Analytic Philosophy

WILEY

ANALYTIC PHILOSOPHY

ORIGINAL ARTICLE OPEN ACCESS

What Second-Best Epistemology Could Be

Marc-Kevin Daoust

Département des enseignements généraux, École de Technologie Supérieure, Montreal, Canada

Correspondence: Marc-Kevin Daoust (marc-kevin.daoust@etsmtl.ca)

Received: 17 October 2022 | **Accepted:** 16 July 2024

Funding: This study was supported by the Fonds de recherche du Québec—Société et culture (Grant 268137) and the Social Sciences and Humanities Research Council (Grant 756-2019-0133).

Keywords: bounded rationality | epistemic ideals | imperfection | perfect rationality | second-best epistemology

ABSTRACT

According to the Theory of the Second Best, in non-ideal circumstances, approximating ideals might be suboptimal (with respect to a specific interpretation of what "approximating an ideal" means). In this paper, I argue that the formal model underlying the Theory can apply to problems in epistemology. Two applications are discussed: First, in some circumstances, second-best problems arise in Bayesian settings. Second, the division of epistemic labor can be subject to second-best problems. These results matter. They allow us to evaluate the claim, made by many philosophers, that second-best problems have import in epistemology (and the specific conditions under which the Theory finds applications). They also allow us to see that talk of "approximating an ideal" is ambiguous, and to clarify the conditions in which approximating an epistemic ideal might be beneficial.

Ideals are common in epistemology. Epistemic ideals often include impeccable reasoning, perfect evidence-responsiveness, full coherence, great cognitive capacities and so forth.[1] However, ordinary agents are not capable of meeting such ideals. Accordingly, some philosophers wonder why epistemic ideals matter to ordinary agents like us, who are imperfect in many relevant ways.[2]

Here is a tentative response to this worry: We should care about epistemic ideals because we can *approximate* them. We have an epistemic ideal (e.g., an ideal epistemic figure or a set of perfect epistemic requirements). Then, we have a guideline for determining what non-ideal agents should do: Approximate the ideal the best you can. The more you approximate the epistemic ideal, the better. Call this the Approximation Thesis.

Some researchers doubt that the Approximation Thesis is true. Their skepticism is based on the Theory of the Second Best. For instance, Egan says that, in accordance with the Theory of the Second Best, non-ideal agents do not necessarily fall under an obligation to "incorporate all of their reliably-obtained information about the world into a single unified corpus of beliefs that's active in guiding all of their behavior all of the time" (Egan 2008, 62), even if ideal agents do. Titelbaum (2015, 291) roughly says that, in non-ideal situations where various epistemic norms conflict with each other, we might need an epistemic theory of the second best (in contrast with an ideal theory). DiPaolo (2018, forthcoming) and Daoust (2021a) argue that noncompliance with ideal coherence norms can be a consequence of second-best problems in epistemology. Staffel (2019, 112) roughly says that, if epistemic ideals have multiple sources and do not always concur, then the approximation of epistemic ideals can give rise to second-best problems.[3] Karlan (2020) thinks that the study of reasoning shortcuts for boundedly rational agents lends support to the claim that there are second-best problems in epistemology. Lawson (2021) argues that second-best problems arise for the accumulation of (legal) knowledge. That is, in imperfect situations, limiting one's accumulation of knowledge can be the second-best option.

So, what is the Theory of the Second Best these philosophers refer to? Essentially, the Theory of the Second Best is a formal contribution to the field of utility (or welfare) optimization (Lipsey and Lancaster 1956). Its main conclusion is that, in non-ideal circumstances, approximating an ideal might be suboptimal.

Epistemologists are right to say that this *general conclusion*, or *interpretation*, finds import in epistemology. Yet, it's still unclear how the Theory's *mathematical model* applies to epistemic norms. The distinction between Theory's mathematical model and the Theory's conclusion matters. As Wiens rightly stresses:

> The original "general theory of second best" consists of an abstract mathematical model ("the model"), a theorem of that model ("the theorem"), and a mapping from the model's mathematical objects to concepts in economic theory ("an interpretation")... [Yet theorists in other fields] typically leave aside the math and focus on the economic interpretation of the model... Any interpretation of a model is bound to be limited in what it can show. (Wiens 2020, 5)

Also, if we stay at the level of interpretation, there are some ambiguities and caveats for applying the Theory to epistemic norms that go unnoticed. Focusing instead on the Theory's mathematical model allows us to see the difficulties more clearly. This, in turn, allows us to better evaluate the Theory's relevance for epistemology, and the contexts in which is it useful.

My goal is to analyze whether the Theory's mathematical model finds applications for epistemic norms. The paper is divided as follows: Drawing on Ng (2004, chap. 9) and Wiens (2020), I describe the mathematical result behind the General Theory of the Second Best (Section 1). Then, I apply the result to some models in Bayesian epistemology (Section 2) and I provide a second-best analysis of the division of epistemic labor (Section 3).

These results are instructive. First, they allow us to evaluate the claim (made by various philosophers) that second-best problems have import in epistemology. As we will see in Section 2, some possible applications of the Theorem are not straightforward. If we stay at the level of interpretation, we can't see these difficulties clearly. Second, they allow us to see that talk of "approximating an ideal" is ambiguous, and allows for multiple interpretations. Third, they show that a plausible claim concerning the approximation of epistemic ideals is subject to counterexamples. Fourth, they invite us to reconsider the role of idealizations in epistemology, but at the same time, they provide some of the conditions in which epistemic ideals can be relevant to agents like us.

1 | From Economics to Epistemology

The authors first present the mathematical result behind the General Theory of the Second Best. Section 1.1 clarifies what "approximating an ideal" can mean. Sections 1.2 and 1.3 present the General Theory of the Second Best. Section 1.4 summarizes the conditions in which second-best problems can be observed.

1.1 | Some Brief Remarks on Approximation

Approximation claims are ambiguous. That is, approximating an ideal can mean different things. For instance, consider[4]:

1.1.1 | Approximation in Terms of Closeness

Suppose an ideal is defined in terms of the optimal value of some variables $x_1, x_2, \ldots, x_n$. Then, approximating the ideal can refer to getting as close as possible to the optimal values of $x_1, x_2, \ldots, x_n$.

1.1.2 | Approximation in Terms of Value

Suppose an ideal is defined in terms of the greatest amount of value (say, X) one can get out of a given situation. Then, approximating the ideal can refer to getting an amount of value that is as close as possible to X.

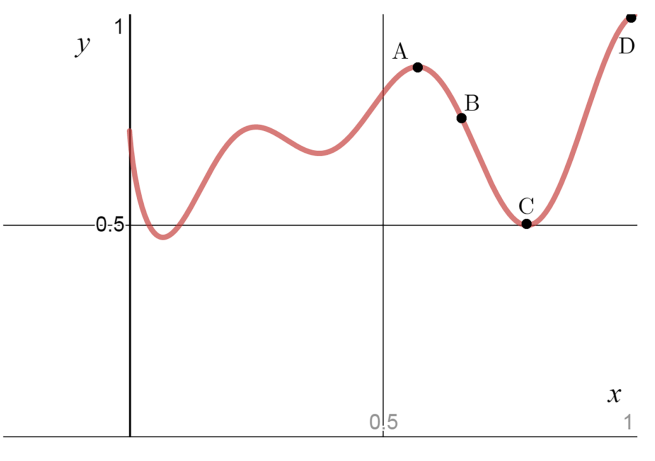
1.1.3 | Approximation of Features

Suppose an ideal is defined in terms of the features or states of affairs of an ideal world. Then, approximating the ideal can refer to meeting as many features of the ideal world as possible.

1.1.4 | Approximation of Relationships

Suppose an ideal is defined in terms of the relationships between variables of an ideal world. Then, approximating the ideal can refer to meeting as many of the relationships that characterize the ideal world.

Some of the above characterizations of approximation can conflict with one other. Take Approximation in Terms of Value and Approximation in Terms of Closeness. Suppose the curve on the following graph represents the amount of value one can get out of a given situation (represented by variable y):



Variable y is a function of variable x.[5] Point D has the greatest value. Suppose one is at point B and has to choose between getting closer to A or getting closer to D. Although point A is not the global optimum, getting closer to A will result in greater value. Getting closer to D will first result in a loss in value (until point C), followed by an increase in value (between C and D). This means that getting closer to an ideal can, at least temporarily, be disvaluable. In fact, suppose one cannot attain the ideal (say, one can only reach values of x that are located between points A and C, which excludes the global optimum D). Then, getting closer to the ideal would be a bad idea—given the range of values one can reach, getting closer to D is disvaluable. In contrast, getting away from the ideal (say, by getting closer to point A) is optimal. Upshot: There are different ways of understanding approximation claims, and some of them conflict with each other. In the above case, Approximation in Terms of Closeness and Approximation in Terms of Value come apart: Getting closer to the ideal value of x does not necessarily result in an increase of value.[6]

My goal in this paper is not to address all the possible interpretations of approximation. Some interpretations might be true, but could be trivial. Others are not trivial, but are obviously false. In the spirit of the Theory of the Second Best, I wish to address a nontrivial, intuitive interpretation of approximation claims. This brings me to my second point. Some understandings of approximation are more plausible (or intuitive) than others. We do not need sophisticated mathematical models to see that Approximation of Features can be suboptimal. Consider the following example: In an *ideal* world, 100 guests attend your party and you have cake for 100 people. However, in *this* world, no one shows up at your party. Should you still make cake for 100 guests? Obviously, that would be a bad idea, even if having cake for 100 guests is a feature of the ideal world.

But perhaps Approximation of Relationships doesn't raise the same issue. Suppose the ideal world is defined in terms of the following relationships between variables:

- For each party organizer, there are 100 guests.
- For each guest, there is one piece of cake.

The ideal world is defined in terms of relationships between three variables—namely, the number of party organizers, the number of guests and the number of pieces of cake. Then, Approximation of Relationships makes sense: Suppose the first ratio cannot be met (e.g., guests do not show up to your party, and so the ratio of organizers to guests cannot be satisfied). Accordingly, we cannot reach the ideal world. Yet satisfying the ideal ratio of pieces of cake to guests is still optimal—that is, it is optimal to have no cake, which satisfies the second ratio. Here, approximating the ideal makes sense. Even if the first ratio cannot be met, we should still meet the second ratio.

At first sight, there is no obvious problem with Approximation of Relationship. And this is why some philosophers find the Theory of the Second Best interesting: It tells us the conditions under which Approximation of Relationships is problematic.[7]

1.2 | A Mathematical Description of First-Best Scenarios

The General Theory of the Second Best is about optimization—that is, about maximizing or minimizing some allocated resources or states of affairs. The Theory is, at its core, a mathematical model. Lipsey and Lancaster simply give it an economic interpretation. I will start by describing the mathematical model and then I will discuss the various interpretations we can give to it.

In its most abstract form, the model put forth by Lipsey and Lancaster consists in mathematical objects, namely:

1. A **Vector** $(x_1, x_2, ..., x_n)$ that denotes some particular assignments of values of n continuous variables (for $n \geq 3$).
2. A **Function** $F(x_1, x_2, ..., x_n)$ that represents a goal (e.g., welfare, utility, justice, etc.).
3. An **Initial Constraint** $G(x_1, x_2, ..., x_n) = 0$ that limits the joint values that variables $x_1, x_2, ..., x_n$ can take.
4. An **Optimization Technique** for identifying which variables $x_1, x_2, ..., x_n$ optimize F while satisfying one or multiple constraints.

For clarity, here is a possible interpretation of the above notions. Suppose you manage three resources, namely, apples, bananas and cherries. These resources contribute to your welfare. You want to maximize your welfare, and so you need to determine how many apples, bananas and cherries you need. Begin with the Vector. There are three variables, namely, apples (a), bananas (b) and cherries (c). The vector (a, b, c) could denote the number of apples, bananas and cherries one has. So, (1, 0.25, 0.5) means that one has 1 apple, 0.25 bananas and 0.5 cherries.

Then, there is the Function. Suppose your objective welfare function can be represented with the following: $(a{\cdot}b)+(b{\cdot}c)+(c{\cdot}a)$. In other words, the number of apples times the number of bananas contributes to your welfare, and the number of bananas times the number of cherries also contributes to your welfare, and so forth. Formally:

$$F(a, b, c) = (a \cdot b) + (b \cdot c) + (c \cdot a) \tag{1}$$

The Initial Constraint can be many things. Perhaps you cannot have an infinite amount of fruit (there is only so much you can buy). Perhaps the banana vendor accepts payment only in cherries (so if you want bananas, you need to give up some cherries). For our purposes, assume our constraint is

$$G(a, b, c) = a + \left(b^2 \cdot c^2\right) + \frac{b^2 + c^2}{4} - 1 = 0 \tag{2}$$

We assume that any allocation of resources has to meet this constraint.

Recall that this is an optimization problem. We want to identify an allocation of resources that optimizes welfare (relative to some constraints). For such a type of problem, we use the method of Lagrange multipliers. It is used for finding a function's optima subject to some constraints, which is precisely what we are trying to achieve.

This method allows us to express our first-best scenario in terms of ratios of derivatives (for simplicity, I'll also call them ratios of variables). The ratios are given by the following:

$$\frac{F'_i}{F'_n} = \frac{G'_i}{G'_n} \qquad i = 1, 2, \ldots, n-1$$

where F'_i denotes the derivative of F with respect to variable x_i and G'_i denotes the derivative of G with respect to variable x_i.

Given the functional forms of F and G in our fruit example, we would get the following ratios[8]:

$$\frac{b+c}{a+b} = \frac{1}{2c \cdot b^2 + 0.5c} \tag{3}$$

$$\frac{a+c}{a+b} = \frac{2b \cdot c^2 + 0.5b}{2c \cdot b^2 + 0.5c} \tag{4}$$

Equations (3) and (4) state the optimal ratios of variables a, b, and c that any local or global optimum satisfies. If an allocation of resources is optimal, it will satisfy such ratios (but satisfying such ratios does not necessarily entail that an allocation of resources is optimal). As it happens, F reaches its highest point (≈1.218) when our variables take the following values:

$$(a, b, c) \approx (0.577, 0.668, 0.668) \tag{5.1}$$

$$F(0.577, 0.668, 0.668) = 0.577 \cdot 0.668 + 0.668 \cdot 0.668 + 0.668 \cdot 0.577 \approx 1.218 \tag{5.2}$$

Naturally, such values of a, b, and c satisfy the ratios in (3) and (4).[9]

1.3 | Conditions in Which the Result Applies

A first-best scenario is characterized in terms of ratios of variables. In non-ideal worlds, some of the ideal ratios cannot be satisfied. We want to know when it will be optimal (or suboptimal) to approximate the remaining ratios. The Theory of the Second Best says the remaining ratios of derivatives will be suboptimal if and only if some conditions C are not satisfied.

Conditions C have to do with the system of partial derivatives that result from optimization in second-best scenarios (see step (A9) in the Appendix for details). For present purposes, there is no need to be that specific. With respect to the project of this paper, all we need is to identify some *good signs* that the result applies, such as:

1. Some of the function's variables have more effect when combined. For instance, suppose you want to get to work. Gas won't get you to work. A car with an empty gas tank won't get you to work. But a car with gas in it will get you to work. So, cars and gas are effective (or have more effect) when combined.
2. Some of the function's variables are substitutes. For instance, suppose you want to get to work. Since you possess car A, you don't need car B. However, when car A is damaged, you might need car B. So, car A and car B are substitutes.
3. Some of the function's variables can cancel or attenuate each other out. For instance, suppose you want to be happy. Driving your car will make you happy, and drinking alcohol will also make you happy. However, driving your car while drinking alcohol will make you worse off.

In other words, there are conditions in which the mathematical theorem underlying the Theory will be "triggered" (e.g., will lead to second-best problems). A good sign that functions $F(x_1, x_2, \ldots, x_n)$ and $G(x_1, x_2, \ldots, x_n)$ will trigger the theorem is when, to optimize F under constraint G, you cannot optimize the value of the variables $x_1, x_2, \ldots, x_n$ independently of each other. Optimizing F under constraint G requires taking the value of all the variables $x_1, x_2, \ldots, x_n$ into account simultaneously. Call this the nonseparability constraint.

Again, consider our fruit example. As discussed in the previous section, some ratios of variables tell us which values of a, b, and c optimize $F(a, b, c)$ under constraint $G(a, b, c)$. Suppose we add a constraint preventing us from attaining the ideal world. Say, suppose the following holds in a non-ideal world:

$$\frac{b+c}{a+b} = \frac{2}{2c \cdot b^2 + 0.5c} \tag{6}$$

If (6) is true, the first-best ratio stated in (3) cannot be satisfied. Should we still satisfy the remaining first-best ratio—namely, the one stated in (4)? This amounts to a new optimization problem, which takes both the initial constraint (2) and the additional constraint (6) into account.

Under such constraints, the function's highest point is ≈1.108, which is less than the first-best optimum of ≈1.218. This second-best optimum is observed when variables a, b and c take the following values:

$$(a, b, c) \approx (0.328, 0.626, 0.946) \tag{7.1}$$

$$F(0.328, 0.626, 0.946) \approx 0.328 \cdot 0.626 + 0.626 \cdot 0.946 + 0.946 \cdot 0.328 \approx 1.108 \tag{7.2}$$

That is, (7.1) expresses the values of a, b and c that optimize F under constraints (2) and (6). However, (7.1) violates the first-best ratio stated in (4). In accordance with the General Theory of the Second Best, if we want to optimize $F(a, b, c)$ in non-ideal circumstances, we sometimes have to violate some first-best ratios of variables. Or, approximating ideal ratios of variables in non-ideal circumstances can be suboptimal.

1.4 | Beyond Economics: Second-Best Results for Everyone

As Wiens (2020, sec. 3) notes, the mathematical model described in the previous section can take any interpretation we want. So while Lipsey and Lancaster gave it an economic interpretation, nothing prevents us from giving it an epistemological one. All we need to do is find problems in which the following four conditions are met:

1. We want to maximize a function *F*.
2. *F* is subject to an initial constraint *G*.
3. *F* and *G* are functions of at least three continuous variables.
4. Some separability conditions are violated in either *F* or *G*.

We don't even need to find particular functional forms that match the above description—that is, we can run the argument without knowing what functions *F* and *G* would look like exactly. If *F* and *G* satisfy the above criteria, they will trigger the mathematical theorem underlying the Theory (Wiens 2020, sec. 3.1).[10]

These four conditions also allow us to make an initial evaluation of the claim that there are second-best problems in epistemology. Specifically, these conditions allow us to determine whether the problems described by epistemologists fit the Theory's formal model. In the Theory of the Second Best, ideals have to do with the constrained optimization of functions (this is the idea behind conditions 1 and 2). But in the work of Egan (2008), Titelbaum (2015), and DiPaolo (2018), epistemic ideals are not defined in terms of the general conditions that optimize a function under certain constraints. For example, Egan is concerned with problems in which the epistemic ideal is characterized by two *features*: (i) agents have maximally reliable belief formation mechanisms and (ii) agents incorporate their information into a unified corpus of beliefs (Egan 2008, 62). This is closer to the conception of the ideal found in *Approximation of Features*. However, Staffel's (2019) theory of approximation seems to meet all four conditions. We will return to this point in the next section.

In the remainder of this paper, I provide two epistemic interpretations of the General Theory of the Second Best.

2 | First Application: Bayesian Epistemology (Or Something Close Enough)

Second-Best Epistemology can be applied to Bayesian norms … indirectly. Let me explain what I mean by that.

Bayesian epistemology typically revolves around some *requirements*. For instance, agents should satisfy the requirement of Probabilism, which roughly states that an agent's credence assignments should satisfy the axioms of probability[11]; they should satisfy some requirements governing probabilistic inference, such as Conditionalization[12]; and they should satisfy some omniscience requirements for a priori or logical truths.[13] However, the classical interpretation of such requirements is *discrete*: Either you satisfy them or you don't. That is, if we take these requirements to be discrete, there is no continuous degree to which one satisfies such requirements. Recall that Second-Best Epistemology works with functions of *continuous* variables, not discrete ones. If a function's variables are discrete, we can't analyze such a function with Second-Best Epistemology. This complicates the application of Second-Best Epistemology to Bayesian requirements—that is, we can't treat discrete Bayesian requirements as variables of a function we wish to optimize.

Another problem for applying second-best problems to Bayesian epistemology is that, typically, this theory of rationality does not make room for initial constraints. To have second-best problems, the function we want to optimize needs to be subject to an initial constraint (this is an essential ingredient for defining the ideal). However, most Bayesian epistemologists don't like constraints. For instance, they often assume that agents are logically omniscient, have infinite memory, infinite processing capacities, can take any credence they want, and so forth. If our agents have all these capabilities, what kind of initial constraint could they face?

Given these initial worries, we can doubt that second-best results will find applications for Bayesian ideals (and this is one of the things we can learn from this paper: Some problems are not a good fit for the second-best framework!). We don't want to be the guy with a hammer desperately looking for a nail.

However, I think there are some theories in the vicinity of Bayesian epistemology that generate second-best problems. Staffel's (2019) framework for degrees of Bayesian rationality is a good point of departure. Her model allows us to analyze the satisfaction of Bayesian requirements indirectly. Also, it is compatible with optimizing a function under initial constraints.[14] So, her model is a good starting point to figure out how we can apply second-best results to Bayesian epistemology.

Here is a (too) brief summary of Staffel's framework. Suppose that Anna has a credence of 0.49 in *P*. However, she knows that *P*'s objective probability is 0.5. She violates the Principal Principle, a Bayesian requirement of rationality. Bob also violates the Principal Principle. He also knows that *P*'s objective probability is 0.5, but he has a credence of 0.1 in *P*. If all we care about is whether agents *fully* satisfy the requirements of rationality, then our judgments about Anna and Bob should not be different. After all, they both violate the Principal Principle. However, it seems that Anna is closer to the perfect satisfaction of the Principal Principle than Bob. How can we account for this?

For Staffel, we can figure out a distance measure between an agent's actual credences and the ones that fully satisfy the Bayesian requirements of rationality. There are different ways to measure this (see Staffel 2019, chap. 3). Following Staffel, we will measure degrees of closeness with the Euclidean distance.[15] Relative to their knowledge, we can measure Anna's and Bob's closeness to the perfect satisfaction of the Principal Principle with the Euclidean distance as follows:

- Distance = (actual credence in *P*—the credence in *P* recommended by the Principal Principle)^2
  - Distance for Anna: $(0.49–0.5)^2 = 0.0001$
  - Distance for Bob: $(0.1–0.5)^2 = 0.16$

Here, a higher value means that the agent is getting farther away from the perfect satisfaction of the Principal Principle.

So, in Staffel's model, Anna is closer to the perfect satisfaction of the Bayesian requirements of rationality than Bob (since $0.0001 < 0.16$). For Staffel, degrees of closeness to the perfect satisfaction of Bayesian requirements matter. She gives practical and epistemic arguments of why approximating some Bayesian requirements of rationality is beneficial (see Staffel 2019, 56–94). In other words, if we want to reap the benefits of epistemic rationality, we should care about such degrees of closeness to Bayesian ideals.

We will assume that F is a function measuring the distance between an agent's actual credences and the ones that perfectly satisfy the Bayesian requirements. In accordance with Staffel, we will also assume that agents should care about minimizing the value of this function. So, we have a function F we wish to optimize. In Staffel's framework, F can be subject to initial and additional constraints. For instance, perhaps the reason why Anna and Bob have imperfect credences in P is that they face some constraints.

This brings us one step closer to home. We have all the essential ingredients for applying second-best results to Staffel's model. We have a function of continuous variables (i.e., credences). We want to optimize it. This function can be subject to an initial constraint, as well as additional ones.

The question, then, is what second-best problems could look like in this framework. Here is a possible interpretation of Staffel's model illustrating this. Suppose that Anna only has credences in the following three propositions: A, B, and ~B. She knows that A and B are probabilistically independent propositions, and that A&B's objective probability is 0.5. In accordance with the Principal Principle, her perfect credence in A&B would be 0.5. In accordance with Probabilism, her credences in B and in ~B should sum up to one. We will assume that, for Anna, Probabilism and the Principal Principle have the same "weight": Getting closer to the credences warranted by Probabilism has the same importance as getting closer to the credences recommended by the Principal Principle.[16] If we limit ourselves to these requirements and her credences in these propositions, the function F measuring the distance between her actual credences and the ones that perfectly satisfy the Bayesian requirements could be this:

$$F(a,b,c) = ((a \cdot b) - 0.5)^2 + ((b + c) - 1)^2 \tag{8}$$

Here, a denotes Anna's credence in A, b denotes Anna's credence in B, and c denotes Anna's credence in ~B.

Two quick remarks on (8). First, the distance measure for Probabilism is $((b+c)-1)^2$. So, relative to Probabilism, agents who have a credence of X in B and a credence of $(1–X)$ in ~B will get a perfect score, which is exactly what we want. Also, this distance measure satisfies some desiderata mentioned by Staffel.[17] However, Staffel discusses other distance measures for Probabilism (see, e.g., Staffel 2019, 77–82). These other distance measures are appropriate when incoherent agents are trying to identify the coherent credence function that is closest to their own credences (Staffel 2019). Here, we are not trying to identify the best path to coherence that incoherent agents can take. This is a separate issue. So, I leave aside distance measures for identifying the best path to coherence. Second, the distance measure for the Principal Principle is $((a \cdot b) - 0.5)^2$. As you can see, I assume that Anna's credences in A and her credence in B determine her credence in A&B. Some might deny this, and suggest instead that there is an additional Bayesian requirement constraining the relationship between her credence in A, her credence in B, and her credence in A&B. We could add such a Bayesian requirement to our function F and revise (8) accordingly.[18] However, the small simplification I made doesn't affect the results.

So, we have a function F we wish to optimize. F could be subject to an initial constraint G. Recall that Staffel's model allows for constrained optimization, including constraints on the combinations of credences one can take (Staffel 2019, 112–116). So, an initial constraint could be part of the explanation why Anna's second-best credences do not fully satisfy the Bayesian requirements.

Our initial constraint G will concern the combinations of credences one can have.[19] For instance, perhaps Anna is epistemically conservative, and can't have "opinionated" credences in many propositions. So, if she is opinionated with regard to proposition A, she has to be less opinionated with regard to propositions B and C. Or perhaps Anna is subject to certain cognitive biases, with the result that her degree of confidence in A is always equal to her degree of confidence in B (her biases lead her to think that A and B are coextensive). Here, readers are free to fill in the details. What matters is that there is a constraint on the joint values of one's credences in A, B, and C. For present purposes, we'll assume that Anna could face the following initial constraint:

$$G(a,b,c) = (a \cdot b) + (b \cdot c) + (c \cdot a) - 1 = 0 \tag{9}$$

Given F and G, there will be some first-best ratio of variables for Anna. Using the method of Lagrange multipliers, we can then express our first-best scenarios. In the above case, we get the following (simplified) equations:

$$b \cdot (a + b) \cdot (a \cdot b - 0.5) = (b + c) \cdot (b + c - 1) \tag{10}$$

$$(a + b) \cdot (a \cdot (a \cdot b - 0.5) + (b + c - 1)) = (a + c) \cdot (b + c - 1) \tag{11}$$

As it happens, our function F reaches its lowest point (0) when Anna's credences are as follows:

$$(a,b,c) \approx (0.772, 0.648, 0.352) \tag{12.1}$$

$$F(0.772, 0.648, 0.352) = 0 \tag{12.2}$$

These values of a, b, and c satisfy the ratios in (10) and (11). And they are compatible with the full satisfaction of Bayesian requirements of rationality.

Then, suppose a constraint on one of the equations prevents Anna from attaining the ideal world. For instance, suppose the following holds in a non-ideal case, and prevents Anna from satisfying (10):

$$b \cdot (a + b) \cdot (a \cdot b - 0.5) = 0.2 + (b + c) \cdot (b + c - 1) \tag{13}$$

If (13) is a constraint on Anna's credences, the first-best ratio stated in (10) cannot be satisfied. We could nevertheless satisfy the remaining first-best ratio stated in (11). Would that be optimal?

If we take the constraint stated in (13) into account, the function's lowest point is ≈0.0198. This is higher than the first-best optimum of 0. This second-best optimum is observed when agents have the following credences in a, b, and c:

$$(a, b, c) \approx (0.892, 0.701, 0.235) \tag{14.1}$$

$$F(0.892, 0.701, 0.235) \approx 0.0198 \tag{14.2}$$

However, such credence assignments do not satisfy the first-best ratio stated in (11). This means that Anna is better off not satisfying the remaining ratio. Thus, if a constraint prevents Anna from satisfying a first-best ratio governing her credence assignments, it could very well be optimal for her not to satisfy the remaining ratios. This is the Theory's main conclusion.

So, there is an indirect way to argue that Bayesian epistemology can be subject to second-best problems. In some interpretations of Staffel's model, imperfect agents who aim at optimizing their degree of Bayesian rationality can violate some first-best ratios of variables.[20] That is, in non-ideal circumstances, agents can entertain credences that are incompatible with first-best ratios of credence assignments. As I said, this is an indirect application of the Theory to Bayesian epistemology. Some "classical" Bayesian frameworks are not a good fit for the Theory of the Second-Best.

3 | Second Application: Epistemic Communities and the Division of Epistemic Labor

In the previous section, I have analyzed some specific functions F and G to see whether they would lead to second-best problems. However, as I have indicated in Section 1.4, the Theory can be relevant *even without knowing the particular functional forms of F and G*. All we need to do is find problems in which four general conditions are met.

I now wish to discuss a case in which the particular functional forms of F and G are unknown. Why? Because we often lack a clear idea of the functions we wish to maximize. Since the Theory can apply to cases in which the particular functional forms of F and G are unknown, it finds more applications.

Suppose you are a "philosopher-monarch" in charge of designing the perfect epistemic community (such as a scientific research community). In other words, suppose you could design any epistemic community you want. What should such a community look like? Should it be diversified, in the sense that different researchers would work with different methods? Should researchers defer to each other? Should they criticize each other?

This is the problem of the division of epistemic labor. Many philosophers have worked on this issue. For present purposes, I will focus on Philip Kitcher's work on ideal epistemic communities and the optimal division of epistemic labor. But since many philosophers have defended claims akin to Kitcher's, the point made in this section should generalize well to other frameworks.[21]

Kitcher (1990) begins by assuming that epistemic communities should aim at getting significant truths and avoiding error. Then, he goes on to argue that research on distinct incompatible theories serves this goal, even if some of these theories are less plausible than others. Specifically, he has argued that pursuing distinct incompatible research programmes is truth-conducive, even when some of them are less plausible than others. Considering the case in which theory X is weakly supported by the evidence (say, the evidential probability that theory X is true is 0.51, and the evidential probability that competing theory Y is true is 0.49), Kitcher says:

> You would (rightly) have dismissed [uniform scientific opinions in favour of theory X] ... as a bad bargain. With the evidential balance between the two theories so delicate, you would have preferred that some scientists were not quite so clear-headed in perceiving the merits of the theories, so that the time of uniform decision was postponed. (Kitcher 1990, 5–6)

His argument relies on decision theory and roughly goes as follows. Suppose that available scientific methods are denoted by X_1, X_2, ..., X_n. A method's ability to solve a given problem is a function of the amount of resources devoted to solving the problem. For instance, the more scientists work on a problem P with sophisticated equipment using a scientific method X_i, the more method X_i is likely to solve problem P. Should scientists take the most plausible or successful method, or should they split their efforts between distinct incompatible methods? This amounts to a familiar maximization problem. For the sake of simplicity, assume there are two competing methods available to researchers. If there are N resources available to solve a given problem (e.g., the number of scientists and research assistants available to work on problem P, their research grants and laboratories, and so forth), we want to maximize the following:

$$\textbf{Maximize:} \quad \Pr\left(X_1 \text{ succeeds} \mid n \text{ resources are allocated to } X_1\right) + \Pr\left(X_2 \text{ succeeds} \mid N - n \text{ resources are allocated to } X_2\right) - \Pr(\text{both methods succeed})$$

Kitcher then argues that, given some plausible assumptions concerning $\Pr_{X1}$ and $\Pr_{X2}$, the above can be maximized by pursuing distinct incompatible research programmes (e.g., by assigning some scientific resources to distinct incompatible methods). This result holds even if one of the methods is less likely to succeed than the other.

As we can see, Kitcher treats the division of epistemic labor as a maximization problem. In later work (Kitcher 1995, chap. 8), his model takes more factors into account. Here is the gist of his later model. Scientists face various decision problems in the course of their enquiries. Scientists are trying to solve problems, and they invest time and resources to do so. They can *borrow* some results from others (instead of trying to reach results on their own). They can also put their own research on hold and contribute to the *critical scrutiny* of others (e.g., review papers, write response papers, try to replicate studies, etc.). Each action has some value, but also some costs. A scientist's goal is to make the best use of his or her resources.

To fully optimize the division of epistemic labor, there could be even more variables to take into account. For instance, factors like the search strategies that are deployed (e.g., strategies for finding a solution to a problem) are relevant too.[22] Communication structures in scientific communities could also be taken into account.[23] But Kitcher's work gives us a *partial sketch* of an ideal epistemic community's maximization function. Following Kitcher, we want to maximize the likelihood that a given epistemic community will reach significant truths (and avoid error) on various problems. This is our function F. Getting significant truths will be a function of at least three factors:

Diversity (*d*). What is the degree to which distinct methods are represented in the scientific community? Perfect diversity would entail that all methods are equally represented by some scientists, while perfect homogeneity would entail that everyone is working with the same method.

Testimony (*t*). What is the degree to which scientists defer to each other? For instance, scientists can form deference networks, where some scientists act as testifiers and others listen to them (by reading their papers, attending their conferences, and so forth). What is the degree to which scientists are active members of such networks?

Critical Scrutiny (*c*). What is the degree to which scientists are critical of each other? For instance, do scientists spend a lot time criticizing each other's ideas, methods and conclusions (by trying to rebut or refute each other, or by trying to find counterexamples to each other's ideas, etc.)?

Each of the above factor increases the truth-conduciveness of science. If scientific communities had infinite resources, they would promote perfect diversity, testimony and critical scrutiny. But scientific communities do not have infinite resources and so they have to make choices. Given the amount of resources available to an epistemic community, it might be impossible to achieve perfect Diversity, perfect Testimony and perfect Critical Scrutiny simultaneously. So, there can be constraints on the joint satisfaction of the above variables. Allocating more resources to the formation of deference networks might require allocating less resources to diversification, or allocating less time for criticizing each other's ideas (and vice versa).

Think of a research agency that can finance the following projects: (i) a conference where scientists will meet and discuss with each other (which improves Testimony), (ii) a new research method that is incompatible with all known previous methods (which improves Diversity) or (iii) a costly experiment for confirming (or disproving) a given theory (which improves Critical Scrutiny). If the funding agency has limited resources, it might have to prioritize some projects over others. This, in turn, means that the agency has to prioritize some variables over others.

Again, we already have most of the ingredients to observe second-best problems. We have a function that we want to optimize, namely, a function of the optimal division of epistemic labor in epistemic communities. Furthermore, such a function depends on at least three variables: Diversity (d), Testimony (t), and Critical Scrutiny (c). We can then characterize a first-best optimum in terms of ratios of variables. Finally, we have an initial constraint G.

Nonseparability is the only missing ingredient. Recall that a good sign that variables are nonseparable is when all variables have to be taken into account simultaneously when optimizing F (under constraint G). The variables cannot be optimized independently of each other. Is this the case here?

Plausibly, there can be cross-effects between Critical Scrutiny and Testimony in F. Members of an epistemic community cannot criticize each other's ideas if they do not read each other, or attend each other's talks, and so forth. In other words, if one wishes to criticize others, one has to know what others think. Critical Scrutiny makes sense only when combined with Testimony, and so there is a clear cross-effect between these variables.

Critical Scrutiny and Diversity also display some cross-effects in F. In homogeneous epistemic communities, Critical Scrutiny will be a lot less efficient. One way in which scientists can fruitfully criticize each other's ideas is to analyze a given problem from distinct perspectives. In diversified epistemic communities, it is easier for agents to come up with interesting objections and counterexamples to each other's ideas, since not everyone has the same background knowledge and methods for solving problems. So, there seems to be a synergy between Critical Scrutiny and Diversity.

Finally, there can be cross-effects between Diversity and Testimony in F. As argued by many philosophers, diversification within epistemic communities makes deference more complicated. Yet deference is an essential component of the practice of testifying to each other. According to Dogramaci and Horowitz:

> When rational reasoners thus coordinate upon the same reliable belief-forming rules, this has the result that rational reasoners are able to serve as each other's epistemic surrogates .... [but if rational reasoners use alternative belief-forming rules, rules that yield distinct views given the same evidence], the enforcement of rational rules of reasoning does not make it safe to trust the testimony of rational reasoners. (Dogramaci and Horowitz 2016, 136–137, 139)

Similarly, Greco and Hedden argue:

> If two agents have the same total evidence but different beliefs about whether *P*, then you cannot defer to each's belief about whether *P* on pain of

> inconsistency .... Suppose you know that one agent has credence n in P while another has credence m in P, and that they have the same total evidence. Then ... you cannot simultaneously adopt credence n in P and credence m in P. (Greco and Hedden 2016, 373)

Roughly, their point is that if members of an epistemic community entertain methods of reasoning that are compatible with each other, deferring to each other is easier.[24] Accordingly, increases in Diversity can affect the value one gets out of Testimony (by making it more complicated). Designing the ideal epistemic community requires taking the interaction between Diversity and Testimony into account.

There can be cross-effects between variables d, t and c. This means that the division of cognitive labor can lead to second-best problems. For instance, the ideal epistemic community can be determined by some first-best ratios of variables. But if a constraint prevents us from satisfying one of the ratios, satisfying the remaining one might be suboptimal.

4 | Discussion

The General Theory of the Second Best says that, in non-ideal circumstances, approximating ideals (understood in terms of ratios of variables) might be suboptimal. The Theory provides counterexamples to a specific interpretation of what "approximating an ideal" means, namely, Approximation of Relationships. I have argued that the Theory can find applications in epistemology. First, some indirect interpretations of Bayesian epistemology can lead to second-best problems. Second, the division of epistemic labor can be subject to second-best problems.

These applications matter. First, many epistemologists think that the Theory's *conclusions* apply in epistemology. I went a step further and tried to determine how the Theory's formal model could apply in epistemology. We saw, for instance, that the Theory's application to Bayesian epistemology is not straightforward. If we stayed at the level of interpretation (e.g., if we didn't try applying the formal model to epistemic norms), we would not have seen that.

As I have said in the introduction, in the past few years, some philosophers have claimed that there are second-best problems in epistemology. However, it was unclear whether the *formal model* underlying the Theory could find applications in epistemology. My goal in this paper was to explore this possibility. As I said in Section 2, some classical Bayesian frameworks are not a good fit for the Theory, and this is why I explored some indirect ways to apply the theory to a Bayesian framework. If you think that these applications are uninteresting, then you have learned something about the putative significance of Second-Best Epistemology. You are now in a better position to evaluate whether epistemologists should have an interest for second-best results. Perhaps Second-Best Epistemology works well for the division of epistemic labor but is not a good fit for some standards interpretations of Bayesian epistemology. I gave you a clearer idea of the concept's possible applications.

Second, these results allowed us to see that "approximating an ideal" can mean many different things. There is no unique or canonical way to express ideals and their approximation. There are at least four different interpretations of what this means, and they are not coextensive. Some, like Approximation of Features, are fairly common among philosophers. Others, like Approximation of Relationships, are more technical and mirror some models found in economics. Without applying the formal model to epistemic norms, we don't see this different way in which we can refer to an ideal and its approximation.

For instance, DiPaolo's second-best analysis concerns the conflict between primary norms of perfection and secondary norms of fallibility (DiPaolo 2018). He writes:

> The epistemic theory of the second best says that when some conditions necessary for perfect belief management aren't met, meeting others may no longer be desirable. Full coherence, perfectly proportioning belief to the evidence, steadfastness, and certain kinds of dogmatism might all be conditions necessary for perfect belief management. But when agents aren't fully coherent ... steadfastness and dogmatism may no longer be desirable. (DiPaolo 2018, 22)

DiPaolo's analysis is faithful to the Theory's *main conclusion*. However, his account of epistemic ideals is very close to Approximation of Features, since it focuses on the *characteristics* of ideally rational agents (e.g., full coherence, perfect evidence-responsiveness, etc.). Approximation of Features doesn't track the mathematical model underlying the Theory. Rather, the Theory is closer to Approximation of Relationships. Without concrete applications of the Theory to epistemic norms, we can't see the different sense in which economists refer to an ideal and its approximation.

Third, these applications show that an intuitive interpretation of what "approximating an epistemic ideal" means is, in fact, subject to counterexamples. We can criticize the recourse to epistemic ideals by engaging with trivial or obviously false interpretations of approximation claims (such as Approximation of Value or Approximation of Features). However, one reason why some philosophers might find Second-Best Epistemology appealing is that it argues against an intuitive thesis concerning the approximation of epistemic ideals.

Finally, these applications invite us to reconsider the relevance of ideals in epistemology. Many contemporary epistemological theories admit a high degree of idealization. We will never come close to being or meeting these ideals. So, why care about them?

Second-Best Epistemology brings good and bad news to those who will reconsider the relevance of idealized epistemic theories. Let's begin with the bad news. We cannot simply assume that, when we cannot meet epistemic ideals, we should approximate them. Second-Best Epistemology is in direct conflict with such an assumption. But there is hope! Second-Best

Epistemology also describes the conditions in which approximation is warranted (e.g., the conditions in which approximating an ideal is optimal). As I indicated in Section 1.3, second-best problems appear when some separability conditions are not satisfied (see also step (A9) in the Appendix). So, if one can argue that some separability conditions are satisfied in a function F and the constraint G it is subject to, one can argue that such functions will not be subject to second-best problems. In other words, in cases where functions are separable, we can use the Theory's main result in favor of the approximation approach. So, approximation does not *always* work, but it might work *in some specific contexts*.

Hence, Second-Best Epistemology can be seen both as a warning and as a source of progress. It is a warning, since it predicts that, under some conditions, approximating an ideal is suboptimal. But it is also a source of progress, since it specifies the conditions under which approximating an ideal is optimal (at least, with respect to a specific interpretation of what ideals and their approximation are).

Acknowledgments

Thanks to Catrin Campbell-Moore, Davide Fassio, Jie Gao, Sophie Horowitz, Liz Jackson, Pavel Janda, Jason Konek, Daniel Laurier, Benjamin Levinstein, David Montminy, Richard Pettigrew, Louis-Xavier Proulx, David Rocheleau-Houle, Ely Mermans, Xander Selene, Julia Staffel, David Waszek and David Wiens for helpful comments on many versions of this project. This research was supported by the Fonds de recherche du Québec—Société et culture (grant #268137) and the Social Sciences and Humanities Research Council (grant #756-2019-0133).

Conflicts of Interest

The author declares no conflicts of interest.

Endnotes

[1] See Wheeler (2018) on ideal and bounded rationality. See Christensen (2004, chap. 6), De Bona and Staffel (2018), DiPaolo (2018), Earman (1992, 56), Smithies (2015), Staffel (2017, 2024), Talbott (2016, Sec. 6.1.A) and Zynda (1996) on the recourse to epistemic ideals. In general, it is unclear what counts as an "ideal." See Gaus 2016, Chaps. 1–2 and Estlund 2014 on this issue. Still, most philosophers agree upon some *examples* of ideal and non-ideal epistemic theories. For example, Bayesian epistemology (Joyce 1998; Pettigrew 2016) is often thought of as an idealized theory, whereas theories of bounded cognition and heuristics (Gigerenzer and Selten 2002; Kahneman 2003; Todd and Gigerenzer 2000) are often thought of as non-ideal theories. I'll briefly come back to this issue in Section 1.1.

[2] See, among others, De Bona and Staffel (2018), DiPaolo (2018), Griffiths, Lieder, and Goodman (2015), Halpern and Pass (2015) Icard (2018), Lorkowski and Kreinovich (2018), Morton (2012, chap. 1), Paul and Quiggin (2018), Staffel (2017, 2019, 2024), and Skipper and Bjerring (2020).

[3] However, she mentions in a footnote that the second-best problems she refers to might be "superficially similar" to the kind of results associated with the Theory of the Second Best (Staffel 2019, 112). More on this point in Section 2.

[4] I make the same distinctions between various interpretations of approximation claims in Daoust (2021b).

[5] The variables could be anything. For instance, x could be the amount of time spent on learning philosophy, and y could be the amount of pleasure one gets out of it.

[6] See Gaus (2016, chap. 1) on a similar point.

[7] Wiens (2020, note 5) makes a similar observation.

[8] The ratios need to be different. To see why, consider an example found in Wiens (2020, §3.1). Suppose we want to maximize the following function F subject to a constraint G:

$$F(e, f, s) = e \cdot f \cdot s$$

$$G(e, f, s) = -1 + f + e^2 \cdot s^2$$

Accordingly, the first-best ratios of variables can be defined as follows:

$$\frac{f}{s} = 2se^2, \frac{f}{e} = 2es^2$$

These two ratios are identical. Surely, if one constraint prevents us from satisfying the first ratio, we also have to depart from the second ratio. But in order to explain this, we don't need the Theory of the Second Best.

[9] See Daoust (2021b, 893 fn16) on why do F and G need to be functions of at least three variables.

[10] Naturally, finding the particular functional forms of F and G would be relevant for other purposes. For instance, this would be relevant for determining the extent to which second-best problems are severe. But this is a task for another day.

[11] Joyce (1998), Leitgeb and Pettigrew (2010a, 2010b), Pettigrew (2016).

[12] Greaves and Wallace (2006), Meacham (2015, 2016), Schoenfield (2017).

[13] Easwaran (2011), Dogramaci (2018).

[14] See Staffel (2019, 112–115).

[15] See Staffel (2019, chap. 5).

[16] This is one interpretation of Staffel's model. There are other possibilities. For instance, we could give each requirement a different weight in F. See Staffel (2019, 103–106) on bundle and piecemeal strategies.

[17] See Staffel (2019, 35–39) on some basic desiderata for distance measures.

[18] That is, they might prefer the following version of F:

$$F(a, b, c, d) = ((a \cdot b) - d)^2 + (d - 0.5)^2 + ((b + c) - 1)^2$$

where d is Anna's credence in A&B, and is governed by a requirement that says that Anna's credence in A&B must be equal to her credence in A times her credence in B (since she knows that these propositions are probabilistically independent of each other). But as I say, we would reach a similar conclusion with this slightly more complicated version of F.

[19] A reviewer makes the following observation: In the literature on bounded rationality, constraints usually concern the mental resources available to agents. For example, agents like us have limited cognitive capacities, and can only make a limited number of inferences. These common constraints are difficult to represent in the Theory's framework. Recall that, in the Theory of the Second Best, constraints govern the joint values of variables in F. Thus, if F is a function of the agent's credences, G is a constraint on the values that the credences can take (and not a constraint on the mental operations that the agent can perform). So, there seems to be an "applicability problem" here. The Theory doesn't seem to apply to many problems discussed in the literature on bounded rationality. Response: This brings us back to the issue of applicability for the Theory of the Second Best. If you are interested in problems of optimization of credences under constraints concerning the mental resources agents have, then the Theory might not be what you are looking for. And one of the merits of this paper is that you are now in a better position to figure out the conditions under which the Theory will be useful to you or not.

[20] Compare with one of Staffel's conclusions: She says that "a surprising upshot of my proposal is that thinkers who are incapable of complying with one principle of rationality must also violate other principles of rationality in order to minimize their irrationality" (Staffel 2019, 95–96). Here we reach a similar result—namely, that thinkers who are incapable of complying with one ratio of variables defining the Bayesian ideal should sometimes violate the other ones. We have generalized Staffel's observation, but under a different account of Bayesian ideals and their approximation.

[21] See, e.g., Longino on diversity and criticism in science (Longino 1990, chap. 4).

[22] See Weitzman (1979), Vishwanath (1992) and Adam (2001).

[23] See Zollman (2007, 2013), Rosenstock, Bruner, and O'Connor (2017), O'Connor and Bruner (2019) and Weatherall and O'Connor (m.s.).

[24] Dogramaci and Horowitz (2016) and Greco and Hedden (2016) go a step further and argue that, if rational agents in an epistemic community have the same evidence, they should have the same methods of reasoning. I doubt that their argument proves this (Daoust 2017, 2022). For present purposes, we can grant them that homogeneity *simplifies* deference within epistemic communities.

[25] The results presented here are adapted from Lipsey and Lancaster (1956). I have tried to make the notation as explicit as possible.

[26] See also Ng (2004, 195–196).

[27] Additive separability is sufficient but not necessary for preventing second-best problems.

References

Adam, K. 2001. "Learning While Searching for the Best Alternative." *Journal of Economic Theory* 101, no. 1: 252–280.

Blackorby, C., R. Davidson, and W. Schworm. 1991. "The Validity of Piecemeal Second-best Policy." *Journal of Public Economics* 46, no. 3: 267–290. https://doi.org/10.1016/0047-2727(91)90008-p.

Christensen, D. 2004. *Putting Logic in Its Place*. Oxford: Oxford University Press.

Daoust, M.-K. 2017. "Epistemic Uniqueness and the Practical Relevance of Epistemic Practices." *Philosophia* 45, no. 4: 1721–1733.

Daoust, M.-K. 2021a. "Should Agents be Immodest?" *Analytic Philosophy* 62, no. 3: 235–251.

Daoust, M.-K. 2021b. "Adversariality and Ideal Argumentation: A Second-Best Perspective." *Topoi* 40, no. 5: 887–898.

Daoust, M.-K. 2022. "Optimizing Individual and Collective Reliability: A Puzzle." *Social Epistemology* 36, no. 4: 516–531.

De Bona, G., and J. Staffel. 2018. "Why Be (Approximately) Coherent?" *Analysis* 78: 405–415. https://doi.org/10.1093/analys/anx159.

DiPaolo, J. 2018. "Second Best Epistemology: Fallibility and Normativity." *Philosophical Studies* 176: 2043–2066. https://doi.org/10.1007/s11098-018-1110-y.

DiPaolo, J. forthcoming. "'I'm, Like, a Very Smart Person.' On Self-Licensing and Perils of Reflection." *Oxford Studies in Epistemology*.

Dogramaci, S. 2018. "Solving the Problem of Logical Omniscience." *Philosophical Issues* 28, no. 1: 107–128.

Dogramaci, S., and S. Horowitz. 2016. "An Argument for Uniqueness About Evidential Support." *Philosophical Issues* 26, no. 1: 130–147.

Earman, J. 1992. *Bayes or Bust? A Critical Examination of Bayesian Confirmation Theory*. Cambridge (MA): MIT Press.

Easwaran, K. 2011. "Bayesianism II: Applications and Criticisms." *Philosophy Compass* 6, no. 5: 321–332.

Egan, A. 2008. "Seeing and Believing: Perception, Belief Formation and the Divided Mind." *Philosophical Studies* 140, no. 1: 47–63.

Estlund, D. 2014. "Utopophobia." *Philosophy & Public Affairs* 42, no. 2: 113–134.

Gaus, G. 2016. *The Tyranny of the Ideal: Justice in a Diverse Society*. Princeton and Oxford: Princeton University Press.

Gigerenzer, G., and R. Selten. 2002. *Bounded Rationality: The Adaptive Toolbox*. Cambridge (MA): MIT press.

Greaves, H., and D. Wallace. 2006. "Justifying Conditionalization: Conditionalization Maximizes Expected Epistemic Utility." *Mind* 115, no. 459: 607–632.

Greco, D., and B. Hedden. 2016. "Uniqueness and Metaepistemology." *Journal of Philosophy* 113, no. 8: 365–395.

Griffiths, T. L., F. Lieder, and N. D. Goodman. 2015. "Rational Use of Cognitive Resources: Levels of Analysis Between the Computational and the Algorithmic." *Topics in Cognitive Science* 7, no. 2: 217–229.

Halpern, J. Y., and R. Pass. 2015. "Algorithmic Rationality: Game Theory With Costly Computation." *Journal of Economic Theory* 156: 246–268. https://doi.org/10.1016/j.jet.2014.04.007.

Icard, T. F. 2018. "Bayes, Bounds, and Rational Analysis." *Philosophy of Science* 85, no. 1: 79–101.

Jewitt, I. 1981. "Preference Structure and Piecemeal Second Best Policy." *Journal of Public Economics* 16, no. 2: 215–231. https://doi.org/10.1016/0047-2727(81)90025-6.

Joyce, J. M. 1998. "A Nonpragmatic Vindication of Probabilism." *Philosophy of Science* 65, no. 4: 575–603.

Kahneman, D. 2003. "Maps of Bounded Rationality: Psychology for Behavioral Economics." *American Economic Review* 93, no. 5: 1449–1475.

Karlan, B. 2020. "Reasoning With Heuristics." *Ratio* 34: 100–108. https://doi.org/10.1111/rati.12291.

Kitcher, P. 1990. "The Division of Cognitive Labor." *Journal of Philosophy* 87, no. 1: 5–22.

Kitcher, P. 1995. *The Advancement of Science: Science Without Legend, Objectivity Without Illusions*. Oxford: Oxford University Press.

Lawson, G. 2021. "The Epistemology of Second Best." *Texas Law Review* 100: 747–770.

Leitgeb, H., and R. Pettigrew. 2010a. "An Objective Justification of Bayesianism I: Measuring Inaccuracy*." *Philosophy of Science* 77, no. 2: 201–235.

Leitgeb, H., and R. Pettigrew. 2010b. "An Objective Justification of Bayesianism II: The Consequences of Minimizing Inaccuracy*." *Philosophy of Science* 77, no. 2: 236–272.

Lipsey, R. G., and K. Lancaster. 1956. "The General Theory of Second Best." *Review of Economic Studies* 24, no. 1: 11–32.

Longino, H. E. 1990. *Science as Social Knowledge: Values and Objectivity in Scientific Inquiry*. Princeton: Princeton University Press.

Lorkowski, J., and V. Kreinovich. 2018. *Bounded Rationality in Decision Making under Uncertainty: Towards Optimal Granularity*. Cham, Switzerland: Springer.

Meacham, C. J. G. 2015. "Understanding Conditionalization." *Canadian Journal of Philosophy* 45, no. 5–6: 767–797.

Meacham, C. J. G. 2016. "Ur-Priors, Conditionalization, and Ur-Prior Conditionalization." *Ergo* 3: 444–492. https://doi.org/10.3998/ergo.12405314.0003.017.

Morton, A. 2012. *Bounded Thinking: Intellectual Virtues for Limited Agents*. Oxford: Oxford University Press.

Ng, Y.-K. 2004. *Welfare Economics: Towards a More Complete Analysis*. London, United Kingdom: Palgrave Macmillan.

O'Connor, C., and J. Bruner. 2019. "Dynamics and Diversity in Epistemic Communities." *Erkenntnis* 84, no. 1: 101–119.

Paul, L. A., and J. Quiggin. 2018. "Real World Problems." *Episteme* 15: 363–382. https://doi.org/10.1017/epi.2018.28.

Pettigrew, R. 2016. *Accuracy and the Laws of Credence*. Oxford: Oxford University Press.

Rosenstock, S., J. Bruner, and C. O'Connor. 2017. "In Epistemic Networks, Is Less Really More?" *Philosophy of Science* 84, no. 2: 234–252.

Schoenfield, M. 2017. "Conditionalization Does Not (In General) Maximize Expected Accuracy." *Mind* 126, no. 504: 1155–1187.

Skipper, M., and J. C. Bjerring. 2020. "Bayesianism for Non-Ideal Agents." *Erkenntnis* 87: 93–115. https://doi.org/10.1007/s10670-019-00186-3.

Smithies, D. 2015. "Ideal Rationality and Logical Omniscience." *Synthese* 192, no. 9: 2769–2793.

Staffel, J. 2017. "Should I Pretend I'm Perfect?" *Res Philosophica* 94, no. 2: 301–324.

Staffel, J. 2019. *Unsettled Thoughts: A Theory of Degrees of Rationality*. Oxford, England: Oxford University Press.

Staffel, J. 2024. "Bayesian Norms and Non-Ideal Agents." In *Routledge Handbook of the Philosophy Evidence*, edited by C. Littlejohn and M. Lasonen-Aarnio. London, United Kingdom: Routledge. https://philpapers.org/rec/STABNA-2.

Talbott, W. 2016. "Bayesian Epistemology." In *The Stanford Encyclopedia of Philosophy*, edited by E. N. Zalta. Stanford, CA: Stanford University. https://plato.stanford.edu/archives/win2016/entries/epistemology-bayesian/.

Titelbaum, M. G. 2015. "Rationality's Fixed Point (Or: In Defense of Right Reason)." *Oxford Studies in Epistemology* 5: 253–294.

Todd, P. M., and G. Gigerenzer. 2000. "Précis of Simple Heuristics That Make Us Smart." *Behavioral and Brain Sciences* 23, no. 5: 727–741.

Vishwanath, T. 1992. "Parallel Search for the Best Alternative." *Economic Theory* 2, no. 4: 495–507.

Weatherall, J., and C. O'Connor. m.s. 2018. "Do as I Say, Not as I Do, or, Conformity in Scientific Networks." *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3391343.

Weitzman, M. L. 1979. "Optimal Search for the Best Alternative." *Econometrica: Journal of the Econometric Society* 47, no. 3: 641–654.

Wheeler, G. 2018. "Bounded Rationality." In *The Stanford Encyclopedia of Philosophy*, edited by E. N. Zalta. Stanford, CA: Stanford University. https://plato.stanford.edu/archives/fall2019/entries/bounded-rationality/.

Wiens, D. 2020. "The General Theory of Second Best Is More General Than You Think." *Philosophers' Imprint* 20, no. 5: 1–26. http://hdl.handle.net/2027/spo.3521354.0020.005.

Zollman, K. J. S. 2007. "The Communication Structure of Epistemic Communities." *Philosophy of Science* 74, no. 5: 574–587.

Zollman, K. J. S. 2013. "Network Epistemology: Communication in Epistemic Communities." *Philosophy Compass* 8, no. 1: 15–27.

Zynda, L. 1996. "Coherence as an Ideal of Rationality." *Synthese* 109, no. 2: 175–216.

Appendix

Our goal is to maximize a function $F(x_1, x_2, ..., x_n)$ of n variables $x_1, x_2, ..., x_n$ (for $n \geq 3$) subject to a constraint $G(x_1, x_2, ..., x_n) = 0$. This is a classical optimization problem. Using Lagrange multipliers, the optimization of $F(x_1, x_2, ..., x_n)$ is given by

$$F'_i - \lambda \cdot G'_i = 0 \qquad i = 1, 2, \ \dots\ , n \qquad \lambda \neq 0 \tag{A1}$$

where F'_i denotes the derivative of F with respect to variable x_i (and G'_i denotes the derivative of G with respect to variable x_i), and λ is a constant multiplier. Naturally, the above can be rearranged as

$$F'_i = \lambda \cdot G'_i \qquad i = 1, 2, \ \dots\ , n \qquad \lambda \neq 0 \tag{A2}$$

Which also gives us the following proportionality conditions:

$$\frac{F'_i}{F'_n} = \frac{G'_i}{G'_n} \qquad i = 1, 2, \ \dots\ , n-1 \tag{A3}$$

These proportionality conditions partly define our ideal world—that is, our ideal world is partly defined by *ratios of derivatives*. For $F(x_1, x_2, ..., x_n)$, there are n–1 ratios of derivatives between our variables (the nth ratio being trivial). If an allocation of resources is (locally or globally) optimal, it will satisfy such ratios.

Suppose a constraint makes it impossible to attain the ideal world. For the sake of simplicity, assume that a constraint disrupts the first ratio of derivatives, as in the following:

$$\frac{F'_1}{F'_n} = k \cdot \frac{G'_1}{G'_n} \qquad k \neq 1 \tag{A4}$$

k could be a constant or a function. For simplicity's sake, we assume it is a constant. More complicated results will be obtained if we assume that k is a function (i.e., steps (A5) to (A9) would be very different if k were a function).

Given (A4), we want to know whether satisfying the remaining ratios of derivatives is optimal or not. This amounts to an optimization problem that includes the new constraint stated in (A4) (e.g., the constraint that prevents us from attaining the ideal world). To solve this new optimization problem, we use Lagrange multipliers (again). After some manipulation, we can determine that the function to optimize is

$$F'_i = \mu \cdot G'_i + \nu \cdot \left[\frac{F'_n \cdot F''_{1,i} - F'_1 \cdot F''_{n,i}}{F'^{2}_n} - k \cdot \frac{G'_n \cdot G''_{1,i} - G'_1 \cdot G''_{n,i}}{G'^{2}_n} \right] \qquad i = 1, 2, \ \dots\ , n \qquad \mu \neq 0, \nu \neq 0 \tag{A5}$$

where F'_i denotes the derivative of F with respect to variable x_i, $F''_{i,j}$ denotes the second derivative of F with respect to variable x_i followed by variable x_j, and μ and ν denote constant multipliers.

After some manipulation, the conditions described in (A5) give us the following system of ratios of derivatives:

$$\frac{F'_i}{F'_n} = \frac{G'_i}{G'_n} \left[\frac{\left[1 + \frac{\nu}{\mu \cdot G'_i} \cdot \left(\frac{F'_n \cdot F''_{1,i} - F'_1 \cdot F''_{n,i}}{F'^{2}_n} - k \cdot \frac{G'_n \cdot G''_{1,i} - G'_1 \cdot G''_{n,i}}{G'^{2}_n} \right) \right]}{\left[1 + \frac{\nu}{\mu \cdot G'_n} \cdot \left(\frac{F'_n \cdot F''_{1,n} - F'_n \cdot F''_{n,n}}{F'^{2}_n} - k \cdot \frac{G'_n \cdot G''_{1,n} - G'_1 \cdot G''_{n,n}}{G'^{2}_n} \right) \right]} \right] \qquad i = 1, 2, \ \dots\ , n-1 \qquad \mu \neq 0, \nu \neq 0 \tag{A6.1}$$

For readability, assume that

$$Q_i = \frac{F'_n \cdot F''_{1,i} - F'_1 \cdot F''_{n,i}}{F{\prime_n}^2}, R_i = \frac{G'_n \cdot G''_{1,i} - G'_1 \cdot G''_{n,i}}{G{\prime_n}^2} \quad \text{(A6.2)}$$

Then, (A6.1) is equal to

$$\frac{F'_i}{F'_n} = \frac{G'_i}{G'_n} \frac{\left[1 + \frac{\nu}{\mu \cdot G'_i} \cdot \left(Q_i - k \cdot R_i\right)\right]}{\left[1 + \frac{\nu}{\mu \cdot G'_n} \cdot \left(Q_n - k \cdot R_n\right)\right]} \qquad i = 1, 2, \ \ldots, n-1 \qquad \mu \neq 0, \nu \neq 0 \quad \text{(A6.3)}$$

Equation (A6.3) gives us the n–1 optimal ratios of derivatives given the additional constraint introduced in (A4). Our goal is to determine whether, under such non-ideal conditions, it is still optimal to satisfy the remaining ideal ratios of derivatives obtained in (A3). That is, we want to know whether (A6.3) and (A3) concur. To do so, we start by assuming that the right side of the equality in (A6.3) is equal to (A3), as in the following:

$$\frac{G'_i}{G'_n} = \frac{G'_i}{G'_n} \frac{\left[1 + \frac{\nu}{\mu \cdot G'_i} \cdot \left(Q_i - k \cdot R_i\right)\right]}{\left[1 + \frac{\nu}{\mu \cdot G'_n} \cdot \left(Q_n - k \cdot R_n\right)\right]} \quad \text{(A7)}$$

Naturally, the equality will be satisfied if and only if

$$\frac{\left[1 + \frac{\nu}{\mu \cdot G'_i} \cdot \left(Q_i - k \cdot R_i\right)\right]}{\left[1 + \frac{\nu}{\mu \cdot G'_n} \cdot \left(Q_n - k \cdot R_n\right)\right]} = 1 \quad \text{(A8)}$$

By assumption, the constant multiplier ν is not equal to 0 (see step (A5), for instance). So, the equality described in (A8) will be satisfied if and only if

$$G'_n \cdot \left(Q_i - k \cdot R_i\right) = G'_i \cdot \left(Q_n - k \cdot R_n\right) \quad \text{(A9)}$$

If (A9) is satisfied, approximating the ideal ratios of derivatives in non-ideal worlds is optimal. If (A9) is not satisfied, then it is optimal to depart from the ideal ratios of derivatives in non-ideal worlds. Hence, second-best problems arise if and only if (A9) is false.

Jewitt (1981) and Blackorby et al. (1991) have argued that (A9) is satisfied if and only if some separability conditions among functions $F(x_1, x_2, \ldots, x_n)$ and $G(x_1, x_2, \ldots, x_n)$ are satisfied.[26] For instance, suppose $F(x_1, x_2, \ldots, x_n)$ and $G(x_1, x_2, \ldots, x_n)$ are additively separable, in the sense that $F(x_1, x_2, \ldots, x_n) = F_1(x_1) + F_2(x_2) \ldots + F_n(x_n)$ and $G(x_1, x_2, \ldots, x_n) = G_1(x_1) + G_2(x_2) \ldots + G_n(x_n)$. Then, there won't be second-best problems.[27]