp}} = \{H_1, H_2\}$ ,  $\Gamma_{\text{obs}} = \{O_1, O_2, O_3\}$ ,  $\Gamma^*_{\text{obs}} = \{O^*_1, O^*_2, O^*_3\}$ , and  $\mathbf{DDC}_{\text{ADD-AD}}(\Gamma_{\text{hyp}}, \Gamma_{\text{obs}} | p)$  is less than  $\mathbf{DDC}_{\text{ADD-AD}}(\Gamma_{\text{hyp}}, \Gamma^*_{\text{obs}} | p)$ , whereas  $\mathbf{DDC}_{\text{ADD-SD}}(\Gamma_{\text{hyp}}, \Gamma_{\text{obs}} | p)$  is greater than  $\mathbf{DDC}_{\text{ADD-SD}}(\Gamma_{\text{hyp}}, \Gamma^*_{\text{obs}} | p)$ .

Let  $\Gamma_{\text{hyp}} = \{H_1, H_2\}$ ,  $\Gamma_{\text{obs}} = \{O_1, O_2, O_3\}$ , and  $\Gamma^*_{\text{obs}} = \{O^*_1, O^*_2, O^*_3\}$ . Consider the following (partially displayed) probability distribution:

O <sub>1</sub>	O <sub>2</sub>	O <sub>3</sub>	O* <sub>1</sub>	O* <sub>2</sub>	O* <sub>3</sub>	H <sub>1</sub>	H <sub>2</sub>	Pr(-   P)
T	F	F	F	T	F	T	F	$\frac{383}{3072}$
T	F	F	F	F	T	T	F	$\frac{1}{3072}$
T	F	F	F	F	T	F	T	$\frac{1}{16}$
F	T	F	T	F	F	T	F	$\frac{511}{3072}$
F	T	F	F	T	F	T	F	$\frac{257}{3072}$
F	T	F	F	F	T	F	T	$\frac{1}{32}$
F	F	T	T	F	F	F	T	$\frac{43}{131072}$
F	F	T	F	T	F	T	F	$\frac{1}{8}$
F	F	T	F	T	F	F	T	$\frac{223}{1024}$
F	F	T	F	F	T	F	T	$\frac{24661}{131072}$

It follows that:

$$(D.2) \quad \mathbf{DDC}_{\text{ADD-AD}}(\Gamma_{\text{hyp}}, \Gamma_{\text{obs}} | p) = 0.375 < 0.3754 \approx \mathbf{DDC}_{\text{ADD-AD}}(\Gamma_{\text{hyp}}, \Gamma^*_{\text{obs}} | p)$$

$$(D.3) \quad \mathbf{DDC}_{\text{ADD-SD}}(\Gamma_{\text{hyp}}, \Gamma_{\text{obs}} | p) \approx 0.174 > 0.160 \approx \mathbf{DDC}_{\text{ADD-SD}}(\Gamma_{\text{hyp}}, \Gamma^*_{\text{obs}} | p)$$

Hence (D.1). QED

## Appendix E

The aim is to show:

- (E.1)  $\mathbf{DDC}_{\text{ADD-AD}^*}$  doesn't meet **Likelihoods**.

Let  $\Gamma_{\text{hyp}} = \{H_1, H_2, H_3\}$  and  $\Gamma_{\text{obs}} = \{O, \sim O\}$ . Consider, first, the following (partially displayed) probability distribution:

O	H <sub>1</sub>	H <sub>2</sub>	H <sub>3</sub>	Pr(-   P)
T	T	F	F	$\frac{27}{100}$
T	F	T	F	$\frac{1}{5}$
T	F	F	T	$\frac{2}{25}$
F	T	F	F	$\frac{9}{100}$
F	F	T	F	$\frac{1}{5}$
F	F	F	T	$\frac{4}{25}$

It follows that:

$$(E.2) \quad \Pr(O | P \ \& \ H_1) = 0.75 > 0.25 = \Pr(\sim O | P \ \& \ H_1)$$

$$(E.3) \quad \Pr(O | P \ \& \ H_2) = 0.5 = \Pr(\sim O | P \ \& \ H_2)$$

$$(E.4) \quad \Pr(O | P \ \& \ H_3) = 1/3 < 2/3 = \Pr(\sim O | P \ \& \ H_3)$$

$$(E.5) \quad \mathbf{DDC}_{\text{ADD-AD}^*}(\Gamma_{\text{hyp}}, \Gamma_{\text{obs}} | p) \approx 0.561$$

Consider, second, the following alternative (partially displayed) probability distribution:

O	H <sub>1</sub>	H <sub>2</sub>	H <sub>3</sub>	Pr(-   P)
T	T	F	F	$\frac{3}{50}$
T	F	T	F	$\frac{1}{10}$
T	F	F	T	$\frac{6}{25}$
F	T	F	F	$\frac{1}{50}$
F	F	T	F	$\frac{1}{10}$
F	F	F	T	$\frac{12}{25}$

Here (E.2), (E.3), and (E.4) hold but:

$$(E.6) \quad \mathbf{DDC}_{\text{ADD-AD}^*}(\Gamma_{\text{hyp}}, \Gamma_{\text{obs}} | p) \approx 0.579$$

Hence  $\mathbf{DDC}_{\text{ADD-AD}^*}$  fails to meet **Likelihoods**. Hence (E.1). QED

## Appendix F

The aim is to show:

$$(F.1) \quad \mathbf{OSE}_D \text{ and } \mathbf{OSE}_R \text{ are not ordinally equivalent.}$$

Let  $\Gamma_{\text{hyp}} = \{H, \sim H\}$  and  $\Gamma_{\text{obs}} = \{O_1, O_2, O_3\}$ , and consider the following (partially displayed) probability distribution:

O <sub>1</sub>	O <sub>2</sub>	O <sub>3</sub>	H	Pr(-   P <sub>1</sub> )	Pr(-   P <sub>2</sub> )	Pr(-   P <sub>3</sub> )	Pr(-   P <sub>4</sub> )
T	F	F	T	$\frac{238259}{2280396}$	$\frac{7975}{32756}$	$\frac{4125}{48184}$	$\frac{631533}{109342720}$
T	F	F	F	$\frac{41480}{570099}$	$\frac{60146}{122835}$	$\frac{1037}{6023}$	0
F	T	F	T	$\frac{238259}{1140198}$	$\frac{55825}{786144}$	$\frac{3375}{24092}$	$\frac{80204691}{109342720}$
F	T	F	F	$\frac{20740}{570099}$	$\frac{17629}{122835}$	0	$\frac{14030541}{54671360}$
F	F	T	T	$\frac{238259}{2280396}$	$\frac{7975}{786144}$	$\frac{4125}{48184}$	0
F	F	T	F	$\frac{269620}{570099}$	$\frac{1037}{24567}$	$\frac{3111}{6023}$	$\frac{222707}{54671360}$

It follows that:

$$(F.2) \quad \mathbf{OSE}_D(\Gamma_{\text{hyp}}, \Gamma_{\text{obs}} \mid p_1 \text{ rather than } p_2) \approx 0.354 > 0.306 \approx \mathbf{OSE}_D(\Gamma_{\text{hyp}}, \Gamma_{\text{obs}} \mid p_3 \text{ rather than } p_4)$$

$$(F.3) \quad \mathbf{OSE}_R(\Gamma_{\text{hyp}}, \Gamma_{\text{obs}} \mid p_1 \text{ rather than } p_2) = 18 < 30.4 = \mathbf{OSE}_R(\Gamma_{\text{hyp}}, \Gamma_{\text{obs}} \mid p_3 \text{ rather than } p_4)$$

Hence (F.1). QED

## References

- Barrett, M., and Sober, E. (forthcoming). “The Requirement of Total Evidence – A Reply to Epstein’s Critique.” *Philosophy of Science*.
- Berofsky, B. (1971). *Determinism*. Princeton: Princeton University Press.
- Bird, A. (forthcoming). “Understanding the Replication Crisis as a Base Rate Fallacy.” *British Journal for Philosophy of Science*.
- Bostrom, N. (2002). *Anthropic Bias: Observation Selection Effects in Science and Philosophy*. New York: Routledge.
- Bradley, D. (2011). “Confirmation in a Branching World: The Everett Interpretation and Sleeping Beauty.” *British Journal for Philosophy of Science* 62: 323–342.
- Brössel, P. (2013). “The Problem of Measure Sensitivity Redux.” *Philosophy of Science* 80: 378–397.

- Carnap, R. (1962). *Logical Foundations of Probability* (2nd ed.). Chicago: University of Chicago Press.
- Ćirković, M., Sandberg, A., and Bostrom, N. (2010). “Anthropic Shadow: Observation Selection Effects and Human Extinction Risks.” *Risk Analysis* 30: 1495–1506.
- Collins, R. (2009). “The Teleological Argument – An Exploration of the Fine-Tuning of the Universe.” In W. Craig and J. Moreland (eds.), *The Blackwell Companion to Natural Theology*. Wiley-Blackwell, pp. 202–281.
- Dawid, A. P. (1976). “Properties of Diagnostic Data Distributions.” *Biometrics* 32: 647–658.
- Earman, J. (1986). *A Primer on Determinism*. Dordrecht: D. Reidel.
- Eddington, A. (1939). *The Philosophy of Physical Science*. Cambridge: Cambridge University Press.
- Epstein, P. (2017). “The Fine-Tuning Argument and the Requirement of Total Evidence.” *Philosophy of Science* 84: 639–658.
- Fitelson, B. (1999). “The Plurality of Bayesian Measures of Confirmation and the Problem of Measure Sensitivity.” *Philosophy of Science* 66: S362–378.
- Fitelson, B., and Hawthorne, J. (2010). “How Bayesian Confirmation Theory Handles the Paradox of the Ravens.” In E. Eells and J. Fetzer (eds.), *The Place of Probability in Science*. Springer, pp. 247–275.
- Forster, M. (1994). “Non-Bayesian Foundations for Statistical Estimation, Prediction, and the Ravens Example.” *Erkenntnis* 40: 357–376.
- Hacking, I. (1965). *The Logic of Statistical Inference*. Cambridge: Cambridge University Press.
- Hájek, A. (2012). “Interpretations of Probability.” *The Stanford Encyclopedia of Philosophy* (Winter 2012 Edition), E. Zalta (ed.), URL = <http://plato.stanford.edu/archives/win2012/entries/probability-interpret/>.
- Kotzen, M. (2012). “Selection Biases in Likelihood Arguments.” *British Journal for Philosophy of Science* 63: 825–839.
- Leslie, J. (1989). *Universes*. London: Routledge.
- Lewis, G., and Barnes, L. (2016). *A Fortunate Universe: Life in a Finely Tuned Cosmos*. Cambridge: Cambridge University Press.
- Maher, P. (2007). “Explication Defended.” *Studia Logica* 86: 331–341.
- Manson, N. (2003). “Introduction.” In N. Manson (ed.), *God and Design – the Teleological Argument and Modern Science*. Routledge, pp. 1–23.



- Manson, N. (2009). "The Fine-Tuning Argument." *Philosophy Compass* 4: 271–286.
- Monton, B. (2006). "God, Fine-Tuning, and the Problem of Old Evidence." *British Journal for the Philosophy of Science* 57: 405–424.
- Olsson, E. (2015). "Gettier and the Method of Explication: A 60 Year Old Solution to a 50 Year Old Problem." *Philosophical Studies* 172: 57–72.
- Roberts, J. (2012). "Fine-Tuning and the Infrared Bull's-Eye." *Philosophical Studies* 160: 287–303.
- Roush, S. (2003). "Copernicus, Kant, and the Anthropic Cosmological Principles." *Studies in History and Philosophy of Science* 34: 5–35.
- Royall, R. (1997). *Statistical Evidence: A Likelihood Paradigm*. London: Chapman & Hall.
- Schaffer, J. (2005). "Contrastive Causation." *Philosophical Review* 114: 327–358.
- Schupbach, J. (forthcoming). "Experimental Explication." *Philosophy and Phenomenological Research*.
- Sober, E. (2000). *Philosophy of Biology* (2nd ed.). Boulder: Westview.
- Sober, E. (2003). "The Argument from Design." In N. Manson (ed.), *God and Design – the Teleological Argument and Modern Science*. Routledge, pp. 27–54. Reprinted with some changes in W. Mann (ed.), *The Blackwell Guide to Philosophy of Religion*, 2004, pp. 117–147.
- Sober, E. (2008). *Evidence and Evolution – the Logic Behind the Science*. Cambridge: Cambridge University Press.
- Sober, E. (2009). "Absence of Evidence and Evidence of Absence: Evidential Transitivity in Connection with Fossils, Fishing, Fine-Tuning, and Firing Squads." *Philosophical Studies* 143: 63–90.
- Sober, E. (2018). *The Design Argument*. Cambridge: Cambridge University Press.
- Stegenga, J. (2018). *Medical Nihilism*. Oxford: Oxford University Press.
- Titelbaum, M. (2010). "Tell Me You Love Me: Bootstrapping, Externalism, and No-Lose Epistemology." *Philosophical Studies* 149: 119–134.
- Weisberg, J. (2005). "Firing Squads and Fine-Tuning: Sober on the Design Argument." *British Journal for Philosophy of Science* 56: 809–821.
- White, R. (2003). "Fine-Tuning and Multiple Universes." In N. Manson (ed.), *God and Design – the Teleological Argument and Modern Science*. Routledge, pp. 229–250.