Adversariality and Ideal Argumentation: A Second-Best Perspective

Marc-Kevin Daoust, Harvard University. mk.daoust@live.ca

Forthcoming in *Topoi*. DOI: 10.1007/s11245-021-09749-y. Please cite the published version.

Abstract: What is the relevance of ideals for determining virtuous argumentative practices? According to Bailin and Battersby (2016), the *telos* of argumentation is to improve our cognitive systems, and adversariality plays no role in ideally virtuous argumentation. Stevens and Cohen (2019) grant that ideal argumentation is collaborative, but stress that imperfect agents like us should not aim at approximating the ideal of argumentation. Accordingly, it can be virtuous, for imperfect arguers like us, to act as adversaries. Many questions are left unanswered by both camps. First, how do we conceptualize an ideal and its approximation? Second, how can we determine what is the ideal of argumentation? Third, can we extend Stevens and Cohen's anti-approximation argument beyond virtue theory? In order to respond to these questions, this paper develops a second-best perspective on ideal argumentation. The Theory of the Second Best is a formal contribution to the field of utility (or welfare) optimization. Its main conclusion is that, in non-ideal circumstances, approximating ideals might be suboptimal.

Keywords: argumentation, ideals, adversariality, norms of perfection, second-best epistemology

Introduction

Consider the following case:

Discussion. Sonia is discussing with Charles. They endorse a different political party for the ongoing election. Each of them present good evidence for why their favourite party has the best platform. Sonia has good reason to think that the Socialists will do a better job at protecting the environment, whereas Charles has good reason to think that the Conservatives have the best policy for protecting civil rights. Sonia and Charles are not trying to win the argument. At some point, they even point out some weaknesses of their own views: Sonia raises an objection against the Socialist platform, and Charles identifies an incoherence in the Conservatives's promises. However, Sonia is not fully receptive to Charles's reasons, and vice versa. They are not fully interested in hearing each other's point of view. After a couple of minutes, they stop talking about politics, and neither of them have substantially changed their mind.

Now, suppose things went differently between Sonia and Charles, as in the following:

Adversaries. Sonia and Charles still endorse a different party for the ongoing election, and they both have good evidence for why their favourite party has the best platform. However, Sonia and Charles have something to prove. *They both want to win the argument*. So, they listen closely to their opponent's arguments and reasons, and actively try to refute them. For strategic reasons, they do not point out the weaknesses of their own views. Motivated by the fear of losing face, their exchange lasts for an hour, and they both address many arguments and objections. Towards the end, Sonia raises a series of excellent arguments against the Conservatives's platform. Her objections are fatal. Charles is short of arguments. Humiliated, he revises his belief that the Conservatives have the best platform.

Were Sonia and Charles more virtuous in Discussion or Adversaries? In Discussion, Sonia and Charles do not take an adversarial stance towards each other. At first sight, adversariality seems like a vicious trait of character. So, there is something to be said in favour of Discussion. However, in Adversaries, Sonia and Charles are more attentive to each other's arguments, and Charles changes his mind after hearing some fatal objections. Being receptive and responsive to arguments and objections are good traits of character. So, from a virtue-theoretic point of view, it is unclear whether agents in Adversaries do worse than in Discussion.

Many theories of argumentation are interested with good argumentation, understood as an *activity* between arguers.[1] Evaluating arguers and the dialectical activity they engage in goes beyond the formal aspects of correct arguments, such as cogency, or the presence of good inferences. For instance, it has been suggested that an ideal arguer possesses virtues—integrity, open-mindedness, humility or intellectual perseverance, and the like.[2]

1 See, for instance, Godden (2016, 345–46). By contrast, some theories of argumentation are interested in the properties of a good argument, understood as an object or a *product*. For instance, a good argument is explicit, cogent and relies on inference principles such as deduction, induction or abduction. Context can also play a role in determining what counts as a good argument. See Walton (2009) for an overview of the literature. For classical work and references in argumentation theory, see Blair (2015) on informal logic, Cohen (1995) on adversarial argumentation and education, Gilbert (2013) on argumentation, agreement and disagreement, Hahn and Oaksford (2007) on Bayesianism and argumentation, van Eemeren and Grootendorst (2010) on pragma-dialectical approaches to argumentation, Walton (2003) on relevance in argumentation, and Walton, Reed and Macagno (2008) on argumentation schemes.

2 See Aberdein (2016) and Cohen (2005) on virtue-based argumentation theory. See Cohen and Miller (2016) on empathy, sympathy and cognitive compathy as features of correct argumentation. See Kidd (2016) on intellectual humility as an intellectual virtue. See Kwong (2016) on open-mindedness and

This paper analyzes the role of adversariality in virtuous argumentation. It focuses on a recent debate between Bailin and Battersby (2016) and Stevens and Cohen (2019). Bailin and Battersby endorse an epistemic account of argumentation, which roughly says that argumentation aims at improving our cognitive system. We can express Bailin and Battersby's thesis in terms of *ideal* argumentation. Ideally, arguers should, after having discussed with each other, have a better understanding of their own views, and be in a better position to reasonably believe various propositions. For them, adversariality plays no role in ideal argumentation, as cooperation is a better approach for improving our cognitive systems. Stevens and Cohen grant that, if we focus on ideal arguers and their perfect virtues, argumentation will be cooperative. However, this idealized picture of virtue doesn't necessarily matter in non-ideal contexts where imperfect agents argue with each other. In non-ideal contexts, approximating the ideal of argumentation might not be virtuous or optimal.

While I am sympathetic to Stevens and Cohen's main conclusion, I think that the relationship between ideal and non-ideal norms is more complicated than it seems. Some important questions are left unanswered by both camps. First, how do we conceptualize an ideal and its approximation? Second, how can we determine what is the ideal of argumentation? Third, can there be anti-approximation arguments outside virtue-theoretic accounts of argumentation?

This paper sheds light on these questions by offering a second-best perspective on ideal argumentation. The Theory of the Second Best is a formal contribution to the field of utility (or welfare) optimization. Its main conclusion is that, in non-ideal circumstances, approximating ideals might be suboptimal (with respect to a specific interpretation of what "approximating an ideal" means). Recently, authors like David Wiens (2020) have suggested that the formal model underlying the Theory can find applications beyond economics. So, I will present this model step by step, and see whether it can apply to

argumentative virtues. See also Godden (2016) and Stevens (2016).

optimal argumentation. Along the way, I will shed light on the above three questions. Specifically, I will make the following claims:

1. There is more than one way to conceptualize an ideal and its approximation;
2. Idealizing does not mean removing all constraints. If we take this idea seriously, the ideal of argumentation could be adversarial to some degree;
3. The second-best perspective suggests that there are good reasons not to approximate ideals even outside virtue-theoretic accounts of argumentation (e.g., in consequentialist or deontological accounts).

Section 1 summarizes Bailin and Battersby's argument, Stevens and Cohen's response, and the questions left unanswered by both camps. Section 2 argues that there is more than one way to conceptualize an ideal and its approximation. Section 3 argues that, in a second-best framework, adversariality can be part of the ideal of argumentation. Section 4 argues that a second-best framework lends support against the approximation of argumentative ideals.

The debate between Bailin and Battersby and Stevens and Cohen instantiates a general problem in ethics, argumentation theory and epistemology, namely, the recourse to ideals.[3] A second-best perspective allows us to see, in accordance with Stevens and Cohen's conclusion, that the approximation of ideals in non-ideal contexts can be suboptimal. But it also raises interesting conceptual issues concerning the nature of ideals. Ultimately, I hope this paper will show that the recourse to ideals is more complicated than it seems, and calls for a closer inspection.

3 See, among others, Aberdein and Cohen (2016), Cohen and Miller (2016), Estlund (2014), Gaus (2016), Hundleby (2013), Staffel (2019), and Stemplowska and Swift (2012) for recent discussion on the role of ideals in ethics, epistemology and argumentation theory.

1. Adversariality and the Ideal of Argumentation

1.1. Bailin and Battersby's Argument Against Adversariality

According to Bailin and Battersby, argumentation aim at improving our cognitive system (Bailin and Battersby 2016, 8). This means that agents "inquire in order to reach a reasoned judgment." (Ibid., 8)

Their analysis focuses on the goal of argumentation practices. While some practices may be structured in an adversarial way, the goal of such practices is to reach reasoned judgments. For instance, consider courtrooms. In Canadian or American trials, lawyers act as adversaries who try to win a debate.[4] However, the judge's (or the jury's) task is to reach a reasoned verdict on whether the defendant is guilty. Thus, the activity might seem adversarial, but the goal of the activity is to reach a reasoned judgment (Ibid., 8).

A similar point can be made for group deliberation, as they indicate in the following:

> In group deliberation, for example, it may be useful to have a participant play the role of devil's advocate to discourage groupthink or deferral to the implicit group hierarchy and to ensure that alternative arguments are given due consideration. Although this may appear to be a case of adversariality, it is really the ideas which are in confrontation. And any arguer is in a position to offer such criticisms and objections as well as to propose arguments, offer supporting arguments, revise arguments, and so on (Ibid., 10).

Although they do not use this vocabulary, we can express Bailin and Battersby's thesis in terms of *ideal* argumentation. Ideally, arguers should, after having discussed with each other, have a better understanding of their own views. They should be in a better position to take a justified stance towards the proposition under discussion. But if non-epistemic factors affected the discussion (such as the need to win the debate), argumentation might fall short of such an ideal. They write:

> In the case of mere winners, that is, when the audience is wrongly persuaded (unjustifiably persuaded), no one is a winner epistemologically. In the case of real winners, that is, when the audience is rightly, or justifiably persuaded, everyone is a winner epistemologically in that all participants have undergone

4 See Wein (2016, 2) on why there are exceptions to this claim.

> an improvement to their cognitive systems, including those who have changed their minds (Ibid., 9).

Bailin and Battersby think that a practice's goal determines what this practice requires. At the end of their paper, they write that "habits of mind or virtues such as open-mindedness, fair-mindedness, and a willingness to follow an argument where it leads can be seen as... required by its epistemic goals [of cognitive improvement]" (Ibid., 11). In other words, ideal argumentative practices aim at cognitive improvement, and the goal of cognitive improvement tells us which practices (or habits of mind) are required for good argumentation.[5]

1.2. Stevens and Cohen's Reply

Stevens and Cohen grant that, if we focus on ideal arguers and their perfect traits of character, argumentation will be cooperative. They agree that perfectly virtuous arguers want to further the aim of argumentation, namely, cognitive betterment of all the participants. They even agree with Bailin and Battersby that specific dispositions, such as open-mindedness and reasons-responsiveness, are part of ideal argumentative practices. (Stevens and Cohen 2019, 5)

However, this idealized picture of virtue doesn't necessarily matter in non-ideal contexts where imperfect agents argue with each other. They say:

> The social aspects of argumentation disrupt the apparently analytic connections between the argumentative *telos* and the cooperative stance. The arguers with whom we argue are flawed human beings; the contexts in which we find ourselves arguing are full of complex contingencies; and, let us admit, we ourselves inevitably fall short of ideal virtue. The easy, straightforward analytical answer turns out to be too easy.... A normative theory of argumentation that provides guidance only for ideal people in ideal circumstances hardly is ideal. (Ibid., 6-7)

5 Bailin and Battersby are responding to Cohen's paper titled "Missed Opportunities in Argument Evaluation" (Cohen 2015). In this paper, Cohen analyzes missed opportunities in argumentation, such as failing to notice a good argument in favour of a thesis, or failing to signal it to an opponent. There are some contexts in which agents who fail to notice or signal a good argument in favour of a thesis are subject to criticism. However, there also seems to be contexts in which missed opportunities are permitted. Cohen tries to make sense of such a datum by analyzing roles, impartiality, and adversariality in argumentation. See Wein (2016) for a reply to Bailin and Battersby's argument.

In non-ideal contexts, it can be virtuous to take an adversarial stance towards each other. For instance, imperfect agents like us are better at finding good arguments and objections when faced with opponents. That is, we make a better contribution to collective discussion when "we are set against other reasoners" (Ibid., 8). Here, Stevens and Cohen echo some observations made by Mercier and Sperber, who write:

> Imagine two engineers who have to come up with the best design for a bridge.... Ella favors a suspension bridge, Dick a cantilever bridge.... They can each build a case for their favored option. Ella would look for the pros of the suspension bridge and the cons of the cantilever; Dick would do the opposite. They would then debate which option is best, listening to and evaluating each other's arguments. To the extent that it is easier to evaluate arguments presented to you than to find them yourself, this option means less work for the same result. (Mercier and Sperber 2017, 220-1)

The gist of Mercier and Sperber's point is this: Agents like us (i.e., with our psychological traits) are not particularly good at evaluating the pros and cons of every possible option. However, we reach reliable and cost-effective results in competitive setups. So, the adversarial model of argumentation can be relevant in non-ideal contexts, since it makes a better use of our cognitive capacities.

Here is another example. In non-ideal situations like ours, some contextual factors can't be changed. Perhaps in the ideal world, there are no courtrooms, union negotiations, or debates clubs. In this world, these structures exist, and they will not disappear in a foreseeable future. These structures embrace the adversarial logic. Arguers who choose a cooperative stance in these contexts would be at a great disadvantage. As Stevens and Cohen rightly stress, "one arguer, deviating from the expected norm and staying purely cooperative because she wants to be virtuous, would destroy the requisite equilibrium and be counter-productive to the accomplishment of the argumentative *telos*" (Stevens and Cohen 2019, 9).

Thus, Stevens and Cohen grant that, in an ideal world, adversarial argumentation has no place. But they stress that, in non-ideal settings (like our world), it might be virtuous to take an adversarial stance towards each other during argumentation. They indicate that virtuous arguers "choose roles that enhance the chances for the epistemic betterment of all,

given the specific participants and the actual circumstances... certain argumentative tasks require certain argumentative virtues more than others." (Ibid., 14) So, suppose agents know that, in contexts C, cognitive improvements are greater when agents act as adversaries. Then, virtuous agents will act as adversaries in contexts C.

1.3. Three Questions Left Unanswered

I am sympathetic towards Stevens and Cohen's main conclusion (i.e., in non-ideal conditions, virtuous agents do not necessarily argue as perfect agents in ideal conditions would). Yet, I think that the nature of ideal norms, as well as the relationship between ideal and non-ideal norms, are more complicated than it seems. The exchange between Bailin and Battersby (2016) and Stevens and Cohen (2019) leaves many questions unanswered.

First, how do we conceptualize an ideal and its approximation? Both papers seem to take an ideal to be defined by a set of features, like (i) all the agents are open-minded, (ii) all the agents cooperate with each other, etc. But why not define the ideal differently? In general, philosophers are not very explicit when it comes to defining ideals.

Second, how can we determine what is the ideal of argumentation? Both papers say that the ideal of argumentation is collaborative, but it is unclear why we need to accept this. Perhaps some interpretations of the ideal of argumentation allow for adversariality, even under the assumption that cognitive improvement is the *telos* of argumentation.

Third, is Stevens and Cohen's anti-approximation argument limited to virtue theory? Part of their argument is premised on a virtue-theoretic account of argumentation. But perhaps approximation of ideals could make sense in deontological or consequentialist accounts of argumentation.

A second-best analysis of adversariality and argumentation will provide partial answers to the above three questions.

2. How Do We Conceptualize an Ideal and Its Approximation?

Approximation claims concerning ideals are ambiguous. There are various ways in which one can refer to approximating an ideal.[6] For instance, consider:[7]

Approximation in Terms of Closeness. Suppose an ideal is defined in terms of the optimal value of some variables x_1, x_2,..., x_n. Then, approximating the ideal can refer to getting as close as possible to the optimal values of x_1, x_2,..., x_n.[8]

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.[9]

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.[10]

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.[11]

Note that these characterizations are not coextensive. They can conflict with one other. So, they are not different ways of expressing the same idea.

---

6 And not all interpretations of approximation make sense. For instance, consider Kwisthout et al.'s discussion of Bayesian modeling in cognitive science. For them, a problem for Bayesian modeling in cognitive science is computational intractability. Very roughly, the problem is that Bayesian modeling sometimes proposes solutions that are too costly. The proposed solution cannot be implemented with a reasonable amount of time and resources. Some reply that Bayesian computations can be approximated using inexact algorithms. However, Kwisthout et al. (2011, 780) doubt that this interpretation of "approximation" is mathematically founded.

7 I make the same observations in Daoust (m.s.).

8 See, e.g., Gaus (2016, chap. 1).

9 Some of Staffel's (2019) distance measures also reflect Approximation of Value. See also Gaus (2016, chap. 1).

10 See, e.g., DiPaolo (2019) and Räikkä (2000). DiPaolo and Räikkä say that they are concerned with second-best problems. However, The Theory of the Second-Best is concerned with Approximation of Relationships, not Approximation of Features.

11 See, e.g., Lipsey and Lancaster (1956) and Wiens (2020).

Consider, for instance, Approximation in Terms of Closeness and Approximation in Terms of Value. Suppose Bob has a blood clot. His welfare is a function of a variable x, namely, the amount of anticoagulant he receives (in millilitres). The optimal value of x is 10: Bob will be completely cured and be in no pain if he receives 10 millilitres of anticoagulant. If he doesn't do anything, he will be in great pain, but he will survive. However, if Bob gets 12 or more millilitres of anticoagulant, he will agonize for hours and die. Getting 12 millilitres of anticoagulant is a good approximation of the ideal dose. However, if Bob got this dose, he would be worse off than if he got no treatment. So, while x=12 is closer to the ideal value of x (10) than x=0, the disvalue of x=12 is greater than the disvalue of x=0. Hence, Bob's case shows that Approximation in Terms of Closeness and Approximation in Terms of Value can conflict with each other.[12]

Many philosophers, along with Stevens and Cohen, are interested in Approximation of Features. They compare the features of ideal worlds with the features of the optimal worlds available to us, and come to the conclusion that approximating the ideal world might be suboptimal. Consider the following case. In the ideal world, I would visit my grandma on Sunday, and I would call her in advance and tell her that I will visit her. Now, suppose that the ideal world is unavailable to me, but that I have the following options:

Option 1. (i-) I do not visit grandma and (ii) I call her in advance, telling her that I will visit.

Option 2. (i-) I do not visit grandma and (ii-) I do not call her in advance to tell her that I will visit.

Surely, if I am not going to visit grandma, I should not raise her hopes. So, Option 2 is better than Option 1. However, while Option 1 and the ideal world have one attribute in common (I call grandma in advance and tell her that I will visit her), Option 2 and the ideal world have no attribute in common. As we can see, it would be absurd to claim that Option 1 is the best available option just because it is a better approximation, *in terms of features*, of the ideal world.

12 See Gaus (2016, chap. 1) for a similar argument.

By contrast, the Theory of the Second Best is concerned with a different interpretation of what ideals consist in. The Theory is concerned with Approximation of Relationships. In accordance with familiar principles found in economics, the Theory starts by identifying a goal (denoted by a function F), the variables (a, b, c,...) at play for achieving this goal, and at least one initial constraint limiting its achievement (a function G=0). Then, the Theory identifies the optimal relationships among the variables.[13]

For illustration purposes, suppose your goal is to make a good cake. The variables at play for this goal are the almonds, bananas and chocolate available to you, and you have a limited budget for buying these ingredients. Then, "F" will be the function of how good your cake is, it will be a function of variables (a, b, c), namely, almonds, bananas and chocolate, and "G=0" will reflect a constraint on your budget for buying the ingredients.

Recall that, for economists, finding ideals is an optimization problem. The problem consists in figuring out the values of some variables that minimize or maximize a function. We want to optimize the allocation of resources—that is, we want to find the allocation of resources that optimizes a given goal. This means that, of the values of a, b and c that satisfy the constraint, we want to find the ones that optimize our function. In other words, of the values of a, b and c that satisfy G=0, we want to find the ones that optimize F.

The method of Lagrange multipliers can be used for finding a function's optima subject to some constraints, which is exactly what economists are doing. Using Lagrange multipliers, the optimization of a function $F(x_i, x_2,...x_n)$ subject to a constraint $G(x_i, x_2,...x_n)$ is given by

$$F'_i = \lambda \cdot G'_i \qquad i = 1, 2, \ldots, n \qquad \lambda \neq 0$$

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.

The method of Lagrange multipliers allows us to express our ideal world in terms of ratios of derivatives (or, for simplicity, ratios of variables). The ratios are given by the following:

13 See Wiens (2020), Lipsey and Lancaster (1956), and Ng (2004). See also Daoust (m.s.).

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

These ratios follow directly from the method of Lagrange multiplier How? Simply divide $F'_i=\lambda\cdot G'_i$ by $F'_n=\lambda\cdot G'_n$ for $i=1,2,\ldots,n-1$, and you get the ratios stated above.

Again, suppose your goal is to make a nice cake (F). As we said, F is a function of three variables (a, b, c). You need to take a monetary constraint (G=0) into account. Then, the ideal world will be expressed by two ratios of variables, as in the following:

First ratio: $\frac{F'_a}{F'_c}=\frac{G'_a}{G'_c}$

Second ratio: $\frac{F'_b}{F'_c}=\frac{G'_b}{G'_c}$

To make a long story short, there is not a 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 the above ratios, are more technical and mirror optimization principles found in decision theory. Accordingly, rejecting the approximation of ideals is more complicated than it seems, since there are many different ways of interpreting what such an approximation amounts to. Perhaps one interpretation of ideals and their approximation can succeed.

3. How Can We Determine What Is the Ideal of Argumentation?

Bailin and Battersby agree with Stevens and Cohen that the ideal of argumentation is collaborative. But why think that? Depending on how we understand ideals and their approximation, things get complicated. In a second-best framework, adversariality could be part of the ideal of argumentation.

3.1. The Ingredients of A First-Best Scenario

Let's see how the Theory can provide a new perspective on ideal argumentation. First, in order to apply the Theory to a given problem, we need at least three elements:

- We want to maximize a function F;
- F is a function of at least three continuous variables;
- F is subject to an initial constraint G.[14]

Bailin and Battersby (2016, 8) think that virtuous argumentation aims at improving our cognitive systems, that is, "to inquire in order to reach a reasoned judgment" (Bailin and Battersby 2016, 8). Stevens and Cohen (2019, 4-5) seem to agree with this assumption. However, both papers also mention that, as arguers, we want to reach the right answer, or be accurate (Bailin and Battersby 2016, 9; Stevens and Cohen 2019, 5). So, F could be a function describing the improvements of our cognitive systems, or the improvement of accuracy.[15] Stevens and Cohen suggest that these aims are consistent with each other. Specifically, having better cognitive systems is instrumental for accuracy: When agents improve their cognitive systems, they become more accurate. As they say, "arguers will want to get [the right answer]... by being open-minded and reasons-responsive" (Stevens and Cohen 2019, 5). In view of the foregoing, I will assume that our *ultimate* goal is to maximize accuracy. Being virtuous, improving our cognitive systems, or being reasonable are means for achieving this final goal.

14 See Wiens (2020, secs. 2–3) and Daoust (m.s.).

15 For full belief, one has an accurate belief in P if and only if one's belief is true. For credence, accuracy comes in degrees. If P is true, then the more one's credence in P is closer to 1, the more one has an accurate credence in P. If P is false, then the more one's credence in P is closer to 0, the more one has an accurate credence in P.

Now, we need to find at least three variables that matters for optimizing accuracy in argumentative contexts.[16] Here is a tentative list of factors and traits of character that can improve or reduce accuracy in argumentative contexts:

Ability to provide and respond to reasons (r): To what degree are agents capable of processing and responding to their reasons? Do they ignore some of their evidence? Do they share all the relevant reasons they have? How well do they reason from their evidence?

Motivation for winning the argument and other non-epistemic factors **(w)**: Why are agents engaged in argumentation? Are they trying to figure out the truth of a proposition, of do they also have other goals, like winning? How important are these other goals—that is, to what extent are agents motivated by winning?

Interaction and receptiveness among arguers **(i)**: Are agents interacting with each other, and to what extent? For instance, are agents elaborating their arguments separately, and letting a third party decide which argument is more plausible? Or do they discuss with each other, and take into account each other's arguments and objections? If they discuss with each other, to what extent are they receptive and listen to what the others are saying?

Most of these variables can be found the exchange between Bailin and Battersby and Stevens and Cohen. For instance, Stevens and Cohen (2019, 5) say that perfectly virtuous agents are reasons-responsive. They also note that adversariality can come in degree (Ibid., 8). They point out that weak forms of adversariality (where agents care about winning, but also care about finding the truth) are less problematic than strong ones. Bailin and Battersby (2016, 2, 4, 11) stress the importance of cooperation and open-mindedness in perfect argumentative practices, which echoes receptiveness. They also argue that competitiveness can undermine the goal of accuracy (e.g., winning the argument doesn't

16 Why do F and G need to be functions of at least three variables? Recall that the Theory of the Second Best defines the first-best scenario in terms of ratios of derivatives. In our cake example, this is given by two ratios—namely, $\frac{F'_a}{F'_c}=\frac{G'_a}{G'_c}$ and $\frac{F'_b}{F'_c}=\frac{G'_b}{G'_c}$. Then, it analyzes cases in which some ratios of derivatives are satisfied, while other ratios are not. For instance, if a constraint refrains us from satisfying the first ratio, should we still satisfy the second one? However, if F and G are functions of less than three variables, you will not have two ratios of derivatives. If you only have two variables, you have at most one ratio, and the question "if a constraint refrains us from satisfying the first ratio, should we still satisfy the second one?" makes no sense.

necessarily entail being right or more reasonable) (Ibid., 2). These variables can also be found in the broader empirical literature on argumentation. For instance, Mercier and Sperber (2017, 220–21, 224–27) say that interactions and adversariality among agents allow for cost-effective argumentation.

I do not want to imply that *only* the above variables are relevant for optimizing accuracy in argumentative practices. Plausibly, there are other variables at play. Traits of character such as compathy, empathy, and humility could be other relevant variables for determining how an argument is likely to lead agents to the right answer.[17] Yet these three initial variables will give us a *partial picture* of ideal argumentation.

These variables can be normalized, i.e., treated as continuous measures ranging from 0 to 1 (Wiens 2020, sec. 3.1). For instance, suppose variables r, w, and i can take values between 0 and 1. For instance, if r=0, then agents are absolutely incapable of providing and responding to reasons, and if r=1, then agents are fully capable to provide and respond to reasons. Similarly, we can assume that, when w=0, non-epistemic factors like winning the argument are maximally important, and when w=1, they do not matter at all. Normalizing the variables simplifies the model without any loss of generality. Various social indicators are normalized (the Gini coefficient, the Human Development Index, etc.).

If we normalize the range of possible values of r, w and i, we can locate various "idealtypes"[18] of argumentation. Suppose (r, w, i) is a vector representing a type of argumentation determined by some values of r, w, and i. Then, (1, 1, 1) could be the idealtype of collaborative argumentation. Agents would greatly interact and be in a position to provide and respond to reasons, and they would entirely ignore non-epistemic factors. They would only care about the truth. By contrast, (1, 0, 1) could be the idealtype of adversarial argumentation. Agents would give maximal importance to some non-epistemic factors, like winning the argument. There could be other idealtypes. For instance, (1, 0, 0)

17 See note 2.

18 An idealtype refers to a case in which the essential features of a phenomenon are most accentuated. In accordance with the Weberian tradition, an idealtype can be disvaluable or unethical. There can be, for instance, idealtypes of dictatorship, corruption, mafia, and so forth.

could be the idealtype of something called "non-conversational argumentation." Agents would try to win the argument by to convincing a third-party (the voters, a judge, a jury, etc.) that they are right, but they would never interact with their opponent or the third party. They would limit themselves to making a positive case for their own view.

Finally, F could be subject to an initial constraint. For instance, even in idealized scenarios, agents do not have an infinite amount of time, or they do not have infinite cognitive capacities for processing all the reasons and objections they hear, and so forth. These types of constraints mirror the ones found in economics, where ideals do not presuppose that resources are infinite, or that technology is fully advanced. There can be an ideal division of resources matching *the ones we have access to*, or an ideal tax structure relative to the technology within our grasp.[19] The same goes for argumentation.

3.2. Some Degree of Adversariality Can Be Part of the Argumentative Ideal

Given what we said in the previous subsection, there is no reason to exclude that the ideal of argumentation will allow for some degree of adversariality. Given F and constraint G, assigning a value of less than 1 to w could optimize our function.[20] This would mean that, relative to the initial constraints agents face, *some* degree of adversariality (or *some* importance given to winning the argument) can optimize F.

Suppose, for illustration purposes, that our functions F and G look like the following:

(1) $F(r,w,i)=(r\cdot w)+(w\cdot i)+(i\cdot r)$

(2) $G(r,w,i)=r+(w^2\cdot i^2)+\frac{w^2+i^2}{4}-1=0$

Note that functions F and G could be different. In section 4, I will explain how we can identify second-best problems without knowing the particular functional forms of F and G.

19 Also, think of Kitcher's (1990) ideal division of epistemic labour. Kitcher roughly says that diversity can be essential for optimizing an ideal community's division of epistemic labour. But Kitcher does not assume that ideal epistemic communities have access to infinite resources. He takes the bounded resources available to an epistemic community as an initial constraint.

20 Recall that, when w=0, non-epistemic factors like winning the argument are maximally important, and when w=1, they do not matter at all.

So, we could run the argument without identifying the particular functional forms of F and G. I picked these equations to give concrete support to the claim that the ideal of argumentation can allow for some degree of adversariality.

Let's take a closer look at the general features of our functions. Begin with F. Every variable in F contributes to accuracy. That is, all things being equal, an increase in either variables contributes to overall accuracy. If there were no constraint, optimizing F would mean that all the variables take a value of 1. In other words, if there were no initial constraint, the collaborative conception of argumentation would be optimal. Bailin and Battersby, as well as Stevens and Cohen, would agree with this feature of F.

As for our initial constraint G, it has the following properties: First, it entails that we cannot fully realize all the variables simultaneously. It is impossible that (w, r, i)=(1, 1, 1). Think of a context in which agents lack the *resources* for being perfectly reasons-responsive, fully receptive, and concerned only with truth-related factors. Perhaps they have a limited amount of time and mental energy to argue with each other, or they cannot change some facts about their psychology, and so forth. In such contexts, there has to be some trade-offs between the variables.

Second, for every increase in the ability of providing and responding to reasons, agents need to give up something: Either agents will care more about non-epistemic factors, or they will interact less with each other. Think of Sonia and Charles's cases discussed in the introduction. In Discussion and Adversaries, Sonia and Charles are sensitive to arguments and objections. This could mean that, in both cases, r takes a high value. But in Discussion, they are not very receptive to each other (and, hence, i takes a low value). In Adversaries, they give a high importance to winning the argument (and, hence, w takes a low value). Hence, these cases could illustrate that an increase in r comes at the cost of a decrease in either i or w.

Using the method of Lagrange multipliers discussed above, we can define our ideal ratios of variables as follows:

(3) First ratio: $\frac{w+i}{r+w}=\frac{1}{2i\cdot w^2+0.5i}$

(4) Second ratio: $\frac{r+i}{r+w}=\frac{2w\cdot i^2+0.5w}{2i\cdot w^2+0.5i}$

As it happens, our function F reaches its highest point (≈1.218) when our variables take the following values:

(5.1) $(r,w,i)\approx(0.577,0.668,0.668)$

(5.2) $F(0.577,0.668,0.668)\approx 1.218$

Of course, such values of r, w, and i satisfy the ratios stated in (3) and (4).

Note that variable w, which represents the importance of non-epistemic factors (like winning), takes a value of less than one (w=0.668). This means that, in order to optimize F, we gave *some* importance to non-epistemic factors. The upshot here is that, under one or multiple initial constraints like G, the goal of accuracy can be optimized by allowing *some* importance to winning the argument. Accuracy through argumentation will be optimized if agents try, to some extent, to win the argument. So, in a second-best framework, there is no reason to exclude the possibility that some degree of adversariality is part of the ideal scenario.

Here is another way to put it. Suppose we added one constraint to our model, namely, w=1. That is, suppose we really wanted to make sure that agents give absolutely no importance to non-epistemic factors. If we make sure that w=1, what is the highest value of our function F? This amounts to a new optimization problem, with the following additional constraint:

(6) $G_2(r,w,i)=1-w=0$

In other words, we would optimize F under the constraints stated in (2) and (6). Under such constraints, our function F is optimized when our variables take the following values:

(7.1) $(r,w,i)\approx(0.522,1,0.427)$

(7.2) $F(0.522,1,0.427)\approx 1.172$

We can then compare (7.2) with (5.2). Without the constraint w=1, our function reached a higher point (e.g., 1.218>1.172). In other words, adding the constraint w=1 lead to a suboptimal result. Thus, in this particular case, it is better to allow for some degree of adversariality.

Once again, to be clear: The equations in (1)-(7.2) merely provide an *example* of what the ratios of variables defining an argumentative first-best could look like. I have no reason to think that these functions F and G are the "real" functions representing the optimization of accuracy under constraints. And since there are probably other variables that matter for optimizing F, this is, at best, a partial sketch of our function. But this shows that a first-best scenario can include some degree of adversariality. All I wish to stress is that, a priori, we cannot rule out this possibility.

4. Anti-Approximation Arguments Beyond Virtue Theory

In this section, I suggest that the Theory of the Second Best can provide a general framework against approximation claims in argumentation. The framework doesn't presuppose any commitment to consequentialism, deontologism, or virtue theory. It can be used in accordance with various philosophical doctrines. For instance, economists use second-best framework to analyze the distribution of goods optimizing *welfare* (Lipsey and Lancaster 1956). Wiens (2020, sec. 3) illustrates how the framework can be applied to Rawls's (2009) deontological ideals. In the previous section, we used it to shed light on the balance of argumentative virtues that optimize accuracy. The framework is flexible and can be used in accordance with different moral theories.

As we saw in section 3, a first-best scenario is characterized in terms of ratios of variables. For instance, the ideal argumentative scenario could be characterized by the ratios stated in (3) and (4). In non-ideal worlds, some of the ideal ratios cannot be satisfied. For instance, perhaps in this world, we cannot satisfy the ratio stated in (3). Then, we want to know whether it is optimal (or suboptimal) to satisfy the remaining ratios, like (4). The Theory says that, in some conditions, it is suboptimal to satisfy the remaining ratios. In

other words, there are conditions in which, when we cannot satisfy at least one of the ideal ratios, we are better off not satisfying the remaining ones. What are these conditions?

We can determine whether a function F is subject to second-best problems without knowing its particular functional form. As Wiens (2020, sec. 3) notes, a function F is subject to second-best problems if the following four conditions are met:

- We want to maximize F;
- F is subject to an initial constraint G;
- F and G are functions of at least three continuous variables;
- Some "separability conditions" are violated in either F or G.

In other words, if the above four conditions are met, it can be suboptimal, in non-ideal conditions, to approximate the ratios defining the ideal world. Regardless of their particular functional forms, if F and G satisfy the above criteria, they will trigger the mathematical theorem underlying the Theory (Wiens 2020, sec. 3.1).

The first three conditions were discussed in section 3. As we saw, F could be a function of accuracy. In argumentative contexts, F could a function of at least three variables, such as responsiveness to reasons, the importance agents give to non-epistemic factors, and the interaction and receptivity of agents. And insofar as "ideal" is not a synonym of "infinite resources" or "perfect agents with unlimited cognitive capacities," F can be subject to an initial constraint G, such as the resources available to agents.

We already have our first three ingredients for arguing that there can be second-best problems in argumentative contexts. All we need to know is whether some "separability conditions" are violated in either F or G. But what does this mean, exactly?

Nonseparability is a technical notion. Here is an accessible summary of what nonseparability consists in.[21] A good sign that functions $F(x_1, x_2,..., x_n)$ and $G(x_1, x_2,..., x_n)$ are nonseparable is that, in order to optimize F under constraint G, one cannot optimize the value of the variables $x_1, x_2,..., x_n$ independently of each other. Optimizing F under

21 Mathematically speaking, the violation of separability conditions have to do with the system of partial derivatives that result from optimization in second-best scenarios. See Daoust (m.s.).

constraint G requires taking the value of all the variables x_1, x_2,..., x_n into account *simultaneously*. Think of the following cross effects between variables:

Synergy. 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.

Substitution. 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.

Attenuation. 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.

We have good reason to think that some variables in argumentative contexts display cross effects. Accordingly, we have good reason to think that there are plausible second-best problems in argumentative contexts. In the remainder of this section, I discuss three plausible cross effects that we could observe between variables r, w, and i.

4.1. Cross Effects Between Responsiveness to Reasons and Interaction in F

When it comes to optimizing accuracy, there can be cross effects between variables i and r. Specifically, these variables can be synergetic. Interaction provides new perspectives that can be beneficial to arguers, in the sense that it allows them to develop better argumentative schemes. Citing Deanna Kuhn's (Kuhn 1991) study on the ability, for isolated agents, to find arguments and objections in favour of their own point of view, Mercier and Sperber say:

> We encountered Deanna Kuhn's study of argumentation demonstrating the difficulty most people have in finding counterarguments.... The same study showed that even in support of their own point of view, people often give rather weak arguments. (Mercier and Sperber 2017, 223)

Then, Mercier and Sperber argue that the same kind of agents (e.g., agents who are not more knowledgeable, more competent, etc.) do better when they interact with each other. They write:

> The experiments presented earlier, which prompted psychologists to deplore the poor quality of the reasons put forward by participants, did not take place in a typical dialogic context. When a normal interlocutor is not swayed by a reason, she offers counterarguments, pushing the speaker to provide better reasons. An experimenter, by contrast, remains neutral. She may prompt the participant for more arguments, but she doesn't argue back. If reason evolved to function in an interactive back-and-forth, strong arguments should be expected only when they are called for by an equally strong pushback. (Mercier and Sperber 2017, 227)

The gist of Mercier and Sperber's argument is that, given the facts of our psychology, interaction improves our capacity to find good arguments and objections in favour of our points of views. This suggests that variables i and r display cross effects for optimizing accuracy in argumentative contexts.

4.2. Cross Effects Between Responsiveness to Reasons and Interaction in G

Naturally, if we take Mercier and Sperber's argument seriously, coupling i and r can affect the resources available for arguing with each other. Reasons-responsive agents who interact with each other have to spend more time and cognitive efforts to process arguments and objections, since they must process new arguments and objections that they would otherwise not be aware of.

The costs associated to combining interaction and reasons-responsiveness are *not merely additive*. Interaction without argumentation doesn't cost anything, since agents lack the capacity to engage in any kind of reasons-based exchange. Argumentation without interaction doesn't cost much, since agents have a limited access to arguments and objections (i.e., they only have access to their own arguments and objections). But when

high interaction is combined with high reasons-responsiveness, arguing with each other can become costly.[22]

Note, of course, that these costs might very well be worth it. As we saw in the previous subsection, reasons-responsive agents who interact with each other do a very good job at finding good arguments and objections in favour of our points of views. So, *for equal results in terms of accuracy or reasonableness*, combining r and i is cost-effective. But this is compatible with the presence of a cross effect between i and r in terms of costs.

4.3. Cross Effects Between Responsiveness to Reasons and the Motivation for Winning in F

When it comes to optimizing accuracy, combining responsiveness to reasons and low motivation for winning makes agents more likely to present objections and arguments that do not favour their own views. Since arguers mainly care about the truth, they do not "hide" arguments and objections. By contrast, when agents care about winning, they do not reveal objections against their own views, or do not reveal arguments that could be useful for their opponents. In other words, adversaries waste some of their abilities to provide and respond to reasons.

Consider Charles and Sonia's cases discussed in the introduction. In Discussion, Charles and Sonia do not care about winning, and they freely discuss objections against their own views. By contrast, in Adversaries, they care about winning. If they are well prepared and they want to increase their chance of winning the argument, they choose to ignore some objections and arguments favouring their opponent. This is one of the main differences between Discussion and Adversaries.

So, there is a plausible cross effect between responsiveness to reasons and the motivation for winning. Improving the ability of *collaborators* to provide and respond to reasons contribute directly to accuracy. By contrast, improving the ability of *adversaries* to provide and respond to reasons might not contribute as much to accuracy, since adversaries do not make full use of their ability to provide reasons.

22 Compare: Driving your car for an hour won't cost you much. Drinking a bottle a vodka won't cost you much. But drinking a bottle of vodka just before driving your car could be extremely costly. Here, the costs associated with these activities are somehow synergetic.

4.4. Taking Stock

This partial survey of three possible cross effects between variables r, w, and i suggests that the separability conditions are violated in either F or G. Accordingly, it can be suboptimal, in non-ideal argumentative contexts, to approximate the ratios defining ideal argumentation. This conclusion can be reached without knowing the particular forms of F and G.

Perhaps you are curious, and you want to know what happens, mathematically, when the variables in either F or G are nonseparable. So, here is a simple example showing what can happen. It is based on the equations introduced in section 3.

In section 3, we identified, for illustration purposes, two ratios that could define our ideal world. To simplify the reading, here they are again:

(3) First ratio: $\frac{w+i}{r+w}=\frac{1}{2i\cdot w^2+0.5i}$

(4) Second ratio: $\frac{r+i}{r+w}=\frac{2w\cdot i^2+0.5w}{2i\cdot w^2+0.5i}$

These ratios define the relationships between our variables in the ideal world.

Suppose the following constraint refrains us from attaining the ideal world:

(8) $\frac{w+i}{r+w}=\frac{2}{2i\cdot w^2+0.5i}$

If (8) is true, it is impossible to satisfy the first-best ratio stated in (3). *However, we could still satisfy the remaining first-best ratio*—namely, the one stated in (4). Should we satisfy this ratio? Is this a good idea? To answer these questions, we need to solve a new optimization problem. In this new problem, we still aim at optimizing F, but this time, we take both the initial constraint (2) and the additional constraint (8) into account.

Under such constraints, F's highest point is ≈1.108. This second-best optimum is observed when variables r, w and i take the following values:

(9.1) $(r, w, i) \approx (0.328, 0.626, 0.946)$

(9.2) $F(0.328, 0.626, 0.946) \approx 1.108$

That is, (9.1) expresses the values of r, w and i that optimize F under constraints (2) and (8). However, the values of r, w, and i in (9.1) violate the first-best ratio stated in (4). So, we *could* satisfy (4), but this would be *suboptimal*. Thus, the Theory's warning applies here: If we want to optimize a function, like F, in non-ideal circumstances, we sometimes have to depart from the first-best ratios of variables defining the ideal world. Or: approximating ideals can be suboptimal. We have reached Stevens and Cohen's main conclusion, but with a completely different argument.

5. Conclusion

This paper focused on a recent debate on the relevance of ideals for determining virtuous argumentative practices. A second-best perspective can shed light on the (ir)relevance of argumentative ideals for imperfect agents like us.

For Bailin and Battersby, adversariality plays no role in ideally virtuous argumentation —rather, it is collaborative. Stevens and Cohen grant that ideal argumentation is collaborative, but stress that imperfect agents like us should not aim at approximating the ideal of argumentation. At least three questions were left unanswered by both camps. A second-best perspective partially answered them. Here is a summary of the questions and answers discussed in sections 2 to 4:

1. *How do we conceptualize an ideal and its approximation*? Response: There is more than one way to do this. In fact, there are at least four different ways to define what ideals and their approximation are. Philosophers tend to be interested in the *features* of ideal worlds, whereas economists and decision theorists are interested in the *relationships* between variables at play in ideals.
2. *How can we determine what is the ideal of argumentation*? Response: In a second-best perspective, the ideal of argumentation is a matter of optimization under constraints.

Contra Bailin and Battersby, as well as Stevens and Cohen, I cast doubt on the claim that the ideal of argumentation is perfectly collaborative. There can be some initial constraints on the ideal scenario. Ideal doesn't necessarily mean "unlimited resources," or "perfect agents with unlimited time and cognitive capacities." If we take initial constraints seriously, the ideal of argumentation could be adversarial to some degree.

3. *Can there be anti-approximation arguments beyond virtue theory*? Response: Yes. A second-best framework shows that, in some conditions, approximation of ideals can be suboptimal. In section 4, I primarily focused on traits of character optimizing accuracy (e.g., receptiveness, reasons-responsiveness, and cooperation). However, the model underlying the Theory does not presuppose any commitment to virtue theory. So, when it comes to argumentation, second-best problems could be identified in deontological or consequentialist frameworks, provided that we identify relevant variables matching these moral theories.

I hope to have shown that argumentation theorists can appeal to ideals in different ways, and that defining ideals is no simple task. We should pay closer attention to how we use ideals in argumentation theory.

Acknowledgments. I warmly thank two anonymous referees for helpful comments on this project.

Funding. This study was funded by the Social Sciences and Humanities Research Council (grant #756-2019-0133).

Conflict of Interest. Marc-Kevin Daoust declares that he has no conflict of interest.

Ethical approval. This article does not contain any studies with human participants performed by any of the authors.

References

Aberdein, Andrew. 2016. “The Vices of Argument.” *Topoi* 35 (2): 413–422.

Aberdein, Andrew, and Daniel H. Cohen. 2016. “Introduction: Virtues and Arguments.” *Topoi* 35 (2): 339–43. https://doi.org/10.1007/s11245-016-9366-3.

Bailin, Sharon, and Mark Battersby. 2016. “DAMed If You Do; DAMed If You Don’t: Cohen’s ‘Missed Opportunities’” OSSA11 Conference Archive (90). https://scholar.uwindsor.ca/ossaarchive/OSSA11/papersandcommentarie/90.

Blair, J. Anthony. 2015. “What is informal logic?.” In *Reflections on theoretical issues in Argumentation Theory* (pp. 27-42). Springer, Cham.

Cohen, Daniel H. 1995. “TArgument is War... and War is Hell: Philosophy, Education, and Metaphors for Argumentation.” *Informal logic 17*(2). DOI: 10.22329/il.v17i2.2406

———. 2005. “Arguments That Backfire.” In *The Uses of Argument*, edited by D Hitchcock and D Farr, 58–65. OSSA, Hamilton.

———. 2015. “Missed Opportunities in Argument Evaluation.” In *Reflections on Theoretical Issues in Argumentation Theory*, edited by Frans H. van Eemeren and Bart Garssen, 121–130. Switzerland: Springer International Publishing. https://www.doi.org/10.1007/978-3-319-21103-9_9.

Cohen, Daniel H., and George Miller. 2016. “What Virtue Argumentation Theory Misses: The Case of Compathetic Argumentation.” *Topoi* 35 (2): 451–460.

Daoust, Marc-Kevin. m.s. “What Second-Best Epistemology Could Look Like.” Unpublished Manuscript

DiPaolo, Joshua. 2019. “Second best epistemology: fallibility and normativity.” *Philosophical Studies* 176 (8): 2043-2066.

Estlund, David. 2014. “Utopophobia.” *Philosophy & Public Affairs* 42 (2): 113–134.

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

Gilbert, Michael A. 2013. *Coalescent Argumentation*. Routledge.

Godden, David. 2016. “On the Priority of Agent-Based Argumentative Norms.” *Topoi* 35 (2): 345–357.

Hahn, Ulrike, & Mike Oaksford. 2007. “The rationality of informal argumentation: a Bayesian approach to reasoning fallacies.” *Psychological review*, *114*(3), 704-732.

Hundleby, Catherine. 2013. “Aggression, Politeness, and Abstract Adversaries.” *Informal Logic* 33 (2): 238–262.

Kidd, Ian James. 2016. “Intellectual Humility, Confidence, and Argumentation.” *Topoi* 35 (2): 395–402.

Kitcher, Philip. 1990. “The Division of Cognitive Labor.” *The Journal of Philosophy* 87 (1): 5–22.

Kuhn, Deanna. 1991. *The Skills of Argument*. Cambridge: Cambridge University Press.

Kwisthout, Johan, Wareham, Todd and Iris van Rooij. 2011. “Bayesian intractability is not an ailment that approximation can cure.” *Cognitive Science* 35 (5): 779-784.

Kwong, Jack MC. 2016. “Open-Mindedness as a Critical Virtue.” *Topoi* 35 (2): 403–411.

Lipsey, Richard G., and Kelvin Lancaster. 1956. “The General Theory of Second Best.” *The Review of Economic Studies* 24 (1): 11–32.

Mercier, Hugo, and Dan Sperber. 2017. *The Enigma of Reason*. Cambridge (MA): Harvard University Press.

Ng, Yew-Kwang. 2004. *Welfare Economics: Towards a More Complete Analysis*. Palgrave Macmillan.

Räikkä, Juha. 2000. “The problem of the second best: Conceptual issues.” *Utilitas* 12 (2): 204-218.

Rawls, John. 2009. *A Theory of Justice: Revised Edition*. Cambridge (MA): Harvard University Press.

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

Stemplowska, Zofia, and Adam Swift. 2012. “Ideal and Nonideal Theory.” In *The Oxford Handbook of Political Philosophy*, edited by David Estlund, 373–90. Oxford: OUP.

Stevens, Katharina. 2016. “The Virtuous Arguer: One Person, Four Roles.” *Topoi* 35 (2): 375–383.

Stevens, Katharina, and Daniel Cohen. 2019. “The Attraction of the Ideal Has No Traction on the Real: On Adversariality and Roles in Argument.” *Argumentation and Advocacy* 55 (1): 1–23.

Van Eemeren, F., and Rob Grootendorst. 2004. *A Systematic Theory of Argumentation: The Pragma-Dialectical Approach*. Cambridge: Cambridge University Press.

Walton, Douglas. 2003. *Relevance in Argumentation*. Routledge.

———. 2009. "Argumentation Theory: A Very Short Introduction." In *Argumentation in Artificial Intelligence*, edited by G Simari and I Rahwan, 1–22. Boston (MA): Springer

Walton, Douglas, Reed, Chris & Fabrizio Macagno. 2008. *Argumentation schemes*. Cambridge: Cambridge University Press.

Wein, Sheldon. 2016. "Commentary on DAMMIT-- Dominant Adversarial Model: Minded Instead of Terminated (A Commentary on 'DAMmed If You Do, DAMmed If You Don't' by Sharon Bailin and Mark Battersby)" OSSA11 Conference Archive (44). https://scholar.uwindsor.ca/ossaarchive/OSSA11/papersandcommentarie/44.

Wiens, David. 2020. "The General Theory of Second Best Is More General Than You Think." *Philosophers' Imprint*. 20 (5): 1-26. HDL: 2027/spo.3521354.0020.005