Good and *ought* in argumentation: COVID-19 as a case study

Andrés Soria-Ruiz[1], Mora Maldonado[2] and Isidora Stojanovic[3]

[1] Nova Institute of Philosophy, School of Social Sciences and Humanities, Universidade NOVA de Lisboa. Avenida de Berna 26C, 1069-061 Lisbon. aruiz@fcsh.unl.pt.

[2] Departament de Traducció i Ciències del Llenguatge, Universitat Pompeu Fabra. Edifici Roc Boronat, Roc Boronat 138, 08018 Barcelona. mora.maldonado@upf.edu

[3] Institut Jean-Nicod, Département d'Études Cognitives, ENS, EHESS, CNRS, PSL Research University, Pavillon Jardin, 29 rue d'Ulm, 75005 Paris. isidora.stojanovic@cnrs.fr

Abstract

The present paper concerns arguments whose conclusions take the form of a prescription such as *you ought to do such-and-such*, which have centered much public discussion and policy since the beginning of the COVID-19 pandemic. We aim to tackle a hitherto under-explored characteristic of many such normative arguments, namely, the relationship between evaluative and deontic propositions, depending on whether they occur as premises or conclusions in such arguments. In order to investigate how willing people are to argue from what is good to what one ought to do, and the other way round, we conducted an Inferential Judgment Experiment. Participants were presented with arguments involving deontic and evaluative propositions, and had to judge whether they could infer the conclusion from the premise. The stimuli that we used are tightly related to the argumentation surrounding the pandemic, regarding the measures of preventing the spread of COVID-19. The results of the study show that there is a robust inferential connection between evaluatives and deontics, but at the same time, a significant asymmetry as well. We explore several theoretical approaches to the relationship between the deontic and the evaluative realm, and test their predictions against the results of our study.

1   Introduction: evaluative and deontic propositions

The topic of this paper are arguments with normative conclusions, as illustrated by instances of argumentation that have become prevalent during the COVID-19 pandemic. In particular, we aim to explore the inferential connection between evaluative and deontic concepts[1], more specifically, between 'good' and 'ought'.

Note, to start, that contrary to Hume's famous *dictum*, it does not seem difficult to derive an *ought* from an *is*:

---

[1] In this paper, we subsume under the umbrella term 'normative' all expressions/arguments with consequences for the behavior and attitudes of people; chiefly, this includes *evaluative* and *deontic* expressions. *Evaluative* terms are, roughly, those that make reference to value, that is, to what is good or bad. *Deontic* terms are those that make reference to obligations, permissions and prohibitions. Most notably, the latter include deontic modals, such as *ought*, *must* or *may* (see Carr 2017).

(1) a. Wearing a face mask in public is good.
  b. Ergo, you ought to wear a face mask in public.

This argument looks perfectly sound; and yet, there is no 'ought' in the premise. One would immediately respond, however, that 'is good' is an evaluative expression, and hence that this is not a counterexample to the claim that one cannot derive an 'ought' from an 'is', since evaluative and deontic expressions belong to the same domain of broadly normative expressions. But this is our point of departure: while evaluative and deontic expressions are often subsumed under the same broad class, there are important, often overlooked, differences between them.[2] Nevertheless, whether evaluatives and deontics also differ in their argumentative properties and roles, and if so, how, are questions that, to our knowledge, have never been properly explored.

1.1 Similarities and differences between deontic and evaluative language and concepts

There are three possible approaches to adopt regarding the relation between the deontic and the evaluative realm:

A1: The differences between evaluatives and deontics that exist in other domains have no impact on argumentation. Evaluatives and deontics are, from an argumentative point of view, pretty much interchangeable.
A2: While evaluatives and deontics exhibit different argumentative patterns, there are interesting logical connections between the two.
A3: There is no special or interesting logical relationship between evaluatives and deontics.

A1 and A3 are two extreme positions; we present them here as mere conceptual possibilities. A1 makes two predictions: first, appropriately similar evaluative and deontic judgments should be overall interchangeable in argumentation. For example, every judgment about what is good should be substitutable for a judgment about what ought to be the case or what one ought to do, and *vice versa*, without affecting the overall validity of an argument. Secondly, it should be possible to derive 'ought' from 'good' and 'good' from 'ought'. Insofar as Moore (1903) aimed to reduce *right* to *good* (by treating the former as denoting instrumental value), he may be seen as a defender of A1; similarly, Hare's (1952) prescriptivism may be seen as an attempt to account for evaluative and deontic expressions (chiefly *good*, *right* and *ought*) as particular species of

---

[2] See e.g. Tappolet 2013, Sven Ove Hansson 2001; see also Hurka 2014, Ch. 2 for a useful survey of the distinction between 'good' and 'ought' in mid-XXth century British ethical philosophers.

prescriptive language.[3] In stark contrast, A3 rejects the existence of any inferential or logical connection whatsoever between evaluatives and deontics. To our knowledge, A3 has never been defended or even endorsed, but we are including it as part of the conceptual space.

It is more common to find endorsement of (some version of) A2 in the literature. A2 is a weaker thesis than A1, since it is not committed to the first prediction above: if evaluatives and deontics exhibit different argumentative patterns, we should not expect these expressions to be interchangeable *salva validitate*. Moreover, A2 does not imply that one can derive *ought* from *good* or the other way around. However, A2 is compatible with the existence of some inferential connections between deontic and evaluative expressions (in either direction), contrary to A3.

Indeed, it is intuitive to think that something akin to A2—or perhaps A1—lurks beneath the usual reaction against the purported counterexample to the *is-ought* gap above. 'Good' and 'ought' belong to the same broad class of normative expressions, so the impression that in (1) one has jumped from the descriptive to the normative realm is an illusion—the premise was *already* normative. In the remainder of this section, we consider the motivation for A2 by considering similarities and differences between evaluative and deontic expressions and concepts.

Consider first the similarities between the two families of concepts. Broadly speaking, there are reasons to think that deontics and evaluatives belong to the same conceptual realm. This is supported by the observation that both classes of expressions seem to admit the same contextual qualifications and parameters. Specifically, they show a clear dependence on different "flavors" of evaluation, they are circumstance-dependent in broadly the same ways, and they are action-guiding. Let us review these characteristics in turn.

First, we speak of categorical and hypothetical value; similarly, we can think of categorical and hypothetical obligations or duties.[4] Categorical values and obligations concern invariable and universal ends such as attaining well-being; instrumental values and obligations are those that subserve means to particular, variable ends. Linguistically, judgments of categorical value/obligation are most often expressed with unqualified uses of the relevant evaluative/deontic expressions; while judgements of hypothetical value/obligation introduce qualifications such as 'in order to φ'. See in this regard the contrast between (2) and (3):

     (2) a. You ought to visit your family.

---

[3] It bears mentioning, however, that Hare observed important semantic differences between 'good' and 'ought'. For instance, it seems possible to say 'You did that quite well, but not yet quite as you ought' without contradiction (1952, p.152). Regardless, his chief purpose was to argue that the same considerations that make 'good' a value word make 'ought' a value word as well.

[4] We clearly speak of aesthetic value; and while we might not speak of aesthetic duties, the very notion of the canon of beauty reflects the idea that people may want to comply with rules and norms when it comes to aesthetic matters.

       b. Visiting your family is good.
   (3) a. You ought to visit your family in order to prevent the spread of COVID-19.
       b. Visiting your family is good in order to prevent the spread of COVID-19.

Clearly, we may agree that families involve (more or less) categorical values and obligations, but these can be overridden by a hypothetical goal, i.e., that of preventing the spread of COVID-19.

Secondly, values and obligations are circumstance-dependent: wearing a mask might be a bad idea for people with certain respiratory conditions; and whether you ought to stay at home, for instance, depends on whether you are safe there.

Finally, values and obligations are action-guiding: ascribing value to something invites the thought that one is motivated to orient one's actions towards it; the same goes for judging that one ought to do something.[5] Relatedly, influential accounts of normativity in metaethics, such as *buck-passing* (Scanlon 1998, Rowland 2019) or *expressivist* accounts (Stevenson 1937; Ayer 1946; Gibbard 1990, 2003) have been applied to account for both evaluative and deontic expressions. Often, authors in these metaethical traditions speak indistinctly of deontic and evaluative normativity. These considerations invite the conclusion that, at an important level of description, the deontic and the evaluative belong to the same realm. It is thus suggestive to expect that, when deontic and evaluative concepts appear in argumentation, one can in some instances move from one class of concepts to the other, and vice versa.

However, there are important differences between both classes of concepts as well. Some of these differences were highlighted by classical ethicists. Hurka (2014) discusses Sidgwick, Moore and Ross as authors all of whom thought that there existed interesting inferential connections between *good* and *ought*, even though there exist important points of contrast between the domain of the evaluative and the deontic. For example, in *The Methods of Ethics* (1874/2019) Sidgwick calls 'right' an *imperative* concept, while 'good' is merely *attractive*. Another important difference between 'ought' and 'good', highlighted by Sidgwick, Ross or Prichard is the well-known *ought implies can* principle, stemming from Kant: if someone ought to do something, then they can do it. But from the fact that an action or state of affairs is good, it does not follow that it can be brought about. Some other authors, like Russell (1910/2009), took 'good' to be the widest normative term, while 'ought' applies only to voluntary acts. Relatedly,

---

[5] There are exceptions to this, in both domains. For example, 'it would be good if I were taller', or 'it ought to be Christmas everyday' are things one might say even though it is not clear how they could have an impact on action. (We owe these examples to an anonymous reviewer.) Evaluative and deontic concepts are very varied and have many different uses; we are happy to recognize that not all of these concepts in all of their uses are action-guiding in the way we characterize that property here, but for our purposes we will assume that, in many cases, both classes of concepts have important practical consequences, and that this feature can be used to mark a contrast with other types of concepts. In previous work, we have proposed some linguistic criteria that help decide whether a use of a term, be it evaluative or not, has action-guiding implications: see Soria Ruiz & Stojanovic (2019) for discussion.

'good', at least on its surface, denotes a simple property of objects.[6] 'Ought', on the other hand, denotes a relation between an agent and whatever they are obliged to do: obligations are imposed upon specific individuals. Moreover, simple evaluatives such as 'good' and 'bad' take a wide range of arguments: individuals, objects, actions or states of affairs can be 'good'.[7]

> (5) a. Jacinda Ardern is a good politician.
>   b. These masks are bad.
>   c. Washing your hands is good.
>   d. It's good that the WHO issued new policies promoting physical exercise.

By contrast, 'ought' can be applied only to actions and states of affairs.

Finally, another point of contrast between evaluative and deontic expressions concerns their gradability: in general, evaluative terms are gradable, while deontic terms are not (Tappolet 2013, Stojanovic 2017). In other words, values come in degrees, while obligations are 'on/off'; either they are in place or not at all. The gradability of evaluative expressions is attested by the admission of adjectival modifiers, such as 'very', 'slightly' or the comparative form. Evaluatives admit them clearly, while deontics sound odd with them (marked with #).

> (4) a. Wearing a mask in public is {very / slightly} good.
>   b. Wearing a mask in public is better than washing your hands.
> (5) a. # Wearing a mask in public is {very / slightly} forbidden.
>   b. # Wearing a mask indoors is more forbidden than wearing a mask outdoors.

This implies that to call something 'good' is not simply to ascribe to it a discrete property that objects either have or lack (such as, e.g., being dead or being married); it involves ascribing a scalar property to a certain degree, and thus implies a tacit reference to a contextually determined threshold or standard (see Cresswell 1976; Kennedy 2007; Kennedy and McNally 2005 for classic references). By contrast, to ascribe an obligation or a prohibition is to attribute a property

---

[6] This may involve something of a simplification. If we analyze semantically the property denoted by 'good', we see that it is a *relational* property, and that it is so in more than one way. First, as with any gradable predicate, the semantics of 'good' is sensitive to scales and thresholds (see Stojanovic 2017, p. 127). Second, 'good', at least in some uses, involves a hidden argument that denotes a beneficiary: for instance, sleeping ten hours a day may be good for a baby or a toddler, but not for an adult (see Stojanovic 2016, Soria Ruiz and Faroldi 2020). Also, whatever the semantic denotation of 'good', its uses can pragmatically imply a variety of things, including obligations. As an anonymous reviewer suggests, saying "Tidying up your room is good" to your child may be interpreted as a request that they (ought to) tidy up their room.

[7] When it comes to other evaluative adjectives, we see some variability. Thus predicates such as 'tasty' only take individuals as arguments, while other evaluative predicates, such as 'important', are more like 'good': e.g. The *Tractatus* is (an) important (book), washing your hands is important, and it's also important that a policy promotes physical exercise.

that something either has or lacks; there are no degrees of prohibition [obligation, permission], something is either forbidden [obligatory, permitted] or not.[8]

These observations allow for the possibility that there are interesting logical, inferential or conceptual connections between deontic and evaluative concepts, even if they are ultimately different concepts.

1.2   The inferential connection between 'good' and 'ought'

For the purpose of the present paper, we will limit our attention to the relationship between 'good' and 'ought', and, more specifically, between propositions of the schema 'φ-ing is good' and 'you ought to φ' (where φ stands for an action-type or a policy). The three approaches above can be put into correspondence with the following hypotheses about the argumentative relation between 'good' and 'ought':

H1: 'φ-ing is good' and 'you ought to φ' are logically equivalent and interchangeable in argumentation.
H2: 'φ-ing is good' and 'you ought to φ' are logically asymmetric; either 'φ-ing is good' implies 'you ought to φ' (but not the other way round) or 'you ought to φ' implies 'φ-ing is good' (but not the other way round).
H3: 'φ-ing is good' and 'you ought to φ' are logically independent.

Again, H1 and H3 have to our knowledge not been explicitly defended, but we are including them for reference. Before we move on, we should point out that H2 is more specific than A2; H2 concerns the relation of logical implication between 'good' and 'ought', whereas A2 simply acknowledged the existence of logical connections between the evaluative and the deontic domain. These might be different from the relation of logical implication described in H2, and they could involve other evaluative and deontic concepts ('bad' and 'forbidden', for example).

We should also mention that, even though, to our knowledge, the inferential connection between 'good' and 'ought' has not been explored, the philosophical and linguistic literature contains well-known definitions of *ought* in terms of scalar goodness, that is, in terms of *betterness*. For example, Lewis (1973 Ch.5) holds that 'ought φ' is true iff φ is *better* than $\sim\varphi$. Along similar lines, Sloman (1970)'s definition is that 'ought φ' is true iff φ is the *best* alternative (where the alternatives to φ are contextually determined; see also Wedgwood 2017). More recently, Lassiter

---

[8] The matters are somewhat more complicated. Although many of the paradigmatic deontic expressions are not gradable, Lassiter (2017, Ch.8) offers evidence for the gradability of some deontic expressions, such as the modal 'ought' or the verb 'to matter'. For example, 'ought' admits comparatives ('British Royals ought to be beaten with sticks now, even more than they ought to have been in 1776'). Moreover, there are gradable deontic adjectives, such as 'important' (Portner & Rubinstein 2016). Conversely, some evaluative expressions, such as so-called extreme adjectives ('excellent'), fail the standard tests for gradability (see Morzycki 2012).

(2017, Ch.8) takes issue with such definitions. His argument against Lewis and Sloman's view is, roughly, that it makes bad predictions for *supererogatory* acts. Supererogatory acts are acts that, while good, lie beyond the call of duty. For instance, suppose that your friend is ill and needs your help to clean her apartment. Given this, it's true that you ought to help your friend. It would be even better, however, if you helped your friend and also baked her a cake. However, baking a cake is not something that you ought to do—what you ought to do is help her. Baking is optional. Lassiter argues convincingly that both Sloman and Lewis's definitions of 'ought' wrongly predict that, if helping your friend *and* baking the cake is better than helping and not baking, then helping and baking is also what you ought to do.

To prevent this, Lassiter settles for a weakened (that is, unidirectional) version of Sloman's definition, which he calls Sloman's Principle:

Sloman's Principle: if ought φ, then φ is the best among its alternatives.

Thanks to its unidirectionality, this principle allows for the possibility of *best* alternatives that are not part of one's duties, thus allowing for supererogation. Interestingly, however, Lassiter does not discuss any connections between *ought* and *good*, only between *ought* and *better*.

The hypotheses above make predictions about the way in which sentences containing 'good' and 'ought' are expected to behave in argumentation. H1 predicts that speakers will have no problem going from premises of the form 'φ-ing is good' to 'you ought to φ' as well as in the reverse direction. Consequently, they are expected to accept any argument that takes them from 'good' to 'ought' or from 'ought' to 'good'. H2 predicts that speakers will only accept one of those directions: either they will accept as valid arguments that take them from 'φ-ing is good' to 'you ought to φ' and reject arguments that take them from 'you ought to φ' to 'φ-ing is good', or the other way around. Finally, H3 predicts that speakers are expected to reject both directions throughout.

1.3 Assessing the hypotheses empirically

In order to test these hypotheses, and thereby the plausibility of the different approaches to the place of evaluatives and deontics within a broader theory of normativity, we have run an Inferential Judgment Experiment, where we instructed participants to decide whether an argument involving deontic and evaluative expressions made sense or not.

As experimental items, we have focused on examples concerning the current COVID-19 pandemic. This choice was motivated by the observation that the global sanitary crisis to which the world has been exposed has given rise to a tremendous amount of argumentation, at the

public, private and scientific level. The need of urgently taking new measures, such as imposing lockdowns, curfews, and instauring various other means of prevention, be they strict requirements or mere recommendations, has been accompanied by a need of justifying such new policies, and those, in turn, typically take the form of an argument that largely uses evaluative and deontic language. While such arguments can often be complex, at least from a linguistic point of view, we have opted for a simplified version. Our experimental study deployed as stimuli arguments such as the following:

> (6)
> a. You ought to wear a mask in public to prevent the spread of COVID-19. Therefore, wearing a mask in public is good to prevent the spread of COVID-19.
> b. Wearing a mask in public is good to prevent the spread of COVID-19. Therefore, you ought to wear a mask in public to prevent the spread of COVID-19.

Note that when participants are asked to assess such arguments, they will typically have a personal opinion regarding their premises and conclusions. Who hasn't wondered whether it is really such a good idea to wear a mask on the street, or while jogging outdoors? Who hasn't asked themselves whether they ought or not wear gloves when shopping? Who hasn't reflected upon the policies and guidelines that were designed and implemented in the face of global, unprecedented uncertainty about *what one ought to do*? The COVID-19 pandemic, and the measures against it, are topics that have spurred the most practical concerns, deliberations and decisions in the last year.

However, precisely because of this, the participants in our study may easily let their first-order opinions about the actions and policies under consideration override their judgment about the logical or argumentative connection between 'ought' and 'good'. For instance, participants may be inclined to assent to (6a) or (6b) simply because they assent to the idea that wearing masks in public is good and/or something that they ought to do. This is why, in addition to instances of argumentation involving specific policies and actions, we have included in our study stimuli over which the participants cannot be opinionated, because they lack sufficient information in order to form an opinion of their own, such as the following:

> (7)
> a. Following the safety procedure is good to prevent the spread of COVID-19. Therefore, you ought to follow the safety procedure to prevent the spread of COVID-19.
> b. You ought to follow the safety procedure to prevent the spread of COVID-19. Therefore, following the safety procedure is good to prevent the spread of COVID-19.

Even if examples such as (7a) and (7b) are still purportedly about the COVID-19 pandemic, they resemble more the kind of stimuli preferably used in experiments of the sort, which aim to avoid

eliciting any personal involvement from the participants. As we shall shortly see, having stimuli of both types in our study permits us to see how this personal interference impacts speakers' judgments about the arguments involving 'good' and 'ought'.

2   An experimental study of 'good' and 'ought' in argumentation

In order to assess the logical relation between the deontic 'ought' and the evaluative 'good', we used an Inferential Judgment Task method, in which participants have to decide whether one can justifiably draw a certain conclusion from a given premise (see Hansen & Chemla, 2018 for a similar methodology). In this case, participants were asked to decide whether a series of arguments involving COVID-19 measures 'make sense' or not, and in some cases, to provide a short justification.[9] An example of a trial is given in **Figure 1**.

Participant's responses are taken to reflect inference patterns. We assume that if a participant considers that a certain argument 'makes sense', it is because they can infer the conclusion from the premise. We further assume that if two arguments are not supported to the same extent, then they must underlie different inferential processes, that is, they cannot be logically equivalent.We then test (a) whether participants support arguments involving 'good' and 'ought' (e.g., 6), and (b) whether they support (6a) and (6b) to the same extent.[10]

---

[9] Participants had to provide justification for a subset of five critical items, randomly chosen. These justifications were treated as debrief information, and analysed only in an exploratory manner.

[10] It is important to note that, in the context of our experiment, the acceptance of a certain argument is not indicative of its source: participants might accept both (6a) and (6b) but for different reasons. As a result, our experiment is unable to disprove H2.

Figure 1. Illustration of a critical trial in the *ought>good* Directionality condition.

2.1 Participants

43 English-speaking participants were recruited online using Prolific and were paid 0.8£ for their participation, which lasted approximately 5 minutes. Participants who had not answered correctly more than ⅔ of each type of control trial (see below) were excluded from the analysis. This led to the exclusion of 10 participants.

2.21 Design and Materials

Arguments were of the form [*p. Therefore, q*]. For simplicity, we represent these arguments as *p>q*, where *p* is the premise and *q* the conclusion. In critical trials, arguments involved the propositions '*φ*-ing is good' and 'you ought to *φ*'. We manipulated which of these propositions was the premise (*p*) and which one was the conclusion (*q*) (henceforth, the Directionality factor). There were two possible Directionalities: *good>ought* or *ought>good* (henceforth G>O and O>G).

Arguments also varied in their content (*φ*). There were 12 different contents, including both specific actions and non-specific actions (see **Table 1**). Specific actions were action-types that participants could be antecedently opinionated about (e.g., washing your hands), whereas

non-specific actions were policies or guidelines whose content was unknown to participants (e.g., following the safety protocol), and thus could not be opinionated about. We thus obtained 24 directionality/content combinations. Each participant was presented with a subset of 12 critical trials such that there were always 6 trials per Directionality controlled for specificity of the action, but without repeated content. All propositions included the phrase 'to prevent the spread of COVID-19' in order to ensure that the statements were read instrumentally (i.e. as specific pandemic-related measures) and not categorically.

Action	Content (φ)
specific	Wearing a mask in public
specific	Disinfecting your groceries
specific	Drinking bleach
specific	Washing your hands
specific	Going to house parties
specific	Avoiding people
non-specific	What Mr. Jackson proposed
non-specific	The new traffic measures
non-specific	The policies first adopted by Celsinsky
non-specific	Following the committee's advice
non-specific	Adopting the University guidelines
non-specific	The safety protocol

Table 1. Possible Contents in Critical Items.

To ensure that the participants were sensitive to classical inference patterns, we additionally included Control trials. Controls featured either valid entailments such as (8a) (upward inferences) or invalid entailments such as (8b) (downward inferences). Participants were presented with six controls of each type, randomly selected from a list of 24 statements. Control performance served as our exclusion criterion.

(8)
a. Cases are on the rise in every European country. Therefore, cases are on the rise in France.
b. Cases are on the rise in France. Therefore, cases are on the rise in every European country.

2.32 Control Truth-Value Judgment Task

Responses in an inferential task like the one presented here might be influenced by participants' first-order opinions about the truth of the premises and/or conclusions. For example, participants might be more likely to accept an argument whose premise they believe to be true than one with a false premise, regardless of the structure of the argument. This may be further boosted by the specific nature of the content used in our experiment, which regards a topic of public debate, namely, the COVID-19 pandemic and the measures undertaken against it. In other words, if the contents of our stimuli were purely abstract actions (e.g., 'being kind' or 'arriving late'), participants might find it easier to abstract away from their first order judgments about the premises and/or conclusion and assess the relevant inferences on their own merits.

In order to evaluate whether the truth of the premises had an impact on the inference patterns, we included a control truth-value judgment task (TVJT) at the end of the experiment. Participants were told they were going to see a series of statements and they had to decide whether the statements were true, false or whether they could not tell (i.e., 'I am not sure' option). Participants were then presented with the same premises of the twelve critical arguments that they had seen during the inferential task, and were asked to decide whether the statements were true, false or whether they could not tell. Participants' responses during this TVJT would then be used to estimate the role influence of perceived truth-value on inference acceptability.

2.4 Procedure

Participants were directed to a web-based Inferential Judgment experiment, implemented using JavaScript. Participants were told they were going to see statements about the pandemic, and they had to decide whether these statements made sense or not. Control and critical items were then presented in a random order. In the second part of the experiment, participants were introduced to the Control Truth-Value Judgment task. They were told they were going to see similar statements to the ones they had seen before, but now they had to decide whether the statements were true, false or whether they could not tell (i.e., 'I am not sure' option).
All materials, including instructions and both critical and control items, can be found in the following OSF repository.

3   Results

Figure 2 shows the mean proportion of 'make sense' responses for critical trials as a function of the Directionality condition. **Figure 3** breaks this down depending on the participants' truth value judgments during the second part of the experiment (TVJT). That is, as a function of whether participants perceived the premises to be true, false or whether they weren't sure.

We ran a logistic mixed effect model to evaluate the effect of Directionality (G>O; O>G), of Truth Value Judgment (true, false, neither) and of the Directionality:Truth Value interaction on

participants' responses. Directionality and Truth Value Judgments were sum-coded. Responses were coded as a binary variable (1 if 'makes sense'; 0 otherwise). We included by-participant random intercepts and by-Directionality slopes. P-values were obtained on asymptotic Wald tests, and the standard alpha level of 0.05 was used to determine significance. All analyses were carried out in R (R Core Team, 2019), using the lme4 package (Bates et al 2014). Data and scripts for the analyses can be found here.

The results of the model are reported in **Table 2**. The intercept value shows the grand mean of 'makes sense' responses across participants and conditions (i.e., for both Directionalities, and for all three Truth Values). The fact that this intercept is positive and significant reveals that the log-odds of 'makes sense' responses is significantly higher than what one would expect by chance. That is, participants tend to accept the critical arguments, that is, arguments involving 'good' and 'ought' such as (6a-b) above, regardless of their Directionality or Truth Value Judgment.

The Directionality value shows the difference between responses for O>G arguments and the grand mean. The effect of Directionality is negative and significant, revealing that the proportion of 'makes sense' responses was significantly lower in the O>G directionality than in the G>O one. The model also reveals a significant effect of Truth Value Judgments, suggesting that the truth value that participants assign to the premise of an argument has an impact on whether they accept the argument.

Finally, our model also assesses the interaction between Directionality and Truth Value Judgment (i.e., whether participants' sensitivity to Directionality depends on the truth value assigned to their premises), but this was only marginally significant.

	Value (β)	SE	z-score	p-value
(Intercept)	1.68	0.34	4.95	<0.001***
Directionality	-0.71	0.29	-2.5	0.01**
Truth Value (assigned)	1.69	0.25	6.73	<0.001***
Directionality:Truth Value	-0.42	0.25	-1.67	0.09˙

Table 2. Model output.

A second post-hoc analysis was then run to determine whether the effect of Directionality interacted with participants' *certainty* with respect to the truth of the premises; that is, with the difference between assigning a proper truth value ('true' or 'false' judgments) and being uncertain ('I am not sure' judgments). This second model specifically tested an interaction

between Directionality and Certainty in participants' inferences. In this case, the results reveal a significant Directionality:Certainty interaction ($p<0.01$), indicating that the effect of Directionality is stronger for those arguments about whose premises the participants were uncertain.
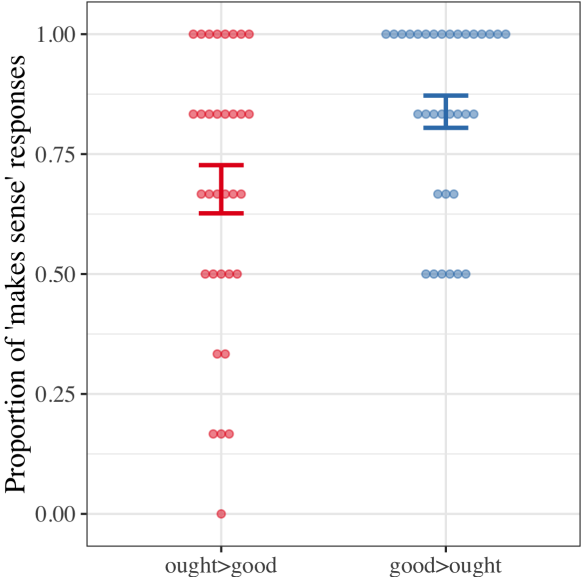

Figure 2. Proportion of 'makes sense' responses in the Inferential Task as a function of the Directionality condition. Error bars represent standard error on by-participant means; dots represent individual participant means.
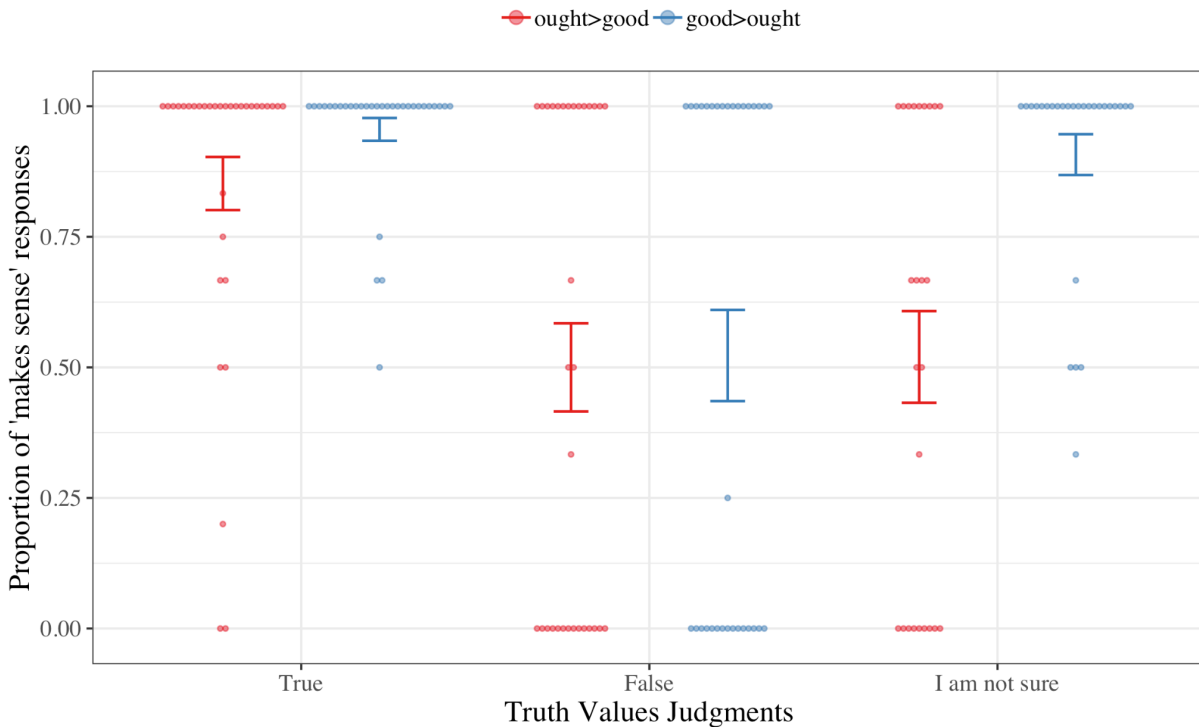
Figure 3. Proportion of 'makes sense' responses in the Inferential Task as a function of the Directionality condition and the Truth Value Judgment assigned during the TVJT. Error bars represent standard error on by-participant means; dots represent individual participant means.

4   Discussion

Our results show that, overall, participants were sensitive to the G>O/O>G asymmetry in the following direction: arguments showcasing the G>O direction of inference were accepted to a higher degree than arguments exemplifying the O>G pattern. However, these results must be taken with a grain of salt. First, it is not the case that participants rejected the O>G pattern of inference. The results are above 50% acceptance in both directions. This means that, overall, inferences in both directions were accepted by most participants. Secondly, the main effect of Directionality was largely driven by arguments about whose premises the participants were unsure. As shown in **Figure 2**, the starkest contrast in the acceptability of G>O/O>G inferences is for arguments such that when asked about their truth value in the control TVJT, the participants replied 'I am not sure'.[11] By contrast, the items that were classified as true were such

---

[11] Premises that participants classified as true/false, on the one hand, and 'I am not sure', on the other, did not necessarily coincide with the items that we classified as specific (e.g., washing your hands) vs. non-specific (e.g., following the safety procedure).

that participants generally accepted both directions of inference (in both cases, results are at ceiling), while for items that the participants classified as false, both directions were around 50%.

This might suggest, we think, that participants have their "inferential capacities" overridden, or at least biased, by their first-order judgment about the premise (recall that the control TVJT asked participants about the truth of the premise—not the conclusion—of the items they had previously seen). In other words, if someone thinks that washing one's hands is good to prevent COVID-19, they're likely to accept inferences of the form: 'Washing your hands is good to prevent the spread of COVID-19. Therefore, you ought to wash your hands to prevent the spread of COVID-19'. And similarly, if someone agrees that one ought to wash one's hands to prevent the spread of COVID-19, then they are likely to assent to an inference of the form: 'You ought to wash your hands to prevent the spread of COVID-19. Therefore, washing your hands is good to prevent the spread of COVID-19'.

On the other hand, for premises about whose truth participants were unsure, participants were significantly more sensitive to the difference between the G>O and O>G directions. It is suggestive to think that, in these cases, participants paid less attention to the content of the premises/conclusion than to the form of the argument. In those cases, participants agreed more often with the G>O than with the O>G direction.

We conclude that, while both directions were found to be acceptable, our study provides very strong evidence that the inference from 'good' to 'ought' is stronger than the inference from 'ought' to 'good'. At a very minimum, we take these results to speak against hypothesis H3, namely, the hypothesis that 'φ-ing is good' and 'you ought to φ' are logically independent. The fact that participants accepted both directions of entailment suggests that such schemas are *not* independent.

That leaves us with H1 and H2. H1 is the hypothesis that 'φ-ing is good' and 'you ought to φ' are logically equivalent, and thus interchangeable; H2, that 'φ-ing is good' and 'you ought to φ' are logically asymmetric, that is, either 'φ-ing is good' entails 'you ought to φ' but not the other way around, or 'you ought to φ' entails 'φ-ing is good' but not the other way around. Our results, however, are in tension with both hypotheses.

On the one hand, the fact that there exists a significant asymmetry in the degree to which participants accept G>O and O>G inferences suggests that 'φ-ing is good' and 'you ought to φ' are not logically equivalent. If they were, we would not expect to see a main effect of Directionality. This speaks against H1.

On the other hand, while our results reveal an asymmetry between G>O and O>G inferences, they are not compatible with the view that O>G arguments are invalid. Both G>O and O>G inferences were found to be overall acceptable (significantly above 50%). This contrasts with our control results, where participants generally *rejected* invalid entailments (e.g., 'Cases are on the rise in France. Therefore, cases are on the rise in every European country'). This indicates that the participants do not interpret O>G inferences in the same way in which they interpret an invalid entailment. Rather, our results seem to suggest that G>O and O>G trigger *different* types of inferences.

Thus, we think that there is room for more nuanced versions of H2, as well as other potential explanations. In the remainder of the paper, we shall discuss two hypotheses that, we think, can account for our data. The two hypotheses that we explore are mutually compatible. What is more, they do not exclude that there could be other explanations as well, some of which may appeal to contents that are pragmatically implicated, rather than semantically entailed. For reasons of space, we shall set such pragmatic explanations aside.

4.1 A difference in the context-sensitivity of *good* and *ought*

In the introduction, we discussed the possibility that *good* and *ought* are context-sensitive expressions. There, we insisted on the similarities between their context-sensitivity: both expressions admit of similar "flavors" of evaluation, and are *circumstance*-dependent. However, it could be that some of these parameters are determined differently for 'good' and 'ought'. In particular, 'good' and 'ought' could be circumstance-sensitive in different ways. Suppose that 'ought' denotes what is beneficial in the current circumstances, while 'good' denotes a stable, long-lasting benefit. This implies that what is good *in general* is something that ought to be the case *in the current circumstances*, but not necessarily the other way round: what ought to be the case *in current circumstances* need not be good *in general*. Recall the two basic directions of inference:

> (6)
> a. You ought to wear a mask in public to prevent the spread of COVID-19. Therefore, wearing a mask in public is good to prevent the spread of COVID-19.
> b. Wearing a mask in public is good to prevent the spread of COVID-19. Therefore, you ought to wear a mask in public to prevent the spread of COVID-19.

Intuitively, when one thinks of (6a), it is possible to accept the premise because we are in very special circumstances: one ought to wear a mask because we are in a particularly acute phase of the pandemic, and we are in a state of a global sanitary crisis. But one may reject the conclusion because, other things being equal, being 'good' seems to require a more general assessment; that

is, something is good if it is beneficial *in the long run*, or *in general circumstances*.[12] The fact that wearing a mask is beneficial in the actual, special circumstances does not guarantee that it is beneficial in general, or that it has benefits beyond what is required by the current circumstances. By contrast, if something is good, then it has benefits that apply to circumstances in general, and thus (most likely) apply to the current circumstances as well.

The simplest way to capture this circumstance-sensitivity would be to say something like the following:
- 'φ-ing is good' is true iff φ-ing is beneficial in general
- 'One ought to φ' is true iff φ-ing is beneficial in the current circumstances

Clearly, the inference from something being beneficial in general to something being beneficial in the actual circumstances is stronger than the inference from something being beneficial in the actual circumstances to it being beneficial in general.

Note however, that even though this is a way of modelling an asymmetry between 'good' and 'ought', it does not yet account for our results. Recall that we observed that both directions, G>O and O>G, are largely acceptable. Is this hypothesis compatible with the overall acceptability of these inferences? We think that it is. But showing this requires saying something about inferences involving generics.

If, as suggested by this hypothesis, the contrast between *good* and *ought* reduces to the contrast between going from a generally-quantified statement to a particular instance and *vice versa*, our hypothesis would be bolstered if such inference patterns showed a similar asymmetry. Consider the following inference:

> (9) a. Dogs have 4 legs. Therefore, this dog has 4 legs.
> b. This dog has 4 legs. Therefore, dogs have 4 legs.

Obviously, (9a) holds promise for being a sensible inference, but less so the other way around. However, there is a sense in which an inference like (9b) might be taken to be acceptable. After all, we learn about general properties of things *via* observation of particular instances. Depending on the property or feature involved, the observation of a single instance might be sufficient to make a generalization.

---

[12] The idea that 'good' tends to refer to what is good in general, rather than simply what is good right here and right now, may involve some oversimplification. As previously stressed, 'good' is a highly context-sensitive expression, and in certain contexts, it can of course be meant and interpreted as referring to the current circumstances. For instance, if we are attending a concert outdoors, you might say "It is good that it's not raining", meaning, roughly, that it is good for the purpose of attending the concert that it is not raining. Such circumstance-specific readings are, of course, possible, and do not undermine our more general point.

This contrast opens up the possibility that the inference from a generalization to a particular instance is almost as strong as an entailment (because generalizations admit exceptions, of course), but in the other direction, it becomes a weaker form of inductive inference.[13]

We think that this might account for the observed asymmetry between G/O, repeated here:

(6)
a. You ought to wear a mask in public to prevent the spread of COVID-19. Therefore, wearing a mask in public is good to prevent the spread of COVID-19.
b. Wearing a mask in public is good to prevent the spread of COVID-19. Therefore, you ought to wear a mask in public to prevent the spread of COVID-19.

In (6a), according to our hypothesis, we would have a *quasi*-entailment from a generalization (*wearing a mask is beneficial in general to prevent the spread*) to a particular instance (*wearing a mask is beneficial in these circumstances to prevent the spread*). By contrast, in (6b) we would find the opposite, an inference from a particular instance (*beneficial in these circumstances*) to a generalization (*beneficial in general*).

Pursuing this hypothesis further would require testing the inferential patterns involving *good* and *ought* against the relevant *generic>particular* and *particular>generic* control cases. For the moment, we must leave this possibility for future work.[14]

4.2 The prescriptive character of deontic 'ought'

---

[13] Note that (9b) feels substantially more acceptable than (9c):
    (9) c. This dog is violent. Therefore, dogs are violent.
Presumably, it is part of our world knowledge that, sometimes, observing a single instance of a property licenses a generalization, and sometimes it does not. The type of inference in (9b) may even be seen as an inference to the best explanation: the best explanation of a dog having four legs is likely that this is a normal genetic trait of this species; while the best explanation of a dog being violent likely involves specifically its psychological and physical state - such as whether it is afraid, hungry, or infected by rabies.

[14] An additional reason why we take this to be a promising hypothesis, and why we plan to pursue it in future work, is that generalizations from a particular instance are especially strong and frequent when one is generalizing from one's own experience. This is very clear, for instance, in the generic force of statements formed with the pronoun 'one'. As Moltmann (2006, p. 258) points out: 'Suppose the speaker is standing at the entrance, looking at the picture. In that case, an utterance of ['One can see the picture from the entrance'] is appropriate as an immediate expression of the speaker's own experience, while at the same time making a generalization'. Further evidence on generalizations made on the basis of first-person experience has been amply documented in psychology; see e.g. Snyder, Stephan & Rosenfield (1978). What is more, agents are particularly keen on generalizing their positive opinions or experiences; see Gershoff, Mukherjee & Mukhopadhyay (2008). This could explain why the O>G inference is so widely accepted in the cases in which the participants judge the premise to be true. That is to say, if a person believes that they ought to do something in the current circumstances, they will be inclined to believe that everyone ought to do it in similar circumstances; at least, that is what research in psychology appears to show.

A different hypothesis to account for the observed asymmetry is the idea that evaluatives and deontics differ in their illocutionary force: while evaluatives are declarative statements, deontics are a kind of prescriptive statement. Prescriptive statements are those that issue commands, directions or recommendations, among other things. This includes, most notably, imperatives and hortatives:

(10) a. Wear a mask in public!　　　　　(imperative)
　　　b. Let's wash our hands afterwards.　　(hortative)

Contrary to declaratives, prescriptive statements cannot be embedded in certain environments, such as e.g., the antecedent of conditionals (* marks ungrammaticality):

(11) a. * If wear a mask in public! then, p.
　　　b. * If let's wash our hands afterwards, then, p.

However, they are perfectly fine in the consequent of a conditional:

(12) a. If p, then wear a mask in public!
　　　b. If p, then let's wash our hands afterwards.

Relatedly, prescriptive statements appear preferably as the conclusion, rather than as premises, of arguments.[15] Consider the contrast between these two arguments (# indicates that the construction is infelicitous):

(13) a. # Wear a mask in public! Therefore, p.
　　　b. # Let's wash our hands afterwards. Therefore, p.

It sounds odd to follow up a command or recommendation with a declarative consequence. The most natural direction is the opposite, that is, to conclude an argument with a prescription:

(14) a. p. Therefore, wear a mask in public!
　　　b. p. Therefore, let's wash our hands afterwards.

Deontic modals, and in particular, 'ought', are subject to similar—though not identical—distributional constraints. First, 'ought' is typically not found in the antecedent of a conditional (even if it is not outright ungrammatical in that position):[16]

---

[15] This could be related to the idea that often in argumentation, premises tend to contain background not-at-issue information, while conclusions are genuinely asserted and at-issue. We thank an anonymous reviewer for a suggestion along these lines.

[16] A search in the Corpus of Contemporary American English (COCA, https://www.english-corpora.org/coca/) for the combination 'if + ought' with one word in between revealed 300 occurrences, but very few were antecedents of

(15) a. If one ought to wear a mask in public, then *p*.
　　 b. If we ought to wash our hands afterwards, then *p*.

By contrast, it is perfectly common as a consequent:

(16) a. If *p*, then one ought to wear a mask in public.
　　 b. If *p*, then we ought to wash our hands afterwards.

Secondly, 'ought'-statements are most commonly found as conclusions rather than as premises of arguments. Inference patterns in (17) sound significantly less natural than those in (8), which we have indicated by '??'.

(17) a. ?? You ought to wear a mask in public. Therefore, *p*.
　　 b. ?? We ought to wash our hands afterwards. Therefore, *p*.

(18) a. *p*. Therefore, you ought to wear a mask in public.
　　 b. *p*. Therefore, we ought to wash our hands afterwards.

When we do find 'ought' as a premise, the conclusion is most naturally interpreted as an inference to the best explanation. For example, take *p* in (17a) to be 'Masks must be efficient in filtering out potential contaminants'. We then likely understand (17a) as conveying that the efficiency of masks in filtering the contaminants explains why the prescription of wearing masks in public has been issued in the first place.

We think that these observations go some way towards explaining the observed asymmetry between O>G and G>O inferences. If 'ought' appears most naturally as a conclusion, rather than as a premise, then this might be driving some participants' rejection of O>G inferences.

Nevertheless, this hypothesis stands at odds with the overall acceptability of both patterns. To account for that, we would still need to assume that, at some general level of description, deontics and evaluatives make similar argumentative moves, thereby allowing speakers to make inferences in both directions. However, if we combine the hypothesis with the observation that deontics and evaluatives have roughly the same "flavors", are context-sensitive in similar ways, and both belong in a broader class of normative expressions, then we can have our cake and eat it, too. That is, we predict overall acceptance of both inference patterns, but together with the

---

conditionals. Most of them were with 'if' following the verb 'wonder', as in 'I wonder if I ought to… '. In those cases, 'if' is not a conditional operator, but an interrogative marker (equivalent to 'whether'). And even those instances that look like conditionals have quite a different flavor from those that appear in argumentation and that interest us here. Here is one such example:
　　 i. If anyone ought to know what the princess was doing every afternoon, she should.

observation that 'ought' occurs more naturally as a conclusion than as a premise, we also predict that G>O inferences will be seen as more acceptable than O>G inferences

Finally, this hypothesis is, to some extent, compatible with the idea explored in Sect. 4.1 that 'good' and 'ought' are circumstance-sensitive in different ways. Similarly to deontics, imperatives may be seen as circumstance-specific in that they issue commands and requests to be performed "here and now", so to speak. If 'good', by contrast, denotes things to be done not only here and now but in general, then it is easy to see how one can go from 'good' to an imperative, but not the other way around.

5   Conclusion and prospects for future research

Argumentation in public policy abounds with normative expressions. When a new policy, or a change in the existing ones, is proposed, what we typically see are arguments whose conclusions take the form of a deontic statement, or a prescription, namely, what one *ought to do*. In this paper, we have explored arguments with normative force with a special focus on the interaction between two kinds of normative expressions: evaluatives vs. deontics, that is, what is *good* vs. what one *ought* to do. While philosophers in metaethics have been interested in the relationship between the evaluative and the deontic realm, and relatedly, between values, on the one hand, and norms and obligations, on the other, this interest has not reached deep into the philosophy of argumentation. In deontic logic, there have been explorations of the relationship between 'ought-φ' and 'φ-ing is the best option', yet hardly anything has been said on how 'ought-φ' relates to 'φ-ing is good'. Here, we have undertaken first steps in trying to understand this relationship.

Our research brings together theoretical and empirical investigation. The experimental study that we have presented aimed at understanding how people move between what is *good* and what *ought* to be the case, when it comes to argumentation. The study has shown, in a nutshell, that both directions are deemed to be acceptable. Nevertheless, it has also revealed an interesting asymmetry: arguing from what is good to do to what one ought to do is deemed *more* acceptable than the other way round. Interestingly, this asymmetry shows up most clearly in those cases in which participants have no personal opinion regarding the truth of the premises or conclusions.

We have presented two sets of considerations that may be able to explain the observed results. The first relies on the idea that 'good' and 'ought' are sensitive to context in slightly different ways. The latter typically makes reference to the current circumstances, the former purports to apply more generally. This explains why the inference from 'good' to 'ought' is overall more robust, while the converse, although less robust, is still deemed acceptable. The second relies on the idea that evaluative and deontic statements differ in their illocutionary force; the former are assertions, the latter, prescriptions. Since prescriptions are more likely to occur as conclusions,

rather than premises, of an argument, we are again able to explain the observed asymmetry. The two sets of observations are mutually compatible, and do not preclude that there may be other elements relevant to understanding the relationship between the two types of propositions and statements. In particular, there may well be some pragmatic effects that have gone unnoticed. To pursue all these explanations further, we would want to compare the inferences involving 'good' and 'ought' with other types of inferences, and to do so both on a theoretic and an empirical level. We hope to undertake this task in the future.

Acknowledgements

References

Ayer, A. J. (1946). Language, Truth and Logic. London: Penguin.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. arXiv preprint arXiv:1406.5823.

Carr, J. (2017). Deontic Modals. In McPherson, T. & Plunkett, D. (eds.), *The Routledge Handbook of Metaethics*. Routledge. pp. 194-210.

Cresswell, M. J. (1976). The Semantics of Degree. In B. Partee (Ed.), Montague Grammar (pp. 261–292). New York: Academic Press.

Gershoff, A., Mukherjee, A., & Mukhopadhyay, A. (2008). What's Not to Like? Preference Asymmetry in the False Consensus Effect. Journal of Consumer Research 35(1), 119- 125.

Gibbard, A. (1990). Wise Choices, Apt Feelings. Clarendon Press ; Oxford University Press.

Gibbard, A. (2003). Thinking how to live. Harvard University Press.

Hansen, N., & Chemla, E. (2017). Color adjectives, standards, and thresholds: An experimental investigation. Linguistics and Philosophy, 40(3), 239-278.

Hansson, S. O. (2001). The structure of values and norms. Cambridge University Press.

Hare, R. M. (1952). The Language of Morals. Oxford University Press.

Hurka, T. (2014). British ethical theorists from Sidgwick to Ewing. Oxford History of Philosophy.

Kennedy, C. (2007). Vagueness and grammar: The semantics of relative and absolute gradable adjectives. Linguistics and Philosophy, 30(1), 1–45. https://doi.org/10.1007/s10988-006-9008-0

Kennedy, C., & McNally, L. (2005). Scale Structure, Degree Modification, and the Semantics of Gradable Predicates. Language, 81(2), 345–381. https://doi.org/10.1353/lan.2005.0071

Lassiter, D. (2017). Graded Modality : Qualitative and Quantitative Perspectives. OUP Oxford.

Lewis, D. (1973). Counterfactuals. Blackwell.

Moltmann, F. (2006). Generic one, arbitrary PRO, and the first person. Natural Language Semantics, 14(3), 257–281.

Moore, G. E. (1903/1993). Principia Ethica. (T. Baldwin, Ed.). Cambridge University Press.

Morzycki, M. (2012). Adjectival extremeness: Degree modification and contextually restricted scales. Natural Language and Linguistic Theory, 30(2), 567–609. https://doi.org/10.1007/s11049-011-9162-0

Portner, P. and Rubinstein, A. (2016) Extreme and non-extreme deontic modals. In Deontic Modality, ed. Charlow, N. & Chrisman, M. Oxford University Press.

R Core Team. (2019). R Foundation for Statistical Computing. R: A language and environment for statistical computing.

Rowland, R. (2019). The normative and the evaluative: the buck-passing account of value. Oxford University Press.

Russell, B. (1910/2009). 'The elements of ethics'. In Russell, B. Philosophical essays. Routledge.

Scanlon, T. (1998). What we owe to each other. Cambridge, Massachusetts: Belknap Press of Harvard University Press.

Sidgwick, H. (2019). The methods of ethics. Good Press.

Sloman, A. (1970). 'Ought' and 'Better.' Mind, LXXIX(315), 385–394.

Snyder, M., Stephan W., & Rosenfield, D. (1978). Attributional egotism. In J.H. Harvey. W. Ickes & R.F. Kidd (eds.), New directions in attribution research. Vol 2, pp.91-117. Hillsdale, NJ: Lawrence Erlbaum.

Soria Ruiz, A. & Faroldi, F. (2020) Moral adjectives, judge-dependency and holistic multidimensionality. *Inquiry* 64, 1-30.

Soria Ruiz, A. & Stojanovic, I. (2019) On Linguistic Evidence for Expressivism. *Royal Institute of Philosophy Supplement* 86, 155-180.

Stevenson, C. L. (1937). The Emotive Meaning of Ethical Terms. Mind, 46(181), 14–31.

Stojanovic, I. (2016) Expressing aesthetic judgments in context. Inquiry 59, 663-685.

Stojanovic, I. (2017). Metaethical Relativism. In McPherson, T. & Plunkett, D. (eds.), The Routledge Handbook of Metaethics, pp. 119-132. Routledge.

Tappolet, C. (2013). Evaluative vs. Deontic Concepts. In International Encyclopedia of Ethics.

Wedgwood, R. (2017). The value of rationality. Oxford University Press.