Biology & Philosophy (2021) 36:40
https://doi.org/10.1007/s10539-021-09817-y

# Bargaining and descriptive content: prospects for a teleosemantic ethics

Karl Bergman[1]

Received: 22 April 2021 / Accepted: 11 August 2021 / Published online: 21 August 2021

## Abstract

Teleosemantics is the view that mental content depends on etiological function. Moral adaptationism is the view that human morality is an evolved adaptation. Jointly, these two views offer new venues for naturalist metaethics. Several authors have seen, in the conjunction of these views, the promise of assigning naturalistically respectable descriptive content to moral judgments. One such author is Neil Sinclair, who has offered a blueprint for how to conduct teleosemantic metaethics with the help of moral adaptationism. In this paper, I argue that the prospects for assigning descriptive content to moral judgments on the basis of teleosemantics are bad. I develop my argument in dialogue with Sinclair's paper and argue that, although Sinclair's account of the evolution of morality is plausible, the teleosemantic account of the descriptive content of moral judgments which he bases thereon suffers from crucial shortcomings. I argue further that, given some minimal plausible assumptions about the evolution of morality made by Sinclair, no assignment of descriptive content is possible. Contrary to prevailing assumptions, the combination of moral adaptationism and teleosemantics suggests that moral judgments lack descriptive content.

**Keywords** Teleosemantics · Metaethics · Evolution of morality · Sinclair · Descriptivism · Non-cognitivism

Recent decades have seen the development of increasingly sophisticated theories purporting to explain moral thought and behavior on evolutionary grounds (e.g. Alexander 1987; Joyce 2007; Kitcher 2007; Boehm 2012; Tomasello 2015), lending growing support to the view that human morality is a biological adaptation. Call this view "moral adaptationism." It is natural to ask what implications moral adaptationism has for metaethical issues concerning the objectivity and mind-independence of moral facts and properties, as well as for first-order normative issues concerning

✉ Karl Bergman
karl.bergman@filosofi.uu.se

[1] Department of Philosophy, Uppsala University, Uppsala, Sweden

which actions are, in fact, morally right and wrong. This paper concerns a particular answer to these questions: a *teleosemantic* answer.

Moral adaptationism has often been taken to have anti-realist implications in metaethics (Ruse 1995; Street 2006), and most ethicists would probably deny that the view has any implications at all for first-order normative questions. However, an emerging literature of *teleosemantic ethics* goes against the grain on both points. Armed with teleosemantic theories of content, writers like William Harms (2000) and Marc Artiga (2015) have argued, *contra* evolutionary anti-realists in meta-ethics, that moral adaptationism supports meta-ethical realism (though see (Joyce 2001) for criticism). And some writers, like Jacob Ross (2019), have argued that teleosemantics together with evolutionary theories of morality can be used to directly address first-order normative questions. These two projects constitute the two legs of what I call "the teleosemantic program in ethics."

The two legs of this program share a common presupposition: that the combination of teleosemantics and moral adaptationism underwrites *descriptivism* about moral judgments, defined here as the view that moral judgments have descriptive content; that they represent the world as being a certain way. My case against the program will target this shared presupposition. I will argue that, given certain plausible assumptions about the evolution of morality, the teleosemanticist should draw the conclusion that moral judgments lack descriptive content.

Let me briefly summarize the teleosemantic program in ethics. According to teleosemantics, the descriptive content of a representation is determined by its teleological function and its historically normal way of performing that function, which is in turn determined by its evolutionary history. Thus, given that we accept the truth of teleosemantics, and given a sufficiently detailed account of the evolution of moral judgments, we can use teleosemantics to answer questions about the content of moral judgments.

There are at least two such questions. First, we have the metaethical question of whether moral judgments *have* descriptive content in the first place. Second, we have the question of *what those contents are*, given that they exist. To answer the second question, I will say, is to *assign contents* to moral judgments. Since teleosemantics identifies the content of a representation with conditions that have explained its past evolutionary success, it promises to allow us to assign contents in purely naturalistic, non-moral terms.[1] Writers who have proposed content assignments of this kind include Sinclair (2012) and Ross (2019).

Both questions are of independent interest, but especially the second question is also of derivative interest, for if we can find a content assignment that correctly identifies the contents of a moral judgment in non-moral terms, that would seem to have direct normative implications. The descriptive content of a judgment is typically understood to give the conditions under which it is *true* or *correct*. If so, the descriptive content of a *moral* judgment, e.g., a rightness- or wrongness-judgment,

[1] A reader who worries about whether this ambition falls foul of Moore's open question argument should keep in mind that by "content," the teleosemanticist has in mind something like correctness conditions, intensionally rather than hyperintensionally individuated.

gives the conditions under which the evaluated action is correctly judged right or wrong—which is to say, the conditions under which it *is* right or wrong. Moreover, if the contents are specified in non-moral terms, this specification will give non-trivial information about the conditions under which actions are right or wrong (cf. Ross 2019, 271).

For an answer to the content-assignment question to be forthcoming at all, the answer to the first, metaethical question must of course be affirmative. My argument in this paper is intended to defend a negative answer to the first question, thereby showing that no answer to the second question is forthcoming. I will develop my argument through a critical engagement with Sinclair's paper, which I believe constitutes a blueprint for how the teleosemantic program in ethics should be pursued but also suffers from crucial shortcomings. Diagnosing these shortcomings will yield a number of constraints on adequate content assignments. I will then argue that, in fact, no content assignment can meet these constraints, and thus, that moral-adaptationism-plus-teleosemantics implies that moral judgments have no descriptive contents at all. If successful, my argument will thus undermine the teleosemantic program in ethics as a whole.

The paper is organized as follows. Section 1 gives a short primer on teleosemantics. Section 2 introduces a moral-adaptationist theory, the bargaining thesis, adapted from Sinclair. The bargaining thesis comprises the minimal commitments I believe must be presupposed by any plausible moral-adaptationist theory. Section 3 discusses Sinclair's view on the contents of moral judgments. It argues, both that this view is inherently implausible and that it is not in fact consistent with the bargaining thesis. Section 4 expands on the argument of Sect. 3 to argue that the bargaining thesis is inconsistent with *any* attempt to assign descriptive contents to moral judgments. If this is right, and if I am correct that the bargaining thesis constitutes the minimal commitments any moral-adaptationist theory must take on board, this invalidates the teleosemantic program in ethics. Section 5 concludes.

## Teleosemantics primer

Teleosemantics is the view that representational content is determined by etiological (evolutionary) function.

There are a number of subtly different versions of teleosemantics, which construe the determination-relation and its relata in subtly different ways. I will base my discussion on the *consumer-based* version of teleosemantics defended by Ruth Millikan (1984). Similar views have been defended by David Papineau (1984), among others. According to consumer-based teleosemantics, a representation's job is to guide downstream systems ("consumers") by adapting their behavior to the state of the world, and a representation's content consists in normal conditions for successful consumer response. These ideas will be elaborated below.

Other versions of teleosemantics—so called "input-based" or "informational" teleosemantics—instead identify content with the conditions that normally prompt tokening of a representation. Proponents of input-based teleosemantics include Fred Dretske (1986) and Karen Neander (2017). I believe that, for present purposes, the distinctions between these two types of views, consumer-based and input-based, are immaterial. Considerations of space prevent me from arguing the point in detail. However, it is symptomatic that, although Sinclair bases his teleosemantic content assignment on both kinds of view, he reaches the same conclusion in both cases (Sinclair 2012, 653–57). I will make my argument on the basis of consumer-based teleosemantics, on the assumption that my conclusions will admit of ready transposition to an input-based framework. The reader should keep in mind that my conclusions are contingent on the correctness of this assumption.

I turn now to an exposition of consumer-based teleosemantics (in what follows, I will freely shift between the terms "consumer-based teleosemantics," "Millikanian teleosemantics," and simply "teleosemantics.").

First, a note on the term "normal." The term is introduced in the teleosemantic vocabulary by Millikan (who stylizes it "Normal") to describe something that happens according to etiological precedent—i.e., a process or condition such that, on most of those occasions where ancestors have made contributions to the persistence and proliferation of the lineage, the explanation for how they made that contribution involves the fact that they underwent that process or met that condition. Intuitively, "normal" contrasts with "flukey" or "accidental": it distinguishes a regular way of having contributed to evolutionary success from other, accidental ways in which this may sometimes be done (Millikan 1984, 33–34, 43–45).

The obtaining of normal conditions does not guarantee successful function performance—flukes can always intervene. Nor is it *necessary* for successful function performance because, again, flukes can intervene and produce successful outcomes by chance. However, normal conditions are necessary to produce successful outcomes *in a normal way*—i.e., according to the precedent set by evolution in the trait's actual past—and unless circumstances have changed significantly since the evolutionary past, the likelihood of achieving success in the absence of normal conditions will be miniscule, presupposing, as it does, a fluke.

Turn now to representations. On the consumer-based view, representations are states whose function is to serve as intermediaries between coadapted systems, *producers* and *consumers*, where the consumer's normal behavior varies as a function of the state of the world and the consumer relies on the producer to monitor the world, to enable it to adapt its behavior thereto. Representations are the producer's means for so doing.

As a consequence, representations do not come alone, but are always organized into *systems* of representations which are (or could be) produced by the same producer and used by the same consumer(s). If an individual representation is a "sentence," the system is the "language" to which that sentence belongs. Within a system, representations contrast with each other in systematic ways, and these systematic differences correspond to systematic differences in how the representations are "interpreted" by their consumers—i.e., how those consumers normally respond to them (Millikan 1989, 287–88).

The basic idea of consumer-based teleosemantics is that the *normal* explanation for how consumers successfully perform their functions avert to the fact that the representations they consume, and which influence their behavior, bear a certain relation to the state of the world. The obtaining of this relation (or "mapping," as

Millikan likes to call it) is thus a normal condition for successful consumer response (Millikan 1984, 98–100; 1989, 286 ff). (Since it is the representation's function to bring this consumer response about, we can also say, equivalently, that the obtaining of this relation is a normal condition for successful performance of *the representation's own* function). And the *content* of the representation is constituted by those conditions that the world must meet in order for this relation to obtain—conditions that are, therefore, normal conditions for successful function performance of the consumer (and the representation itself).

To use the standard example: honeybees make a figure-eight movement, called a "waggle dance," to guide their nestmates to nectar. Prompted by the dance, the observing bee flies to a certain location, normally given as a function of the dance's length (which translates into distance) and angle to the perpendicular (which translates into direction). If the dance has also been produced normally, there will be nectar at that location. The dance's consumer is that mechanism in the bee brain which, in this manner, translates observed dance into behavior. Presumably, the function of this mechanism is to guide bees to nectar. For this to occur, barring a fluke, the dance must relate ("map") to the world in such a way that there actually is nectar at the coordinates defined by the dance's length and angle. That the dance relates thus to the world, hence that there actually is nectar at the location, is a normal condition for successful function performance. Hence, the teleosemanticist concludes: the dance's content is < there's nectar at so-and-so location > . And indeed, this is a content assignment that makes intuitive sense, lending credibility to the approach.

At this juncture, I must beg the reader's patience and dwell momentarily on some of the finer points of teleosemantic content determination: it will prove relevant to my criticism of Sinclair to follow. My formulation above—content being determined by a "certain relation" normally borne between representation and world—invites the question: which relation? For any representation, there will be many such relations, and all but one must be excluded by supplementary principles to avoid content indeterminacy.

For just one example, consider the relation borne between representation and world just in case the ambient air contains oxygen—for most consumers in the animal world, it will be a normal condition for successful function performance that consumed representations bear *this* relation to the world. Such "constant" relations, which give the same condition for every representation in a system, are excluded by Millikan by the requirement that the content-specifying relation be *differential*: for each member of a system of representations, it places a different condition on the world. The content-defining normal conditions are therefore ones that are *specific* to a representation: if you put a different representation from the same family in its stead, you would get a different set of conditions (Millikan 1989, 287–88).

The above problem and Millikan's solution thereto shows that teleosemantics must not commit itself to including *all* the normal conditions for performance of a representation's function in its content, but only a subset thereof. But which subset? Millikan restricts content-constituting normal conditions to those that are part of the "most proximate" normal explanation for a representation's success (Millikan 1984, 100), a move seemingly designed to exclude normal causal precursors of the most ecologically relevant normal conditions from entering into the content. But this move still leaves us with the set of most proximate normal conditions, from which the theorist must pick a (non-strict) subset.

The problem is related to what has become known as the problem of "vertical indeterminacy." Consider again the waggle dance. As stated, it makes sense to attribute to the waggle dance the content < there's nectar at so-and-so location > , and to be sure, the existence of nectar at the relevant location is *a* normal condition for proper functioning of the waggle dance. But even excluding causal precursors and constant background conditions, there are further factors that have, in the past, contributed to evolutionary success. One is that the nectar has not been poisoned. Another is that no predators have been lying in ambush near the nectar. Etc.

As Karen Neander has pointed out, if we count all these factors into the content of a representation, we get implausibly specific contents (Neander 1995, 126–27). However, the way Neander frames the problem, it is related to a corresponding indeterminacy in what counts as *the function* of the representation and its consumer. If the function is *to find nectar*, it doesn't matter whether or not the nectar is poisoned: the function can be served normally even if it is. On the other hand, if the function is *to nourish the hive*, proper functioning requires the nectar to be non-poisonous.

If we abstract from the question of functional indeterminacy and assume that the representation/consumer has a determinate function, is there then any reason *not* to count *every* (specific, proximate) normal condition for proper performance of that function as part of the representation's content? The answer, of course, ultimately depends on whether a principled distinction can be made among the specific, proximate normal conditions and whether making that distinction would yield explanatory advantages (e.g., better capture intuitive content attributions). Absent any such positive reasons, however, it seems that any restriction on this maximalistic principle would be ad hoc. Moreover, the intuitive idea behind consumer-based teleosemantics is that representations *adapt* consumers to the conditions picked out by their content, thus raising the consumer's chance of success if those conditions obtain; and, of course, the consumer's chance of success is higher the more normal conditions obtain. I therefore propose that the teleosemanticist should embrace the principle that, given a determinate consumer function, the representation's content consists in *all* the (specific, proximate) normal conditions for performing the function. Call this the exhaustion principle.[2] We will return to, and make use of, the exhaustion principle in Sect. 3.

We have now seen how teleosemantics purports to explain how descriptive content is determined. But pretheoretically, we do not believe that all representations have descriptive content—think of directive or imperative representations like desires. Teleosemantics purports to account for this datum as well. Plausibly, a desire's function is to motivate behavior that helps bring about the desire's

[2] David Papineau has defended the view that the content-determining conditions are ones that *guarantee* or *ensure* successful function performance (e.g. Papineau 2017, 99). Papineau's principle entails the exhaustion principle. But Papineau's principle is probably too strong. It should be possible for representations to be true or descriptively correct, yet fail to perform their functions, due to the intervention of flukes or abnormal background conditions.

satisfaction (Millikan 1984, 140). There are no specific conditions that the world must meet in order for this function to be successfully performed in a normal way. Rather, a host of other systems—perception and cognition, but also subsidiary motivational states—help ensure that the desire is only translated into action if *some* circumstances conducive to its success obtain. Just like a true belief can be exploited in the pursuit of many different ends, a desire can be satisfied under many different circumstances. In both cases, what enables this flexibility is the complex adaptivity of a sophisticated psychology such as that of humans (Millikan 1989, 295–96).

Consider, for instance, my present desire to have a sugary snack. The function of this desire, on Millikan's reasonable assumption, is to help bring about its satisfaction condition: my having a sugary snack. Now, there are obviously possible circumstances that make it more likely that this desire will be successful in fulfilling its function: there being a sugary snack in my kitchen, me having sufficient funds to purchase one, etc. And there may be constant background conditions that are necessary for *any* normal desire satisfaction—sufficient ambient oxygen, etc. But it seems unreasonable to suppose that there are any *specific* condition that *must* be met for the desire to be *normally* successful—after all, a complex organism such as myself has presumably evolved to take opportunistic advantage of all sorts of circumstances in pursuit of the satisfaction of its desires. If I possess no snacks, and no funds with which to purchase them, I can always devise another strategy: shoplifting, perhaps?

In the case of desires, we thus presumably cannot specify a differential relation in which the representation must stand to the world in order to be normally successful. Desires therefore lack descriptive contents. I shall argue that, for reasons pertaining to how moral judgments most plausibly perform their functions, the same holds for them as well.

With the basic framework of teleosemantics now on the table, we can turn to its application to moral judgments. The idea behind the teleosemantic program in ethics should by now be clear: given an account of the functions of moral judgments and of how those judgments normally perform those functions, we can use the teleosemantic principles of content determination to assign contents to them. The first step for the teleosemantic ethicist is therefore to produce such an account.

## The bargaining thesis

In this section I will present an account of this kind, which I call "the bargaining thesis." The account is, essentially, that of Sinclair (2012); but since I aim to go on and use this account to cast doubt on the prospects of teleosemantic ethics as a whole, I will aspire to show both that the basic tenets of Sinclair's account are plausible and that they are widely shared among evolutionary ethicists. This will serve to give generality to my argument in Sect. 4. There are obvious limits to how far I can realize my aspirations within the scope of this paper, so the argument will necessarily be of a provisional nature.

Note that I do not argue that these basic tenets are *true*. For my purposes, a weaker claim suffices, namely, that *if* morality is an evolved adaptation, then these tenets are true. For the scope of this paper, I simply assume that morality is an evolved adaptation. If it is not, the teleosemantic program in ethics is in trouble for different reasons.

To serve its purpose within the teleosemantic program in ethics, a moral-adaptationist account must be precise enough to say something about the functions and the normal success-conditions of *specific* (token) moral judgments—e.g., my judgment that killing is wrong. But it must also say something about the overall function of the moral faculty, to which individual moral judgments normally contribute, each in its own way. The task before us is thus twofold: find the overall function of the moral faculty; and find the specific contributions of individual moral judgments to this function.

A word on scope. Sinclair discusses two types of moral judgments, rightness-judgments and wrongness-judgments, and I will follow him in this. I will assume that by "rightness-judgments," Sinclair means moral prescriptions or judgments of moral obligation rather than mere permissibility-judgments. I'll further restrict my attention to atomic judgments of the form Φ IS M, where Φ names an action or action type and M is one of these two moral predicates.

Sinclair, following Alan Gibbard (1990), suggests that the function of the moral faculty is "to produce and sustain mutually beneficial patterns of co-operative action and attitude" among the members of a social group (Sinclair 2012, 652). These are "patterns on behalf of a number of individuals that will benefit each individual as long as the others also (mostly) perform their roles," (650) where *benefit* can be understood either directly in terms of evolutionary fitness or in terms of the satisfaction of desires produced by mechanisms that have tended to track fitness-enhancing outcomes; and is measured relative to "a certain baseline of non-cooperative behavior" (650). In what follows, I will refer to such mutually beneficial patterns of action (and attitude) as "MBPAs." A pattern which is *not* mutually beneficial is a mutually *disadvantageous* pattern of action, or an "MdisPA."

The general idea to which Sinclair gives voice is one with broad currency within the evolutionary ethics literature and which resonates with our commonsense understanding of morality. The idea is that the job of morality is to standardize conduct and to ensure that the standardized conduct thus produced is beneficial for the group and its members: cooperation, mutual aid, abstention from violence, and so on. The challenges facing evolution in producing a mechanism fit to perform this task are well-known: what benefits the group and its members long-term is not always what benefits the individual decision-maker here and now, so mechanisms of both psychological and social nature are needed to transform the utility calculus of the individual decision-maker so as to favor pro-social decisions. The exact conditions under which such mechanisms can evolve, and the exact form they take, are still not fully understood—but for present purposes, a broad picture will suffice.

From this point on, I will assume that Sinclair is correct in identifying the production and maintenance of MBPAs as the function of the moral faculty. But this is rather unspecific. There are many possible patterns of action that qualify as mutually beneficial. *Which* MBPAs is morality in the business of promoting?

Sinclair has little to say about this, but what he does say is crucial to the continued argument: if he is right, and I believe he is, then there are (I will argue) in-principle obstacles to executing the program of teleosemantic ethics. What Sinclair says is that many of the MBPAs which morality is in the business of promoting are *solutions to bargaining problems*.

Bargaining problems arise when there are several mutually incompatible cooperative action patterns available to a group of agents, all of which constitute improvement over the uncooperative baseline, but which distribute the costs and windfalls of cooperation differently among the agents. In such situations, the agents face a conflict of interest, and methods are needed to resolve this conflict and coordinate the group around a single cooperative pattern of action—or it will default back to the uncooperative baseline (Sinclair 2012, 650; cf. Schelling 1960). Typical bargaining problems involve the distribution of costs and benefits in cooperative endeavors. In embarking on such endeavors, there are typically many different ways to assign tasks, distributing the windfall of cooperation, etc., all of which stand to benefit each participant relative to the uncooperative baseline, but some of which stand to benefit some participants more than others.

While some bargaining problems can be solved by genetically hard-wiring the solution in the agents, others are evolutionarily novel and must be solved by the agents themselves. This is where moral judgments come in. According to Sinclair, the moral faculty is precisely the psychological endowment whose function is to enable the solution of novel bargaining problems via the institution and maintenance of MBPAs (Sinclair 2012, 349). By so doing, it benefits the individual, who gets to reap the benefits of these MBPAs, thereby contributing to the individual's, and its own, evolutionary success.

It seems uncontroversial that morality is, to a large extent, concerned with promoting cooperation; and that cooperation is the source of most that is good in human life. It thus seems eminently plausible that if morality is indeed an adaptation, a main driver of its evolution has been its ability to promote cooperation—as most of those writing on the subject have indeed presupposed.[3] This may suffice on its own to grant Sinclair the claim that many of the MBPAs which it is the function of morality to promote will be solutions to bargaining problems. Moreover, if we consider the role of morality in governing the distribution of approval and disapproval, rewards and punishments, etc. (more on which below), bargaining takes on an even more central role—for how to distribute these things is also a problem with multiple solutions, over which agents must bargain.

I conclude that Sinclair's views are plausible: both that the function of the moral faculty is to produce and maintain MBPAs and that at least some of those MBPAs are solutions to bargaining problems.

This is the function of the moral faculty as such. By what mechanisms, then, do individual rightness- and wrongness- judgment normally contribute to the performance of this function?

[3] See, for instance, the references in (Smyth 2017), which also constitutes a rare voice of dissent from this view. Often, the presupposition in question is not framed as an explicit hypothesis, but is implicit in the treatment of the subject, where "evolution of morality" and "evolution of cooperation" are treated as more or less interchangeable.

Sinclair, following Gibbard, identifies two such mechanism (Sinclair 2012, 652). First of all, a moral judgment motivates behavior in accordance with the judgment—meaning that a rightness-judgment motivates the behavior it prescribes, whereas a wrongness-judgment motivates abstention from the behavior it condemns. For example, the judgment that *killing is wrong* motivates the agent to abstain from killing, the judgment that *serving your country is right* motivates the agent to serve her country, and so on.

A reader coming from metaethics may wonder whether the view that moral judgments have *any* sort of motivational function is really consistent with descriptivism about those judgments—i.e., the view that they have descriptive content. Here, it must be noted that the teleosemanticist is not committed to the type of Humean psychology according to which a mental state can be motivational *or* descriptive, but not both. On the contrary, teleosemantics—at least its Millikanian form—countenances the existence of hybrid states. Millikan calls these hybrid states *pushmi-pullyu* representations (Millikan 2005), and Sinclair explicitly endorses the view that moral judgments are hybrids of this sort (Sinclair 2012, 656–57).

I do not believe, however, that this further commitment of Sinclair's is necessary to carry out his program. The claim at issue is that moral judgments *normally* motivate certain behaviors, i.e., that they do so when recapitulating the mechanism by which ancestral judgments have contributed to the persistence of the lineage. This claim is consistent with denial of strong motivational internalism, the view that the subject of a moral judgment is *necessarily* motivated to conform with it (cf. Bedke 2009).[4] Furthermore, it is arguably consistent with denying that moral judgments are themselves motivational states. Suppose moral judgments are purely descriptive states but that in normally functioning individuals, the properties they represent actions as instantiating constitute the targets of some independent motivational system. Suppose further that it is, at least in part, because moral judgments have subserved this independent motivational system by identifying its targets that their lineage has persisted. In that case, we could say that moral judgments are normally motivating, because they have been involved in causing these motivations (by directing the independent motivational system) and have persisted for so doing. Thus, the claim that moral judgments normally motivate certain behaviors is consistent with a range of views on the precise psychological mechanisms involved.

We turn now to the second mechanism whereby moral judgments, according to Sinclair, normally contribute to the function of the moral faculty, the production and maintenance of MBPAs. Moral judgments do not only directly motivate behavior on the part of the subject; they also motivate efforts to make *others* conform to the judgment—perform those actions judged right and abstain from those actions judged wrong.

This is a view with precedents in both the metaethical literature (Gibbard 1990, 40–45; Blackburn 1998; Björnsson and McPherson 2014) and the literature on evolutionary ethics (e.g. Mameli 2013). It is also a view that should resonate with everyday ethical experience.

[4] It is thus consistent with the existence of an *amoralist*, a person who makes moral judgments without being motivated by them. It only entails that such a character would be abnormal.

Given that the moral faculty is in the business of bringing about *group-level* standardization of conduct (MBPAs), we should expect there to be a mechanism whereby the moral faculty pushes the behavior, not only of the subject herself, but also of other group members towards a given pattern of action. Even if an MBPA benefits each individual compared to the uncooperative baseline, there are well-known evolutionary reasons why such benefits are not in themselves sufficient to produce adaptive pressure in favor of cooperative behavior: a cheater stands to benefit even more by exploiting an otherwise cooperative group, partaking of the benefits of cooperation while shouldering none of its costs. To counter this mechanism, the moral faculty needs a way to bring cheaters into line.

These considerations are further strengthened in light of morality's role in solving bargaining problems. If morality is supposed to solve bargaining problems, then there can be morally motivated individuals who, despite acting on the basis of their moral convictions, act so as to disrupt or hinder cooperation, simply because they pursue a different solution than their fellows to some given bargaining problem. Much like cheaters, such individuals would have to be brought into line with the cooperation regime of the others.

The details of how a morally motivated individual can influence the behavior of others is a further issue, and one that is of less relevance here. Sinclair himself emphasizes the role of moral discourse and the explicit discursive avowal of moral judgments (Sinclair 2012, 651–53). Another important vector of influence might be the expression of moral emotions like indignation, pride, and gratitude, as emphasized by Gibbard (1990, 40–45). A third mechanism is sanctions, whether in the form of reciprocity (Trivers 1971; Alexander 1987) or overt punishments (Sripada 2005; Boyd et al. 2010). The last mechanism seems fundamental: it is plausible that both discursive and emotional expression can play their behavior-influencing role only because they function as credible *threats* or *promises* of future concrete sanctions. Sinclair, indeed, mentions the role of discursive avowal as a signal of the willingness to sanction, a role it could not play unless morally motivated agents were in fact normally willing to sanction.

To effectively bring about group-level standardization of conduct, other-directed efforts to influence behavior must be coordinated; they must themselves be a collective effort (a theme emphasized by Sripada (2005) and Boyd et al. (2010), among others). Unless coordinated, such efforts will be costly to the individual and unlikely to have an adequate motivational effect on the target. For present purposes, this observation is of relevance not least because it strengthens Sinclair's contention that morality is in the business of solving bargaining problems; for it means that the collective patterns of conduct which morality normally brings about involve these coordinated sanctions themselves, and the issue of how to coordinate sanctions—whom to punish when, how harshly, etc.—is itself likely to involve bargaining problems.

To summarize: the function of the moral faculty is to produce and maintain MBPAs in a group, at least some of which are solutions to bargaining problems. The mechanism by which individual moral judgments normally contribute to this function is by making the subject herself, as well as other group-members, engage in (rightness-judgments) or abstain from (wrongness-judgment) the targeted behavior.

It is upon this package of views, the bargaining thesis, that my critique of Sinclair and the teleosemantic program in ethics is based. I have shown that Sinclair is committed to them, and I have sought to show that they are plausible enough to lend generality to the criticism of Sect. 4.

## Sinclair on the content of moral judgments

In this section, I present and criticize Sinclair's assignment of contents to moral judgments. Showing where Sinclair goes wrong will help us understand the constraints teleosemantics places on content assignment and prepare the ground for the general criticism of Sect. 4.

As we saw in Sect. 1, (consumer-based) teleosemantics identifies the content of a representation with the normal conditions for successful performance of the representation's function. Given the bargaining thesis, it seems clear that a normal condition for a rightness-judgment to perform its function is that the behavior approved of *is actually part of an MBPA*—otherwise, motivating that behavior could not contribute to the production and maintenance of MBPAs (barring a fluke). Similarly, it seems clear that a normal condition for a wrongness-judgment is that the behavior disapproved of is part of an *MdisPA,* a mutually disadvantageous pattern of behavior—since, by motivating abstention from that behavior, the judgment then guides the group's conduct towards MBPAs.

This, at least, is Sinclair's reasoning. On the basis thereof, he assigns the following descriptive contents to moral judgments. To a judgment of the form Φ IS RIGHT, he assigns the content *Φ is part of a pattern of action that is mutually advantageous*; and to a judgment of the form Φ IS WRONG, he assigns the content *Φ is part of a pattern of action that is mutually disadvantageous* (Sinclair 2012, 656).

We will profit from a compact symbolism with which to represent assignments of contents to judgments. I will use small caps to denote judgments, angle brackets for contents, and ‘$=_{C}$’ for the relation *has the content*. With these conventions, Sinclair's assignment becomes:

**Sinclair's assignment**

Φ IS RIGHT $=_{C}$ <Φ is part of an MBPA>
Φ IS WRONG $=_{C}$ <Φ is part of an MdisPA>

*Sinclair's assignment* assigns determinate, naturalistic descriptive contents to moral judgments. If correct, it thus offers a straightforward vindication of metaethical descriptivism, as I have defined it.

It bears noting, in passing, that Sinclair does not himself consider his view a form of metaethical descriptivism. This is down to a difference between how he and I use the word "descriptivism." On Sinclair's definition, "descriptivism" is the view that "moral content is descriptive content" (Sinclair 2012, 641). Though Sinclair assigns descriptive content to moral judgments, and is therefore a descriptivist in my sense, he denies that the contents assigned by him constitute "moral content," and so he is not a descriptivist in his own sense.

Sinclair's point is a Moorean one: someone can rationally endorse the contents assigned by him, while denying the corresponding moral contents (Sinclair 2012, 658). This argument, however, strikes me as inconclusive. There is no general agreement that the differences tracked by Moorean tests are differences of content, and for the teleosemanticist, whose notion of content is best understood in a Russellian rather than a Fregean way—as being intensionally, rather than hyperintensionally, individuated (see note 1)—it is tempting to simply reject the idea that Moorean tests give any indication about content identity or difference (cf. Millikan 1993). But even so, Sinclair's argument points to the fact that someone could endorse the contents given by Sinclair's assignment—by believing them, let's say—yet fail to make *moral* judgments about the action Φ. Thus, if "descriptivism" is understood as the view that moral judgments have descriptive contents such that endorsing them is *sufficient* to make moral judgments, Sinclair's assignment gives no succor to the descriptivist. Still, it is a form of descriptivism in my sense, and for the reasons given in the introduction, I believe that view to be interesting enough on its own to merit investigation.

I now turn to my criticism of Sinclair's assignment. Showing where Sinclair goes wrong will help us understand the constraints teleosemantics places on content assignment and prepare the ground for the general criticism of Sect. 4.

We can distinguish between the *internal* and the *external* adequacy of a content assignment. Internal adequacy requires the assignment to be consistent with the teleosemantic principles of content determination and with the hypothesized normal functioning of those judgments. External adequacy requires the assignment to be independently plausible as a claim about the content of moral judgments. I will argue that Sinclair's assignment fails along both of these dimensions.

I begin with rightness-judgments and with internal adequacy. Now, I do not dispute that Sinclair has successfully identified *a* normal condition for successful function performance of rightness-judgments, given the bargaining thesis. It seems clear that a rightness-judgment cannot contribute to the production and maintenance of MBPAs in a normal way unless the behavior it prescribes is actually part of an MBPA—since, as established, the normal way for a judgment to do so is to motivate the very behavior it prescribes. But this condition does not exhaust the conditions necessary for normal performance.

To see this, consider that MBPAs are, at least sometimes, solutions to bargaining problems; and bargaining problems have multiple incompatible solutions. Accordingly, there can exist distinct, incompatible behaviors, each of which is part of a different MBPA, where all those MBPAs constitute alternative solutions to the same bargaining problem. Call two behaviors "rival" if they are incompatible in this way; and call two rightness-judgments "rival" if they prescribe rival behaviors. If different group members make rival rightness-judgments at roughly the same time, then clearly, not all of these judgments can succeed: one can succeed only if the others fail.

Let's look at a simple example. Peter and Paul have two tasks to accomplish. One of them must tend to the fire while the other goes out hunting. Both tasks must be accomplished if they are to get food and survive the cold night, but each person prefers the relative safety and comfort of watching the fire. This, then, is a classic bargaining problem. It has two solutions, both of which are MBPAs:

(A) Peter watches the fire, Paul hunts.
(B) Paul watches the fire, Peter hunts.

Consider now the two judgments.

($J_A$) It is right for Peter to watch the fire and for Paul to hunt.
($J_B$) It is right for Paul to watch the fire and for Peter to hunt.

made by Peter and Paul respectively. Both these judgments prescribe conduct in accordance with *some* MBPA. Hence, they are both descriptively correct according to Sinclair's assignment ("Sinclair-correct"). *But they cannot both be successful.* Either Peter and Paul settle on (A), and $J_A$ is successful; or they settle on (B), and $J_B$ is successful; or they both stay by the fire and they starve, and neither judgment is successful.

*Sinclair's assignment*, then, specifies only *some* of the (specific, proximate) normal conditions for a rightness-judgment to fulfill its function. This, I claim, implies that it is internally inadequate. In Sect. 1, I argued for the exhaustion principle: given that we have assigned a determinate function to a representation, the content assigned to that representation by teleosemantics should include *all* the (specific, proximate) normal conditions for fulfilling that function. Sinclair's assignment does not do this. Hence, it is internally inadequate.

My argument for the exhaustion principle in Sect. 1 was, admittedly, tentative. However, as I will argue next, the reason just given for thinking that Sinclair's assignment is internally adequate—that rival judgment can both be Sinclair-correct—is also a reason to think that it is *externally* inadequate. If correct, this serves to further strengthen my case against Sinclair as well as to lend indirect support the exhaustion principle.

Here is the argument. On pretheoretic grounds, we should expect an adequate assignment of contents to moral judgments to have certain implications concerning the relations between moral contents. Among other things, we should expect an assignment to entail the following:

**Univocity**. Pairs of incompatible actions, i.e., actions such that performing one precludes performing the other, cannot both be correctly judged right.

*Univocity* tells us that morality should not turn out to place conflicting demands on us.[5] As we saw above, Sinclair's assignment entails that both (A) and (B) are correctly judged right. But (A) and (B) are incompatible, so the assignment violates

[5] A minority of philosophers believe in the existence of *genuine moral dilemmas*, i.e., the possibility that morality may place genuinely conflicting obligations on us (see Sinnott-Armstrong 1988). If they are right, Univocity is false. But Univocity has strong considerations in its favor, especially in combination with the further but trivial-seeming principle of agglomeration (if I ought to $\Phi$ and I ought to $\Psi$, then I ought to $\Phi$ and $\Psi$). For instance, the two principles jointly entail that we may sometimes be morally required to do what's impossible—perform pairs of mutually incompatible actions—and that would violate the widely-accepted principle that *ought implies can*. Similarly, it is reasonable to assume that an agent is blameworthy if she fails to do what is right (morally required). If an agent were required to do something impossible, she would then be blameworthy whatever she did, which seems unfair.

Univocity. In general: there is nothing stopping two incompatible actions from each being part of its own MBPA, and whenever that's the case, Sinclair's assignment will imply that they are both correctly judged right, in violation of Univocity. If this argument is sound (I will discuss a possible response on Sinclair's behalf below, after having dealt with wrongness-judgments), we can conclude that one and the same feature of Sinclair's assignment leads to failure of both internal and external adequacy. But this state of affairs also gives us reason to believe that if another content assignment were to be made on the basis of the bargaining thesis, one that respected the exhaustion principle, this assignment would also respect Univocity—since such an assignment would have to entail, precisely, that rival judgments could not be jointly correct.

Here, a qualification is in order. The function of a moral judgment, according to the bargaining thesis, is to produce and maintain MBPAs *within a social group*. Thus, the thesis would seem to allow that two rival judgments belonging to members of *different* groups can both be successful. If this is reflected on the side of content, the result will be a kind of moral *relativism*, one according to which Univocity is satisfied so long as it is taken to concern the relations among judgments made by members of the same group, but not when it is taken to concern relations among moral judgments in general.

This need not be a problem. Univocity is most plausible when read as a constraint on the relations between judgments belonging to a single individual. If extended to relations among judgments in general, it contradicts moral relativism. This is a view we have reason not to discount a priori: as noted above, Ross's (2019) version of the teleosemantic program in ethics assigns relativistic contents to moral judgments. For now, suffice to say that Sinclair's assignment violates Univocity regardless of how it is read.

I turn now to Sinclair's account of wrongness-judgments, which suffers from a similar defect. Consider that one and the same action Φ can be part of both an MBPA and an MdisPA. A rightness-judgment promoting Φ and a wrongness-judgment condemning Φ will be rivals in the sense that, if made by members of the same group at roughly the same time, they cannot both be successful. Yet both these judgments would be Sinclair-correct.

Consider the judgments $J_R$ and $J_W$:

($J_R$) It is right for Peter to watch the fire.
($J_W$) It is wrong for Peter to watch the fire.

made by Peter and Paul respectively. Peter's action of watching the fire is part of an MBPA, namely (A). Hence, $J_R$ is Sinclair-correct. But this same action is also part of an MdisPA, namely:

(C) Peter and Paul both watch the fire.

Thus, $J_W$ is *also* Sinclair-correct. But these two judgments are rivals in the aforementioned sense, and hence, at most one of them can succeed.

The problem here is perfectly parallel to the problem for rightness-judgments discussed above. The contents assigned by Sinclair's assignment specifies only *some* of the conditions necessary for a judgment to be normally successful. And in this case, too, the mistake also leads to externally inadequacy. Consider the following criterion on the adequacy of a content assignment:

> **Contrariness**. One and the same action cannot be correctly judged both right and wrong.

*Contrariness* tells us that "right" and "wrong" function as contrary predicates. As in the case of Univocity, Contrariness is most plausible as a condition on relations among a single person's judgments. Understood as a condition on the relations among judgments in general, it contradicts relativism, a view we have reason not to rule out a priori. But Sinclair's assignment violates Contrariness regardless of how it is read.

Before continuing, I want to discuss a possible response that Sinclair could make to the two arguments from internal inadequacy given above. The arguments, Sinclair could say, equivocates on the term "correct." It is true that, according to Sinclair's assignment, rival judgments like $J_A$ and $J_B$, or $J_R$ and $J_W$, can both be *descriptively* correct. But descriptive correctness (i.e., satisfaction of descriptive content) may not be the standard of correctness relevant for Univocity and Contrariness. Sinclair, recall, defends a hybrid view according to which moral judgments are both descriptive and motivational states; and though $J_A$ and $J_B$ ($J_R$ and $J_W$) may be descriptively consistent, on his view, they are not motivationally coherent: they guide the subject in incompatible directions. Now, Univocity and Contrariness track our pre-theoretic intuitions about which judgments can and cannot reasonably be made together, and Sinclair could maintain that these pre-theoretic intuitions are expressions, not of the descriptive inconsistency of the relevant judgments, but of their pragmatic incoherence. If so, his view need not violate either principle.[6]

To this response, I respond in turn by pointing out that on the teleosemantic account, descriptive content is *already* "pragmatic": it is determined by the normal conditions for successful performance of a representation's function. The teleosemantic idea is that a descriptively correct representation *is* one with certain pragmatic virtues, a certain aptness for success (cf. Godfrey-Smith 1994). The response suggested above on behalf of Sinclair presupposes that there are determinants of the success-aptness of a representation *beyond* those that are already implicit in descriptive correctness. The question then becomes why these additional success-conditions should *not* be considered part of the descriptive content. The question is essentially the same as the one I raised in my defense of the exhaustion principle in Sect. 1: is there any principled reason to restrict descriptive content to only some of the conditions for normal success? Sinclair has given us no such reasons, and I can think of none.

Meanwhile, there are positive reasons not to thus restrict descriptive content. By including these extra determinants of success in the descriptive content, we get an account according to which descriptive content *on its own* can explain the intuitive force of Univocity and Contrariness. And that explanation seems, on the whole,

[6] I thank an anonymous referee for suggesting this argument to me.

more parsimonious than if said intuitive force should have to be explained by a combination of descriptive content and motivational force. It is also more general, because it doesn't require us to commit to a hybrid view like Sinclair's: a purely descriptivist, non-hybrid view of moral judgments based on the bargaining thesis could explain Univocity and Contrariness in this way, whereas it would not, arguably, have access to the explanation in terms of motivational coherence suggested above. These observations lend further support to the exhaustion principle.

Even if Sinclair can evade the arguments from external adequacy in the way suggested above, then, his assignment still falls foul of the argument from internal adequacy.

Let us take stock. I have argued that one and the same feature of Sinclair's assignment—its permitting rival judgments to both be descriptively correct—leads to failure of internal adequacy and, at the very least, forces us to jump through some explanatory hoops to secure external adequacy.

The argument does not, however, invalidate the bargaining thesis itself as a basis for assigning contents to moral judgments. In fact, it suggests that if there were a way to give an *internally* adequate content assignment on the basis of the bargaining thesis—one that respected the principle that rival judgments cannot both be descriptively correct—that assignment would also meet Univocity and Contrariness and thus, to that extent, be externally adequate as well. The question, then, is what an internally adequate content assignment would look like.

What we need, it seems, is a content assignment that respects the principle that two rival judgments cannot both be descriptively correct or, equivalently, that the descriptively correct judgments should all *cohere* in the sense of guiding conduct in one single direction. At least, internal adequacy requires that the judgments *of members of the same group, made at (roughly) the same time* cohere, for reasons discussed above. Call this "the coherence requirement." In the next section, I will discuss prospects for finding an assignment that meets this requirement. However, my conclusion will be that, given this requirement, we are unlikely to be able to assign descriptive contents to moral judgments at all.

## Against descriptivism

For a content assignment to meet the coherence requirement, it seems necessary that for each social group at each time, there be some single, unique MBPA (including sets of compatible MPBAs), such that the assignment renders correct 1) all and only those rightness-judgments within the group that prescribe behavior that is part of that MBPA; and 2) all and only those wrongness-judgments within the group that condemn behavior that is incompatible with that MBPA; and which renders incorrect all other such judgments.

Let us call an MBPA with these properties "optimal" for the group. What we need, then, is a content assignment of the following form:

**Optimality**

> Φ IS WRONG $=_C$ <Φ is incompatible with the optimal MBPA for my [the judge's] group at this time>.
> Φ IS RIGHT $=_C$ <Φ is part of the optimal MBPA for my [the judge's] group at this time>.

The reader can assure herself that Optimality meets the coherence requirement, as desired. Furthermore, it satisfies Contrariness and Univocity, at least if read as claims about judgments on the individual or group level. That should not be surprising, since Optimality in effect recapitulates, in new terminology, the traditional moral-realist idea of a single "moral law" governing right behavior, although relativized to groups at times.[7]

I have argued that 1) a content assignment based on the bargaining thesis would have to meet the coherence requirement, and 2) to meet the coherence requirement, a content assignment has to be a version of Optimality. It follows that a content assignment based on the bargaining thesis has to be a version of Optimality. But this conclusion, in itself, does not constitute a satisfying consummation of the program of teleosemantic ethics—for two reasons. First, the word "optimal" is a placeholder. To execute the program, we should want to find an informative description that allows us to *identify* the optimal MBPA. Second, and more importantly, although Optimality sets a constraint on any adequate content assignment based on the bargaining thesis, this is consistent with the possibility that there are, in fact, no adequate content assignments based on the bargaining thesis that meet this constraint. In other words, it may turn out that the bargaining thesis is inconsistent with moral judgments having *any* normal success-conditions of the right content-determining kind (specific, proximate). If that is so, the bargaining thesis would turn out to entail that moral judgments lack descriptive content: it would turn out to imply a form of moral non-cognitivism.

I will argue that it is, in fact, so. No content assignment based on the bargaining thesis can satisfy the constraint set by Optimality. To see why, we must consider what other adequacy constraints such a content assignment would have to meet. First, as has been repeatedly stressed, an assignment must identify conditions that are actually plausible normal conditions for performance of the function of moral judgments, given the bargaining thesis. If those normal conditions are to have the form implied by Optimality, then whatever properties characterize an optimal MBPA must be such as to feature in a regular explanation of how ancestral moral judgments have in fact performed their function. But, I claim, no set of properties of MBPAs can play this role, so there is no adequate content assignment.

For a set of properties to play this role, they would have to be properties such that, in the preponderance of historical cases when moral judgments have successfully performed their functions of producing and maintaining MBPAs, they have

[7] This relativization does not mean, however, that Optimality entails relativism. It is possible that the optimal MBPA turns out to be the same for every group.

succeeded in doing so because the MBPA they promoted instantiated those properties. To identify those properties, we should therefore ask what features of an MBPA facilitate its being successfully promoted; facilitate its being adopted as the MBPA to which the group actually conforms. In a word, we can refer to such properties as the *attractions* of an MBPA, or properties that render the MBPA (relatively more) *attractive* (than other MBPAs or than if it had lacked them). Below, I list four features of MBPAs that I take to be attractions in this sense. There may be more.

(1) *Benefits from conformity* An MBPA can be more attractive than another because the group stands to benefit more if it conforms to it. This concerns *total* benefits yielded but also the way in which benefits and costs are distributed. People might, for instance, be antecedently disposed to favor *fair* solutions to their bargaining problems—but there may also be some preference for solutions that disproportionately favor powerful individuals with strong bargaining positions, who might otherwise choose to opt out of cooperation altogether.

(2) *Ease of conformity* Independently of the benefits yielded *if* the group conforms, MBPAs can vary with respect to how *easy* it is to conform to them. Demanding MBPAs will be less attractive because people will be less likely to conform to them, thus increasing the risk that the group reverts to the uncooperative baseline.

(3) *Champions/opponents* If an MBPA has actual champions in the group, it will be more likely to be adopted than if not. The sheer number of champions may be a factor, but also the relative status and bargaining power of those champions. Conversely, an MBPA with opponents in the group will be less likely to be adopted.

(4) *Past continuity* People will, in general, be more disposed to adopt an MBPA if it stands in continuity with the past, in the sense that past conduct and past ways of solving bargaining problems are (largely) consistent with it. There are several reasons for this. As a matter of general psychology, people are disposed to treat like situations in like manner and to expect the same of others. This tendency to defer to precedent is also a feature, I take it, of moral psychology specifically. There are reasons to expect this to be the case: such a tendency would facilitate the solving of novel bargaining problems, both by reducing the incidence of costly relitigation and by enabling people to learn, from experience, which solutions others are most likely to favor. Moreover, real bargaining problems are seldom temporally discrete, one-off occurrences, but are temporally extended and interlocking, and to suddenly change patterns therefore risks disrupting ongoing cooperative projects and cause mismatched expectations among the agents.

Note that several MBPAs, even rival ones, can all be consistent with past conduct. There is never just one way of generalizing a pattern from a finite set of precedents, and novel situations may arise which require a choice between rival actions belonging to distinct MBPAs, all of which are consistent with past conduct. In such situations, generalization is underdetermined.

I hope the reader will share my assessment that a moral judgment is more likely to be successful if the behavior it promotes is part of an MBPA that possesses one of the attractions listed above than if not, all else being equal. But to get the requisite notion of optimality we need more, for there can only be one optimal MBPA for each group and time. Thus, it seems, we need some way of weighing these attractions (as well as any additional ones that I may have omitted from my list) into a single metric and identify optimality as the property of scoring the highest on this metric.

But, as advertised, I doubt that this can be done. The reason for my skepticism has to do with the essentially contested nature of bargaining, which introduces a measure of irreducible contingency into the determination of success and failure for moral judgments, which is in turn incompatible with the existence of a consistent evolutionary precedent that could determine normal conditions for moral judgments.

I will elaborate. “Normal,” as we’ve seen, contrasts with “flukey” or “accidental”: there must be some *regularity* to the mechanism whereby ancestral would-be representations have used normal conditions to secure their evolutionary success, the persistence and proliferation of their lineage. If the posited property of optimality figures in the normal conditions for successful performance of the function of moral judgments, and if this property is to be identified by a weighing-together of attractions, then there must be some way of weighing these attractions together into a metric of optimality such that there is a reliable, regular mechanism that has correlated the ancestral evolutionary success of moral judgments with the fact that the actions they promote are part of the MBPA that scores highest on this metric.

I don’t think this weighing exists, because I do not think that there is sufficient regularity in the manner by which some moral judgments have prevailed over others over the course of ancestral history. Attractiveness is multidimensional, and different attractions are wont to attract different group members more or less strongly. There could thus be situations where no single candidate MBPA is the most attractive to all group members—where different members are attracted to different MBPAs. In such situations, the choice between the rival MBPAs will have to be resolved by a process of negotiation, sensitive to many contingencies including subtle differences in the bargaining position and skill of the participants, information asymmetries in the group, complex Machiavellian machinations, and so on. Can we expect there to be a regular, uniform mechanism whereby determinate features of MBPAs correlate with success across such conflicts? That seems dubitable.

Suppose, to the contrary, that we do have some weighing of attractions *W* which gives us, for any group at any time, the optimal MBPA. Suppose now that in our group at this time, I favor behavior that is part of *A* and you favor rival behavior that is part of some rival MBPA *B*, and suppose that among all MBPAs *A* scores highest on *W*, making it the optimal one. Under normal conditions my judgment should succeed, thus necessitating the failure of yours. But suppose that *B* is also very attractive; perhaps promising great benefits, enjoying the support of powerful group members, and standing in conformity with past precedent. It just happens not to rank quite as high according to *W* as *A* does.

Now suppose that, due to the vicissitudes of social negotiation, your judgment is in fact the successful one, influencing behavior in the group and contributing to the production and maintenance of MBPAs. On the assumptions made, we would have to say that what has occurred is an abnormality, a fluke. But that conclusion seems absurd. Surely, what has occurred here is no more flukey than what happens any time two attractive, but rival, moral judgments duke it out. After all, had my judgment succeeded instead, it would have been due to the same kind of contingency-laden process of social negotiation that led yours to victory. So *W* cannot give the normal conditions for proper performance of a moral judgment's function, and that is true no matter which candidate optimality-metric is substituted for *W*. All that is needed for this conclusion is the assumption that it is at least *possible* to get situations like the one described, where *W* favors an MBPA which has close rivals.

Admittedly, my argument is an argument from ignorance, based on my own inability to imagine what sort of mechanism could underlie a regular correlation between success and features of the social bargaining situation. Perhaps others will be more imaginative than me—if so, I welcome their correctives. If nothing else, my argument can be read as a challenge to the teleosemantic ethicist: if you want to assign contents on the basis of the bargaining thesis, you must produce a metric of optimality with the requisite explanatory properties. You will then have to explain how the seeming contingency and chanciness of bargaining outcomes are really manifestations of an underlying pattern.

If my argument is sound, the conclusion is that no content assignment can satisfy the constraint of optimality while also giving plausible normal conditions for the successful performance of a moral judgment's function. Thus, there are no adequate content assignments for moral judgments. Thus, the teleosemantic ethicist must conclude that moral judgments lack descriptive content. At most, we can make broad-based observations about what kinds of qualities tend to make a moral judgment apt for success, but we won't be able to systematize them in the way requisite to assign descriptive contents.

What emerges here is a picture of moral judgments, not so much as attempts to represent an antecedently given reality, but as rivaling bids in a battle for influence over people's conduct, with no clear evolutionary precedent to adjudicate the legitimacy of the eventual victor. In this battle, the moral faculty does not seek antecedently correct answers but *chooses a strategy* in the hope that it will, in the hustle and bustle of social negotiation, turn out to be successful. This picture, of course, is fundamentally incompatible with moral cognitivism.

## Conclusions

If my argument above is correct, the combination of teleosemantics and moral adaptationism will not support moral descriptivism, and it will not yield any normative-ethical conclusions.

The argument, of course, makes a number of substantial assumptions. Not least, it presupposes the bargaining thesis. Though I have done my best within the given constraints to argue that this thesis should be accepted by any proponent of the teleosemantic program in ethics, a defender of that program has the option of promoting an alternative moral-adaptationist account, one that might possibly support a different conclusion as to the status of moral descriptivism.

Further interpretation of my result is likely to vary with the reader's preexisting metaethical commitments, on the principle that "one man's *modus ponens* is another's *modus tollens*." If the reader is already a convinced metaethical descriptivist, she is unlikely to have her mind changed by the fact that the combination of teleosemantics and moral adaptationism points in a different direction. Rather, she is likely to see this as simply a reason to doubt the conjunction of these theoretical frameworks. A reader already sympathetic towards non-cognitivism, however, may feel ever so slightly vindicated by the result. For my own part, I do not lay claim to either a *modus ponens* or a *modus tollens* interpretation of my result. I'm happy if I have demonstrated an incompatibility among these views.

**Funding** Open access funding provided by Uppsala University. Not applicable.

**Declarations**

**Conflict of interest** Nothing to declare.

## References

Alexander Richard D (1987) The Biology of Moral Systems. A. de Gruyter, Hawthorne, N.Y.

Artiga M (2015) Rescuing tracking theories of morality. Philos Stud 172(12):3357–3374. https://doi.org/10.1007/s11098-015-0473-6

Bedke MS (2009) Moral judgment purposivism: saving internalism from amoralism. Philos Stud 144:189–209. https://doi.org/10.1007/s11098-008-9205-5

Björnsson G, McPherson T (2014) Moral attitudes for non-cognitivists: solving the specification problem. Mind 123(489):1–38. https://doi.org/10.1093/mind/fzu031

Blackburn Simon (1998) Ruling Passions: A Theory of Practical Reasoning. Oxford University Press, Oxford

Boehm C (2012) Moral Origins: The Evolution of Virtue, Altruism, and Shame. Basic Books, New York

Boyd R, Gintis H, Bowles S (2010) Coordinated punishment of defectors sustains cooperation and can proliferate when rare. Science 328(5978):617–620. https://doi.org/10.1126/science.1183665

Dretske F (1986) Misrepresentation. In: Bogdan R (ed) Belief: Form, Content, and Function. Oxford University Press, Oxford, pp 17–36

Gibbard A (1990) Wise Choices, Apt Feelings: A Theory of Normative Judgment. Clarendon Press, Oxford

Godfrey-Smith P (1994) A Continuum of Semantic Optimism. In: Stich S, Warfield TA (eds) Mental Representation: A Reader. Blackwell, Oxford, pp 259–577

Harms W (2000) Adaptation and moral realism. Biol Philos 15:699–712. https://doi.org/10.1023/A:1006661726993

Joyce R (2001) Moral realism and teleosemantics. Biol Philos 16(5):723–731. https://doi.org/10.1023/A:1012280429613

Joyce Richard (2007) The Evolution of Morality. 1. MIT Press paperback ed. Life and Mind. Cambridge, Mass: The MIT Press

Kitcher P (2007) Biology and ethics. Oxford University Press

Mameli M (2013) Meat made us moral: a hypothesis on the nature and evolution of moral judgment. Biol Philos 28(6):903–931. https://doi.org/10.1007/s10539-013-9401-3

Millikan RG (1984) Language, Thought, and Other Biological Categories: New Foundations for Realism. Mass MIT Press, A Bradford Book

Millikan RG (1989) Biosemantics. J Philos 86(6):281–97

Millikan RG (1993) White Queen Psychology; or, the Last Myth of the Given. White Queen Psychology and Other Essays for Alice. Mass MIT Press, Cambridge, pp 279–363

Millikan RG (2005) Pushmi-Pullyu Representations. Language: A Biological Model. Oxford University Press, Oxford, pp 166–86

Neander K (1995) Misrepresenting & malfunctioning. Philos Stud 79:109–141. https://doi.org/10.1007/BF00989706

Neander K (2017) A mark of the mental: in defense of informational teleosemantics. Life and mind: philosophical issues in biology and psychology. The MIT Press, Cambridge, Massachusetts London, England

Papineau D (1984) Representation and explanation. Philo Sci 51(4):550–572. https://doi.org/10.1086/289205

David Papineau (2017) Teleosemantics. In: Livingstone Smith David (ed) How Biology Shapes Philosophy: New Foundations for Naturalism. Cambridge University Press, pp 95–120

Ross J (2019) Teleosemantics and Normative Ethics. In: Timmons M (ed) Oxford Studies in Normative Ethics, vol 9. Oxford University Press, Oxford, pp 271–294

Michael Ruse (1995) Evolutionary Naturalism: Selected Essays. Routledge, New York

Schelling TC (1960) The Strategy of Conflict. Oxford University Press, New York

Sinclair N (2012) Metaethics, teleosemantics and the function of moral judgements. Biol Philos 27(5):639–662. https://doi.org/10.1007/s10539-012-9316-4

Walter Sinnott-Armstrong (1988) Moral Dilemmas. Philosophical Theory. Blackwell, USA

Smyth N (2017) The function of morality. Philos Stud 174:1127–1144. https://doi.org/10.1007/sl1098-016-0746-8

Sripada CS (2005) Punishment and the strategic structure of moral systems. Biol Philos 20(4):767–789. https://doi.org/10.1007/s10539-004-5155-2

Sharon Street (2006) A darwinian dilemma for realist theories of value. Philos Stud 127(1):109–66

Tomasello M (2015) A Natural History of Human Morality. Harvard University Press, Cambridge, Massachusetts

Trivers RL (1971) The evolution of reciprocal altruism. Quart Rev Biol 46(1):35–57. https://doi.org/10.1086/406755