Every science needs a framework of assumptions to epistemically access what it aims to explain. For the moral sciences, the leading candidate is the *morality-as-cooperation* (MAC) framework, which claims that morality functions to solve the problems of cooperation. It has two basic problems, though. The first is that it neglects the non-cooperative elements of morality, which includes normative constraints on our relations with non-cooperators (e.g., severely disabled humans, non-human animals) and ourselves. The second problem is that it neglects the possibility of error in cooperative schemes: it grounds morality in reinforcing cooperative schemes—regardless of whether they are moral or immoral. In this paper, I develop a framework for the moral sciences that addresses these problems. For one, it solves the first problem simply by limiting its claims about morality to the context of cooperation. For another, it solves the second problem by modifying Boyd’s (1988) semantic approach to individuate moral rightness as the referent of the moral term ‘right’. I argue from recent evidence about moral judgment-making that in the context of cooperation, the referent of the moral term ‘right’ and hence, moral rightness itself consists in the property of cooperative functionality for only the broadest recruiting system (which excludes all immoral cooperative schemes). Finally, I argue that this new framework earns its explanatory keep: it appears to be indispensable to the best explanation of moral judgment-making—in cooperation games in the lab and out in the field.

**Keywords:** cooperation; function; normativity; moral progress; moral judgment-making

Every science needs a framework of assumptions to epistemically access what it aims to explain (Duhem, 1914; Quine, 1951). For the moral sciences, the leading candidate is the *morality-as-cooperation* (MAC) framework, which claims that morality functions to solve the problems of cooperation (Curry, 2016; Curry et al., 2019; c.f., Alexander, 1987; Gibbard, 1990; Hausman & McPherson, 2006; de Waal, 2006; Wong, 2006; Prinz, 2007; Kitcher, 2011; Rai & Fiske, 2011; Haidt, 2012; Greene, 2013; Tomasello, 2015). The notion of function here is normative: it admits of dysfunction. Thus, some moral kinds count as *functional* in virtue of successfully solving the problems of cooperation: e.g., moral judgment-making can coordinate with ostracism to solve the tragedy of the commons (Feinberg et al., 2014). And other moral kinds count as *dysfunctional* in virtue of unsuccessfully solving the problems of cooperation: e.g., an act of retribution can fail to solve the tragedy of the commons by creating costly cycles of retribution (Simpson et al., 2017).

---

1 Draft, 16/09/21. This paper has not been peer reviewed. Please do not copy or cite without author's permission.
There are two basic types of objections to the MAC framework. The first is an objection from insufficient explanation: there is more to morality than what the MAC framework aims to help us explain. Buchanan (2020) refers to this as the cooperation dogma: the MAC framework stipulates that the domain of morality is limited to the domain of cooperation. But this appears to be false because most moral codes in actual human practice contain what seem to be non-cooperative elements: e.g., normative constraints on our relations with non-cooperators (e.g., severely disabled humans, non-human animals) and ourselves (e.g., self-care, ascetic practices). Since the purpose of the MAC framework is to provide the moral sciences with epistemic access to what they aim to explain, it fails when it doesn’t provide the moral sciences with epistemic access to these non-cooperative elements of morality. The easiest response to this objection is just to restrict the scope of the MAC framework to the cooperative elements of morality. This way, functional cooperation counts as morally right, but other, non-cooperative functions can count as morally right too. I’ll refer to this new framework as the functional-cooperation-as-morally-right (FCAMR) framework.

The second basic type of objection is an objection from incorrect explanation: the assumption that the MAC/FCAMR framework makes to help us explain morality (or just its cooperative elements) is false. A common version of this objection claims that the MAC framework entails that functional cooperation is morally right in the context of immoral cooperative schemes. For example, it entails that it would be morally right to turn Jewish refugees in to the Nazi authorities since this supports the cooperative scheme in Nazi Germany. But this seems false. After all, most would agree that it is morally right to functionally cooperate only in moral cooperative schemes (to support them) and instead to dysfunctionally cooperate in immoral schemes (to undermine them). This objection is grounded in an appeal to moral intuitions, but it can also be grounded in the need for explanation. After all, any behaviour can simultaneously support and undermine different cooperative schemes: e.g., turning in Jewish refugees supports the cooperative scheme in Nazi Germany but undermines the cooperative scheme of Jewish and non-Jewish dissidents. This raises an explanatory problem: how do we classify a behaviour as cooperatively functional or dysfunctional when it participates in so many conflicting cooperative schemes?

In this paper, my goal is to argue that in the context of cooperation, the explanatory power of the FCAMR framework is maximized by classifying behaviours that are (a) cooperatively functional for the broadest system that recruits them (e.g., the global society, not Nazi Germany) as morally right and (b) cooperatively dysfunctional for the broadest recruiting system as morally wrong. In §1, I’ll argue that the MAC/FCAMR framework is committed to grounding morality in cooperative functionality. In §2, I’ll explain how cooperative functions relate to cooperative schemes. In §3, I’ll argue that the function of moral judgment-making is to identify the behaviours that count as cooperatively functional for the team in the context of cooperation games. In §4, I’ll generalize by arguing that the function of moral judgment-making in the broader context of cooperation is to identify the behaviours that count as cooperatively functional for the broadest system that recruits them. I’ll argue that this indicates that in the context of cooperation, the property of moral rightness consists in the property of cooperative functionality vis-à-vis the broadest recruiting system. In §5, I defend this FCAMR framework by applying it to develop a proto-explanation of moral judgment-making in the field. In §6, I conclude by considering the future of the FCAMR framework.
§1. Functional Cooperation

Moral scientists shy away from using normative language to express the morality-as-cooperation (MAC) framework. One way that they do this is by using the terms ‘problems of cooperation’ and ‘solutions to the problems of cooperation’, which seem to suggest that (a) cooperation creates problems that anything can either solve or not-solve and (b) solving and not-solving are non-normative relations between the problems of cooperation and their solutions. Their reluctance is understandable: normativity—especially moral normativity—is controversial, so using normative language might invite too much controversy. Empirical results that appear to rely on controversial moral assumptions would never make it past review! Still, avoiding normative language only causes confusion. In this section, I’ll argue that the problems of cooperation and their solutions are both fundamentally grounded in functional cooperation, which is a normative kind.

Let’s start by considering a problem of cooperation known as the tragedy of the commons (Hardin, 1968). This problem arises whenever there is a finite public good (the commons), which includes any resource that (a) confers benefits on anyone consumes it, (b) is finite, (c) is renewable above a certain threshold (i.e., its carrying capacity), (d) is accessible to the public, and (e) is depleted whenever it is consumed. A simple example is the megafauna during the last ice age, who provided a common food source to early human populations. The problem is that each group of consumers has an incentive to consume more than their share of the surplus capacity of the commons (i.e., the amount that exceeds the carry capacity). After all, each group on its own can’t consume enough of the commons to deplete it below its carrying capacity but they can consume enough to endow themselves with more than their competitors. Since every group has this incentive, though, they will collectively deplete the commons below its carrying capacity and thus collectively eliminate it—resulting in the tragedy of the commons. This appears to be what caused the extinction of most megafauna by the end of the last ice age.

Behavioural game theorists reproduce the tragedy of the commons in the controlled context of the laboratory using public goods games, which are defined by the following basic set of rules:

**Public Goods Game.** Any public goods game involves multiple rounds that each consist in this series of three steps:

1. A team of players contribute real money into a common pot;
2. Experimenters multiply the real money in the common pot by a certain factor; and
3. Experimenters divide the resulting amount of money equally among the entire team of players (even among those who didn’t contribute).

Let’s consider a public goods game, where there are five players, each player starts with $1, money in the common pot is doubled in each round, and there are 8 rounds (Figure 1). Immediately, we face a problem of generalization. In each round, each player faces a choice to contribute one of the following 101 amounts: contributing $0.00 or $0.01 or $0.02… or $0.99 or $1.00. In each round, the team of five players thereby faces a choice to contribute in one of $101^5$ ways—one for each series of possible contributions by the five players. This is a problem because it’s practically impossible to explain all $101^5$ ways that the game could go. Therefore, a solution to this problem

---

2 My reason for choosing this problem of cooperation is that there are some interesting empirical results that use this game, which we’ll consider in §2.
must identify a few ways that the game could go that are such that explaining them is necessary and sufficient to achieve a general understanding of the gameplay. Neander (2017) argues that functional standards provide a solution to this problem: they classify as functional only the tiny minority of activities whose explanations are necessary and sufficient for a general understanding of any system, and they classify as dysfunctional the vast majority of activities whose explanations are unnecessary and insufficient for a general understanding of any system.

How do we identify which activities are functional and dysfunctional? One popular answer is that we should look at the evolutionary history and classify the activities that are selected for as functional and the activities that are selected against as dysfunctional (e.g., Wright, 1971; Millikan, 1989; Neander, 1991; Garson, 2016). However, there is a serious problem with this answer: functions often change in new contexts and modern society represents an abrupt departure from environments in our evolutionary past, such that we can’t reliably infer the functions of our activities in modern contexts from their functions in historical contexts (Smyth, 2017).³ A better answer, I think, is that we should accept the functional classification that provides the best solution to the problem of generalization. This seems to be what behavioural game theorists do: they consider which specific activities are such that it’s necessary to explain them to achieve the best general understanding of gameplay and they classify those activities as functional. This way, they remain completely neutral on controversial theories about the metaphysics of functions.

Using this method, behavioural game theorists have found that these three functional standards seem to provide the best general explanation of the public goods game:⁴ ⁵

**Collective function.** A team’s behaviour counts as collectively functional if and only if (and because) it maximizes the amount of money that the team earns. Otherwise, it counts as collectively dysfunctional.

**Cooperative function.** A player’s behaviour counts as cooperatively functional if and only if (and because) it maximizes the amount of money that the team earns, because that would achieve the team’s collective function. Otherwise, it counts as cooperatively dysfunctional.

**Selfish function.** A player’s behaviour counts as selfishly functional if and only if (and because) it maximizes the amount of money that the player earns. Otherwise, it counts as selfishly dysfunctional.

For the team to achieve its collective function in our example, it must contribute $5 from all its players to the common pot in each round. By the end of the game, it will have earned \(5 \times 2^8 = 1280\). For the team to achieve its collective function, then, it’s necessary for each player to achieve their cooperative function for the team: each player must contribute all of their holdings in each round (e.g., $1 in the first round), so that the team has contributed all of its holdings (e.g.,

³ In fact, most sciences draw functional inferences without appealing to evidence from evolutionary history (Godfrey-Smith, 1993; Kitcher, 1993; Amundson & Lauder, 1994; Walsh, 1996; Neander, 2016). More on this in §6.
⁴ We’ll learn why precisely three functions are necessary to understand what’s interesting about the game play in §2.
⁵ Behavioural economists usually use the language of ‘interests’, rather than the language of ‘function’, for historical reasons and possibly to avoid confusion between ‘function’ in the normative sense and ‘function’ in the mathematical sense. Hence, they would refer to these three functions as self-interests, cooperative interests, and group interests.
In the first round, each player can know that every other player will contribute $1.00 to the common pot. Then the selfish player can maximize their earnings by contributing $0 and still earning 20% of the common pot: 20% × ($1 × 4) × 2 = $1.60. At the end of the first round, the selfish player will hold $2.60 (the highest value that’s possible for them to hold in the first round) while the cooperative players will each hold $1.60. By comparison, suppose that the one player can know that every other player will defect (i.e., contribute $0). Then they would earn back 2/5 of whatever they contribute to the common pot (after it’s doubled and then divided among the players). Thus, they still minimize their losses by contributing $0. In any case, the dominant strategy for each selfish player is to contribute nothing to the common pot.
The competition between selfish and cooperative functions to contribute nothing and everything, respectively, usually results in the tragedy of the commons. Players usually demonstrate a stronger disposition to selfish (vs. cooperative) functionality in the first round and this disposition grows progressively stronger until the final round. For example, Simpson et al. (2017) report that the average contribution starts at 40% in Round 1 and drops to 20% in Round 8. At an average contribution of 30%, the team would hold $5 \times 1.38 = $40.79. To achieve a general understanding of the tragedy of the commons, though, it’s unnecessary to explain contributions of intermediate amounts, from $0.01 to $0.99. Instead, it’s only necessary to explain contributions of boundary amounts ($0 and $1): by Round 8, a team of perfect cooperators (who contribute $1) would hold $1280 and a team of perfect defectors (who contribute $0) would hold $5. The result is a tragedy: perfect defectors who aim to maximize their earnings in each round achieve a much worse outcome by all three functional standards than perfect cooperators.

The MAC framework couches this explanation in terms of “problems of cooperation” and their “solutions”. But these terms can only be defined in terms of “cooperative functionality”. In the public goods game, the problem of cooperation is the need for the team to find a mechanism that increases functional cooperation amongst players. Then the solution to this problem of cooperation includes any mechanism that is able to satisfy the team’s need to increase cooperative functionality amongst players. Otherwise, if we extend these problems and solutions to any cooperation, then the problems and their solutions are trivial: any contribution to the common pot (even $0) counts as cooperation, such that anything that has an effect on player behaviour would count as increasing some type of cooperation and so, as a solution to the problem of cooperation. Thus, the problems of cooperation and their solutions are ultimately grounded in the function of cooperation: they are problems of functional cooperation and their solutions. And so, the MAC framework is committed to the normative claim that morality is fundamentally grounded in cooperative functionality.

§2. Cooperative Schemes

Cooperation games provide an excellent opportunity for scientists to investigate cooperation in the controlled conditions of the lab but public goods games and most other cooperation games have a critical limitation: they create only one team (the set of all players). Thus, any behaviour can only count as cooperatively functional in one way: vis-à-vis the team. In the field, though, agents are embedded in multiple cooperative schemes, which are often quite incompatible: e.g., those of Nazi Germany and its dissidents. Thus, any behaviour can count as cooperatively functional in many ways: one for each cooperative scheme. This raises important questions for the MAC framework: is morality grounded in cooperative functionality vis-à-vis all cooperative schemes? Or only some of them? And if so, which ones? Let’s build up to answering these questions in later sections by answering an easier question in this section: how do cooperative schemes relate with each other?

---

6 The fact that most contributions are such that explaining them is unnecessary to achieving a general understanding of the public goods games is the reason why our functional classification (which classifies only $0 and $1 contributions as functional) counts as a solution to Neander’s (2017) problem of generalization.

7 It would be incorrect to say that they have achieved greater selfish functionality in the long-run, though, since selfish functionality concerns the behaviours that maximize earnings in each round—not the consequence of maximizing earnings after some set of rounds. Functionalism should not be conflated with consequentialism.
To answer this question, let’s classify the selfish, cooperative, and collective functions under more general categories, which we can then use to taxonomize cooperation schemes. The first point to note is that selfish and collective functions appear to fall under a common category that excludes cooperative functions. After all, selfish functions are functions that players have just in virtue of being players and collective functions are functions that teams have just in virtue of being teams. I propose that we classify these functions as intrinsic:

**Intrinsic function.** A function F of a system S is intrinsic if and only if (a) S can do F and (b) explaining F is necessary for achieving a general understanding of S.
The players minimizing their contributions to the common pot and so maximizing their earnings has an intrinsic function for the players: explaining this activity is necessary for achieving a general understanding of the players qua players. I referred to this as a selfish function, but it’s really just a player-level intrinsic function. Likewise, the team maximizing its earnings (i.e., the total earnings of all the players) has an intrinsic function for the team: explaining this activity is necessary for achieving a general understanding of the entire team qua team. I referred to this as a collective function, but it’s really just a team-level intrinsic function.

However, cooperative functions don’t count as intrinsic functions. After all, they are functions that players have not in virtue of being players but in virtue of being members of teams (an extrinsic relation with another entity, the team), who causally contribute to the intrinsic functioning of the teams. So, I propose that we classify these functions as instrumental:

**Instrumental function.** A function F of a system S for a system S* is instrumental if and only if (a) S can do F, (b) S doing F contributes to causing S* to do its intrinsic function F*, and (c) explaining S doing F is necessary for achieving a general understanding of the etiology of S* doing F*.

The players maximizing their contributions to the common pot has an instrumental function for the team: it is the functional cause of the team maximizing the sum of the earnings of the players and achieving its function. That is, explaining the players all contributing 100% to the common pot is necessary for achieving a general understanding of the etiology of the team achieving its function. I referred to this as a cooperative function, but it’s really just a player-to-team-level instrumental function. Intrinsic and instrumental functions ground the distinction between systemic functionality and systemic success:

**Systemic functionality.** A system S achieves systemic functionality if and only if (and because) S achieves its intrinsic function in virtue of its components achieving all of their instrumental functions.

**Systemic success.** A system S achieves systemic success if and only if (and because) S achieves its intrinsic function, regardless of whether all of its components achieved their instrumental functions.

After all, a system can achieve systemic success without achieving systemic functionality due to (a) redundancies among their internal components or (b) lucky compensations from the external environment (Figure 2). Since explaining these conditions for systemic success is unnecessary for achieving a general understanding of a system, they count as dysfunctional—regardless of the fact

---

8 My distinction between intrinsic and instrumental functions is related to the distinction between etiological functions (e.g., Wright, 1971; Millikan, 1989; Neander, 1991) and organizational functions (e.g., Cummins, 1975; Kitcher, 1993; Mossio et al. 2009) insofar as systems have etiological functions in virtue of their own properties and systems have organizational functions in virtue of their relations with broader systems that recruit them. My distinction is different, though, insofar as I distinguish these functions for serving roles in different generalizations (Neander, 2017) whereas other accounts distinguish these functions by giving them various different reductive analyses.

9 Public goods games are a special case where the players achieving their instrumental functions is necessary for the team’s success.
that they cause systemic success. We didn’t find such cases in the public goods game because it is a special case where the players achieving their instrumental functions is necessary for the team’s success. Still, we have all of the concepts that we need to map the functional structure of the public goods game (Figure 2).

Now let’s apply our taxonomy of functions to map out more complex functional structures in the field. For a salient contemporary example, let’s taxonomize the alt-right (vs. e.g., Nazi Germany). It has a specific intrinsic function qua the alt-right: promoting various forms of racial, sexual, gender, political, and ideological discrimination. These are some of its activities that are such that explaining them is necessary to achieve a general understanding of the alt-right itself (its sub-movements at a time and its changes over time). By comparison, the alt-right qua online social movement has more general intrinsic functions: e.g., recruiting and coordinating weak online entities to overpower powerful offline entities with entrenched interests. These are only a few of its activities that are such that explaining them is necessary and sufficient to achieve a general understanding of any online social movement, from the alt-right to Black Lives Matter (BLM). To back such a broad generalization, these functional activities must be much more general than the functional activities that the alt-right has qua the alt-right.

A cooperative scheme is just the set of instrumental functions that any system imposes on the components that it recruits to achieve its intrinsic functioning. In this sense, the alt-right has a complex cooperative scheme. For example, it is often observed that the effect of alt-right trolling is (a) impressing, entertaining, and engaging real and potential allies for the alt-right and (b) threatening, humiliating, and disengaging real and potential enemies of the alt-right (many of whom are offline or have less online power). From here, it is reasonable to infer that trolling is a functional cause of the alt-right achieving its intrinsic functions, such that it features in the alt-right’s cooperative scheme. Likewise, it is also often observed that the effect of individual-level racial, sexual, gender, political, and ideological discrimination is innovation in discrimination that eventually causes innovations in group-level discrimination by the alt-right. From here, it is reasonable to infer that individual-level discrimination is a functional cause of the alt-right achieving its intrinsic functions, such that it also features in the alt-right’s cooperative scheme.

Simultaneously, the alt-right is embedded in broader social systems and thus participates in their cooperative schemes. For example, it is often observed that the effect of the alt-right promoting its radical brand of discrimination is normalizing the subtler brand of discrimination that is favoured by the conservative establishment. From here, it is reasonable to infer that the alt-right promoting its radical agenda has the instrumental function of normalizing the subtler yet still-discriminatory agenda of the conservative establishment. In fact, they might even serve the instrumental function of using these tactics to fight an online proxy war on behalf of the conservative establishment with movements like BLM. Like selfish players in the public goods game, though, the alt-right may have intrinsic and instrumental functions that conflict. It’s completely possible that the alt-right explicitly disavows itself of any ties to the conservative establishment while the conservative establishment recruits the alt-right into its cooperative scheme to unwillingly perform instrumental functions for it. After all, explaining the alt-right’s tactics may still be necessary to achieve a general understanding of the etiology of intrinsic functionality in the conservative establishment.
Critically, though, the alt-right is also embedded in the global society and thus participates in its cooperative schemes. It isn’t entirely clear what the intrinsic functions of the global society are, but it is plausible that its cooperative scheme involves recruiting its human resources as efficiently as possible in order to achieve its other intrinsic functions. To do this, it needs to recruit its human members for the roles that they are most competent for—regardless of their racial, sexual, gender, political, and ideological features. Since these features aren’t correlated with innate-competence distributions, promoting discrimination via these features decreases the effectiveness by which a society recruits its human resources. Thus, the alt-right qua alt-right has no instrumental function for the global society. But the alt-right qua online social movement does have the instrumental function for the global society, to recruit and coordinate online entities with interests in efficient egalitarian institutions to overpower and disrupt offline entities with entrenched interests in inefficient inegalitarian institutions. Obviously, though, the alt-right is incompetent at achieving this instrumental function, so it counts as a highly dysfunctional component of the global society.

Therefore, activities can not only count as instrumentally functional for multiple social groups but the same activity can also count as instrumentally functional for one social group (e.g., the alt-right) and instrumentally dysfunctional for the other. Consider the implications for our problem: the MAC framework entails that morality is fundamentally grounded in cooperative functionality, but it doesn’t say which instrumental functions for which social groups. A non-partisan solution, which moral scientists might prefer, is that morality is fundamentally grounded in all types of cooperative functions for all types of social groups. While this answer might seem uncontroversial, it totally neglects the functional relations among cooperative schemes, which we just considered. For instance, it neglects the difference between grounding morality in the cooperative schemes of systems that aren’t recruited by any other systems (e.g., Earth’s socio-ecosystem), systems that are functional components of systems that recruit them (e.g., BLM), and systems that are dysfunctional components of systems that recruit them (e.g., the alt-right).

§3. Moral Judgment-Making

We need a principle for determining whether an instrumental function for a social system is part of the grounds for morality. Also, we need to ensure that this principle itself is non-partisan. That is a tall order, but I think that it can be fulfilled. The general strategy will be to argue that (a) there is evidence to believe that our use of the predicate ‘is right’ in moral judgment-making refers to a property of cooperative functionality in the context of cooperation and (b) this indicates that the property of moral rightness consists in that property of cooperative functionality within the context of cooperation. This strategy is a descendent of Richard Boyd’s (1988) recipe for moral realism. We’ll modify his strategy in §4 to develop a more sophisticated framework for morality and I’ll resist Horgan & Timmons’ (2000) objection that it entails a problematic kind of moral relativism in §5. Before we get to that, though, let’s review evidence that in the context of cooperation, moral judgment-making has the instrumental function to solve the problem of functional cooperation.  

---

10 For empirical evidence that inequality impairs cooperation, see Turchin (2015) and Cronin et al. (2015). However, see Buchanan (2020) for evidence that inequality only impairs cooperation in certain contexts and not in others.

11 Another, more popular body of evidence about moral judgment-making considers responses to moral dilemmas (since Greene et al., 2001). But in moral dilemmas, it is quite difficult to identify the functional standards involved. As a result, it seems to me that there is much less that we can infer from this body of evidence.
Moral judgment-making can be divided into two broad types: (a) second-personal, where moral judgments are communicated to the agents whose behaviours they evaluate, and (b) third-personal, where moral judgments are communicated to different agents than those whose behaviours they evaluate. Let’s start with Simpson et al.’s (2017) study on second-personal moral judgment-making. They compared cooperative and collective functionality across three conditions. In the control condition, they found that selfish functionality prevailed: contribution to the common pot approached 30% after 9 rounds. Next, Simpson et al. used two other conditions to increase total contribution. In the sanction condition, they allowed players to sanction (penalize) each other for low contributions: this increased contribution significantly, to ~55%. In the moral judgment condition, they allowed players to send moral ratings (from -3 to +3) of the contributions to the players who made them: this increased contributions even higher, to ~65%.\(^{12}\)

From this result, Simpson et al. (2017) conclude that second-personal moral judgment-making is a more effective solution to the problem of cooperation than material sanctioning. They offer two explanations. First, it contributes more effectively to collective functionality because it exploits the intrinsic motivation among players to acquire moral praise and avoid moral blame. Second, moral sanctioning contributes less effectively to collective functionality because it tends to trigger costly cycles of retribution: material sanctions provoke selfish players to sanction the cooperative players to punish them, provoking the cooperative players to punish them even further, etc. These material sanctions progressively deplete the team’s holdings, resulting in fewer holdings by the end of the game compared to when the team was recruiting second-personal moral judgment-making to solve its problem of cooperation. This kind of explanation has been proposed previously on theoretical grounds (Gibbard, 1990; Joyce, 2006).

Again, though, we face a problem of generalization. Recall that in our example, each player can make one of 101 possible contributions. As a result, each player has \(2^{101}\) possible moral judgments to make about the entire set of possible contributions: they can choose one of two judgments for each contribution.\(^{13}\) For example, it is possible to judge that contributing anything more than $0.50 is right and anything less than $0.50 is wrong, contributing $0.13 and $0.74 is right and anything else is wrong, etc. Which of these moral judgments is such that explaining it is necessary and sufficient for achieving a general understanding of its solution to the problem of functional cooperation? A plausible answer is: the moral judgment that evaluates $1.00 as right and anything less than $1.00 as wrong. After all, this is the only moral judgment that will motivate cooperators and defectors to contribute up to 100%, which will maximize the effectiveness with which these second-personal moral judgments can support group functionality. By comparison, the latter will motivate players to contribute up to 65%—not beyond that—which will limit the effectiveness with which these second-personal moral judgments can support group functionality.

This is evidence that instrumentally functional moral judgments evaluate 100% contributions as right, but there’s a caveat. Moral judgments that maximize the functionality of the social group need not count as instrumentally functional. After all, e.g., it’s conceivable that moral judgment-making only functions to increase contributions past some threshold (e.g., 65%) and then some

\(^{12}\) Simpson et al. (2017) noted that this effect should be even stronger when moral judgments are delivered face-to-face rather than as written messages.

\(^{13}\) If we allow participants to make other kinds of judgments (e.g., “it is neither right nor wrong”), then the problem of generalization only gets worse.
other mechanism functions to increase contributions from that threshold the rest of the way up to 100%. Then moral judgment-making increasing contributions past that threshold might prevent the other mechanism from taking over and more effectively increasing contributions up to 100%. And moral judgments that evaluate any behaviours at or above 65% as right would count as functional insofar as explaining them would be both necessary and sufficient to achieving the best general explanation of moral judgment-making (Neander, 2017). Barring evidence of a mechanism that takes-over from moral judgment-making, though, the fact that moral judgments that evaluate 100% contributions as right would be the most effective solutions to the problem of cooperation is good (but defeasible) evidence that they are instrumentally functional.

So, we may provisionally conclude that moral judgments that evaluate contributions under 100% are instrumentally functional. What would more conclusively indicate this? Evidence that there are no mechanisms that take-over from moral judgment-making would be particularly excellent evidence. Correlational evidence could help too, though. For example, suppose that we found that training players in cooperation game theory increased the average contribution that was evaluated as right. The best explanation might be that the training increased the instrumental functionality of the moral judgments and would continue to do so until moral judgments achieved instrumental functionality by evaluating only 100% contributions as right. Or suppose that we discovered that there was a significant positive correlation between the average contribution that was evaluated as right and performance on other cooperation games (or other tasks). The best explanation might be similar: general improvements in cognition increased the instrumental functionality of the moral judgments and would continue to do so until moral judgments achieved instrumental functionality by evaluating only 100% contributions as right.

It might seem that moral judgment-making functions to motivate people to pursue the activities that are evaluated as right and avoid the activities that are evaluated as wrong (e.g., Gibbard, 1990; Joyce, 2006; Sinclair, 2017), but that inference doesn’t generalize to other results. Let’s consider Feinberg et al.’s (2014) study on third-personal moral judgment-making (gossip). They modified the public goods game by dividing the players into sub-groups and ensuring that they were always assigned to a new sub-group in each round. They found that defectors prevailed under control conditions, which resulted in the lowest average contribution (~20% after 6 rounds). But they used two other conditions to increase contribution. In the first condition, they allowed players to send notes about the behaviours of their co-players to whomever would be playing with their co-players in the next round. This gossip created a moral reputation that followed the players, one round at a time. This increased the average contribution to ~40% after 6 rounds. In the second condition, they also allowed players to respond to these notes by voting to exclude players from the game before starting the round. This created material consequences for bad reputations: gossip would lead to ostracism. This increased average contribution significantly to ~90% after 6 rounds.

As before, we can provisionally infer from the fact that third-personal moral judgment-making is most effective when it evaluates 100% contributions as right that it is instrumentally functional when it evaluates 100% contributions as right—in the absence of defeating evidence that there might be mechanisms that take over increasing contributions beyond some threshold. But notice that moral judgment-making doesn’t instrumentally function to motivate people to contribute 100%: it instrumentally functions to inform players about the cooperative functionality of each player’s performance, which causes players to effectively direct their motivation to ostracize the
defectors, which then increases the risk of ostracism for potential defectors that motivates them to contribute 100%. The motivational effects of moral judgment-making don’t generalize from the second-personal to the third-personal cases, but its identification and information effects seem to.

In public goods games, then, teams recruit moral judgment-making to identify the behaviours that are instrumental functionally functional for the team and then recruit second-personal and third-personal implementations of moral judgment-making to motivationally reinforce the instrumental functionality of the players for the teams. But there’s no reason to limit this conclusion to just public goods games, or even just cooperation games. After all, moral judgment-making can also be recruited to instrumentally function for the alt-right, the conservative establishment, BLM, the progressive movement, etc. to reinforce the instrumental functionality among their components. To do this, these social systems promote moral judgment-making that evaluates the behaviours that achieve their instrumental functions as right and the behaviours that fail their instrumental functions as wrong. So, I propose that we may infer a general conclusion: regardless of whether it’s communicated second-personally or third-personally, the instrumental function of moral judgment-making for any social system is to identify and inform whether any behaviour has the property of being instrumentally (cooperatively) functional for that social system.\footnote{Then second- and third-personal moral judgment-making are distinguished by the mechanisms that relate their informing about cooperative functionality with different motivational effects that promote cooperative functionality.}

### §4. Moral Rightness

Recall that Boyd’s (1988) strategy for individuating moral rightness requires us to identify the referents of moral predicates in moral judgments. But we’ll find that the instrumental functions of moral judgment-making don’t determine the referents of moral predicates in moral judgments. Still, they do provide us with evidence that can help us identify the intrinsic function of moral judgment-making, which does determine the referents of moral predicates in moral judgments. The goal of this section is to use evidence from §3 to identify the intrinsic function of moral judgment-making and then use Boyd’s strategy to individuate moral rightness for the functional-cooperation-as-morally-right (FCAMR) framework. I’ll appeal to theoretical considerations in this section to develop my proposal, but theoretical considerations are insufficient to justify my empirical proposal. So, in §5, I’ll demonstrate that the FCAMR framework that we build in this section seems to be necessary for the best general explanation of moral judgment-making.

Alright, let’s start with this extensional claim that is entailed by my final answer:

**FCAMR1.** In the context of cooperation, intrinsically functional moral judgments evaluate behaviours that are (a) instrumentally functional for some unique system $S$ as right and (b) instrumentally dysfunctional for $S$ as wrong.

This provides a general explanation for what makes moral judgment-making liable to recruitment by other systems. After all, it claims that moral judgment-making already has the intrinsic function to track the instrumental functionality for some unique system (presumably, one seems morally legitimate), which makes it liable to recruitment by other systems to support their cooperative schemes. In other words, its having this intrinsic function explains how it comes to have all of its instrumental functions vis-à-vis so many systems. Moreover, this explanation is general insofar as
it provides a solution to the problem of generalizing over the relations between moral judgment-making and so many systems: it suggests that its relation to just one unique system is such that explaining it is necessary and sufficient to achieve a general understanding of moral judgment-making. Thus, the generality of FCAMR1 is a theoretical consideration that counts in its favour.

Of course, the most salient question that we can ask about FCAMR1 is: which unique system? There is only one system that is meaningfully distinct from every other system in virtue of their instrumental relations: the broadest system that recruits a behaviour into its cooperative scheme, which is the only system that doesn’t participate in the cooperative scheme of any other system. After all, all other systems share the relations of having cooperative schemes that narrower systems participate in (including just the behaviour itself) and participating in cooperative schemes that broader systems have. Since the broadest system that recruits a given behaviour is the only system that’s meaningfully distinct, including it in the intrinsic function of moral judgment-making would provide a meaningful solution to the problem of generalizing over all the systems that recruit the behaviour into their cooperative schemes. This counts as a (defeasible) reason to including it in the intrinsic function of moral judgment-making.\(^\text{15}\) Hence:

**FCAMR2.** In the context of cooperation, intrinsically functional moral judgments evaluate behaviours that are (a) instrumentally functional for the broadest system \(S\) that recruits those behaviours and (b) instrumentally dysfunctional for \(S\) as wrong.

The controlled conditions of the laboratory ensured that the broadest system that recruited players’ contributions to the common pot in the public goods game was the team. By comparison, there are multiple candidates for the broadest systems that can recruit behaviours in the field: e.g., global society, Earth’s socio-ecosystem, and even the Solar System’s emerging socio-ecosystem.\(^\text{16}\)

Following a similar inference as in §3, I suggest that evaluation in itself isn’t relevant here: what’s relevant is that these evaluations identify and thus encode information about the functional status of cooperative behaviours vis-à-vis the broadest system that recruits them.\(^\text{17}\) Hence:

\(^{15}\) This kind of intuition tends to be at least somewhat reliable. I suspect that our conclusion that functional moral judgments evaluate only 100\% contributions is supported by this intuition: 0\% contributions and 100\% contributions are the only two meaningfully distinct contributions (as minimum and maximum) and only the latter was a sensible candidate for being evaluated as right by functional moral judgments. I’m counting on this same intuition here.

\(^{16}\) What is worth noting is that the broadest systems that recruit our behaviours are almost always much larger than the global society, which is just their human component. I suspect that this fact justifies an account that treats non-human species as potential partners for cooperation: the global ecosystem may recruit our interactions with them to perform instrumental functions. For example, maybe human interactions with non-human species that support or restore some ecological functions are such that explaining them is necessary and sufficient to achieve a general understanding of ecological functioning. And maybe, human interactions that exploit non-human species for human consumption are such that explaining them is unnecessary and insufficient to achieve a general understanding of the etiology of ecological functioning. This proposal might provide a cooperative conception of our moral relations with non-human animals and other non-cooperators. However, this proposal requires a full defense, which I reserve for future work.

\(^{17}\) Information encoding is well-defined: for any event \(E\) (e.g. the instantiation of an intrinsically functional moral judgment) to encode information about a contingent fact \(F\) (that a behaviour is instrumentally functional for the broadest recruiting system) is for the conditional probability of \(F\)-given-\(E\), \(P(F|E)\), to be greater than the unconditional probability of \(F\), \(P(F)\)—for the instantiation of \(E\) to increase the conditional probability that \(F\) (Skyrms, 2010; Shea, 2018; c.f. Dretske, 1981).
**FCAMR3.** In the context of cooperation, intrinsically functional moral judgments encode the information that a behaviour is (a) instrumentally functional for the broadest system (S) that recruits the behaviour by evaluating it as right and (b) instrumentally dysfunctional for S by evaluating it as wrong.

Per the definition of information encoding (Footnote 17): intrinsically functional moral judgments increase the conditional probability that any given behaviour is (a) instrumentally functional for the broadest recruiting system when they evaluate it as right and (b) instrumentally dysfunctional for the broadest recruiting system when they evaluate it as wrong.

FCAMR3 is an extensional claim (about moral judgments that are intrinsically functional), but our question requires an intensional answer (about the intrinsic function of moral judgment-making). So, my proposal is that we translate FCAMR3 into an intensional hypothesis:

**FCAMR4.** In the context of cooperation, the intrinsic function of moral judgment-making is to encode whether a behaviour is (a) instrumentally functional for the broadest system (S) that recruits it by evaluating it as right or (b) instrumentally dysfunctional for S by evaluating it as wrong.

Thus, moral judgment-making has related yet distinct intrinsic and instrumental functions: (a) its intrinsic function is to identify and inform whether behaviours are instrumentally functional for the broadest system that recruits the behaviours and (b) its instrumental functions are to identify and inform whether behaviours are instrumentally functional for any system that recruits both the behaviours and moral judgment-making. Systems that are instrumentally functional components of the broadest system will recruit moral judgment-making to learn how to better participate in the broadest system whereas systems that are instrumentally dysfunctional components of the broadest system may recruit it to reinforce their own dysfunctionalities.

Now that we have a theory about the intrinsic function of moral judgment-making, let’s identify the referents of moral predicates in moral judgments. To do this, we’ll need to choose a theory of reference. A suitable choice for our purpose is *informational teleosemantics*, which is a popular family of theories of referent determination in the philosophy of cognitive science. It claims that the referent of a term is just the thing that it is the function of the term to carry information about (Dretske, 1981; Papineau, 1987; Millikan 1989; Neander, 2017; Shea, 2018). But advocates of informational teleosemantics haven’t said whether the referent-determining function is intrinsic or instrumental. However, it’s widely believed that every term has a unique referent, which would suggest that the referent-determining function must also be unique. Since moral judgment-making can have a unique intrinsic function yet have many instrumental functions, its intrinsic function is probably its referent-determining function.

If informational teleosemantics is true, then in the context of cooperation, if the intrinsic function of the moral judgment-making is to carry information about whether cooperative behaviours are functional, then the predicates ‘is right’ and ‘is wrong’ that moral judgments predicate of any behaviour refer to the properties of *instrumental functionality* and *instrumental dysfunctionality* for the broadest system that recruits the behaviour, respectively. Thus:
**FCAMR5.** In the context of cooperation, the predicate ‘is right’ refers to the property of being instrumentally functional for the broadest system (S) that recruits the behaviour and the predicate ‘is wrong’ refers to the property of being instrumentally dysfunctional for S.

Finally, we’re prepared to appropriate Boyd’s (1988) strategy to individuate moral rightness. So, the term ‘right’ obviously refers to the property of rightness. And now we have reason to believe that in the context of cooperation, the term ‘right’ also refers to the *instrumental functionality* of certain cooperative behaviours for the broadest recruiting system. Does that implication violate the widely-held belief that every term has a unique referent? I don’t think so. Boyd (1988) proposes a solution: if a term refers to two properties, then the properties must be identical.

I propose a contextualized version of that solution: if a single term refers to two properties within a context, then the property that the term refers to in all contexts consists in the property that it refers to within that context.\(^{18}\) For example, if ‘red’ refers to the property of being maroon and the property of being red in a particular context, then the property that the term refers to in all contexts (i.e., the property of being red) consists in the property that it refers to within that context (i.e., the property of being maroon). Likewise, if ‘right’ and ‘wrong’ refer to instrumental functionality and dysfunctionality for the broadest system that recruits the behaviours in the context of cooperation and ‘right’ and ‘wrong’ refer to moral rightness and wrongness across all contexts, then:

**FCAMR6.** In the context of cooperation, moral rightness consists in the property of instrumental functionality for the broadest recruiting system (S) and moral wrongness consists in the property of instrumental dysfunctionality for S.

FCAMR6 entails that the cooperative part of morality is fundamentally grounded in cooperative functionality *because* the rest of cooperative morality consists in various solutions to the problem of functional cooperation—in mechanisms that facilitate our selecting morally right behaviours and avoiding morally wrong behaviours.

### §5. Cooperation Landscape

Theoretical considerations help to *develop* the functional-cooperation-as-morally-right (FCAMR) framework, but we can only *justify* the FCAMR framework by showing that it’s indispensable for achieving a general understanding of cooperative morality. We’ve already seen in §3 how useful the FCAMR framework can be for explaining moral judgment-making in public goods games. In this section, I’ll provisionally justify the FCAMR framework by using it to develop a preliminary explanation of moral judgment-making in the field. In particular, I’ll argue that the social practice of moral judgment-making is best understood as a search through the landscape of cooperative schemes for the ideal cooperative scheme, which achieves every instrumental function for the largest recruiting system. I’ll also argue that this preliminary explanation refutes the objections that (a) functional frameworks for morality entail the false conclusion that Buchanan & Powell (2018) call “evo-conservatism” and (b) Boyd’s (1988) strategy entails a problematic kind of moral relativism (contra Horgan & Timmons, 2000).

---

\(^{18}\) Boyd (1988) introduced this kind of assumption to distinguish *synthetic* (referent-based) approaches to defining moral terms from the *analytic* (sense-based) approaches that Moore (1903) refuted with his open-question argument.
Let’s start with the concept of a cooperation landscape, which derives from the general notion of a state-space landscape from computer science. A cooperation landscape is just a set of cooperative schemes that varies continuously in altitude: each peak represents a set of cooperative behaviours with high overall instrumental functionality for the broadest recruiting system whereas each trough represents a set of cooperative behaviours with lower overall instrumental functionality for the broadest recruiting system (Figure 3). A critical distinction here is the distinction between local optima, which are peaks that are taller than their local vicinity, and global optima, which are peaks that are taller than the entire landscape. In terms of this distinction, the intrinsic function of moral judgment-making is to discriminate the global optima from the local optima and select the global optima, which represent morally right schemes for cooperation.

This landscape metaphor is typically used to illustrate the performance of search algorithms: as they navigate through the landscape in their search for the global optima, they often get “stuck” on the local optima because they have insufficient drive for exploration and insufficient tolerance for suboptimality to descend from the local optima into the troughs, from where they can continue searching for other peaks. Hence, moral judgment-making imposes two instrumental functions on its (cognitive and social) components: they should have (a) sufficient drive for exploration and (b) sufficient tolerance for (temporary) suboptimality to leave the local optima and eventually reach the global optima. Barring lucky compensations, moral judgment-making will deviate from its intrinsic function when it lacks sufficient drive for exploration and tolerance for suboptimality. It will get stuck on local optima that are higher than their vicinity but low in the entire landscape.

This metaphor is indispensable for generalizing over the social practice of moral judgment-making at two different levels. First, it’s indispensable for generalizing over moral judgment-making in the components of the broadest recruiting system. For example, the practice of moral judgment-making...
making in the alt-right (one component of the broadest recruiting system) is best understood as deviating from its intrinsic functionality because members of the alt-right deviate from their instrumental functions vis-à-vis moral judgment-making (and the broadest recruiting system): their search for instrumental functionality (for the broadest recruiting system) is impaired by their low drive for exploration (their disposition to maintain the identity and power of the “white race”) and their low tolerance for suboptimality (their unwillingness to tolerate the difficulties of integrating races and cultures into non-races and new cultures). Thus, they get stuck on a local peak (which represents a radical white nationalist and white separatist cooperation scheme), which is very low in the cooperation landscape (Figure 3).

By comparison, the social practice of moral judgment-making in BLM (another component of the broadest recruiting system) is best understood as approaching intrinsic functionality because members of BLM approach their instrumental functions: their search for instrumental functionality (for the broadest recruiting system) is facilitated by their strong drive for exploration (their disruptive approach to any status quo) and tolerance for suboptimality (their willingness to forego the benefits of the status quo before the benefits of any new system become available). This drives them down local optima (which represent the racist status quo) into troughs from which they can find local optima on higher ground (which represent cooperation strategies that better promote racial equality). As their search continues, they can gradually gain altitude in the cooperation landscape, finding more and more instrumentally functional cooperation schemes (Figure 3).

Second, this metaphor is also indispensable for generalizing over the practice of moral judgment-making in the broadest recruiting system itself. After all, social groups often stagnate: they end up functioning to occupy certain positions in the cooperation landscape and then recruiting moral judgment-making to maintain these positions—rather than functioning to continue searching the cooperation landscape for the global maximum. For example, holding a local peak that represents a radical white nationalist and white separatist cooperation scheme is the intrinsic function of the alt-right qua alt-right—not finding the global maximum. And holding a region that represents left-wing approaches to racial equality is the intrinsic function of BLM qua BLM—not finding the global maximum. In other words, the alt-right and BLM both identify with potential solutions to the problems of cooperation—not with finding the perfect solutions to the problem of functional cooperation vis-à-vis the broadest recruiting system.\(^{19,20}\)

Fortunately, though, there are mechanisms that take over the search from these social groups. The broadest recruiting system itself recruits the alt-right, BLM, and other groups that recruit moral judgment-making to perform a multi-track search over the cooperation landscape. It uses the alt-

\(^{19}\) I’d only add that the alt-right qua sociopolitical movement and BLM qua sociopolitical movement both share the intrinsic function to search the cooperation landscape for more instrumentally functional strategies for the broadest recruiting system.

\(^{20}\) Both the alt-right and BLM recruit moral judgment-making to reinforce their intrinsic functions. For example, since the alt-right qua alt-right has the intrinsic function of promoting white supremacy and white separatism, it will assign moral judgment-making among its members the instrumental function of reinforcing its promotion of white supremacy and white separatism. Obviously, though, this contradicts its intrinsic function. By comparison, since BLM qua BLM has the intrinsic function of promoting racial equality in a way that is instrumentally functional for the broadest recruiting system, it will give moral judgment-making the instrumental function of reinforcing its promotion of racial equality. This reinforces moral judgment-making’s own intrinsic search for the behaviors that are instrumentally functional for the broadest recruiting system (which includes promoting racial equality).
right to investigate the local peaks that represent radical white nationalist and white separatist cooperation strategies and it uses BLM to investigate the broader region that represents left-wing strategies for racial equality. So, these social groups acquire the instrumental function to represent various cooperative strategies to the broadest recruiting system. Then the broadest recruiting system adjudicates between these two regions in the cooperation landscape by using a variety of mechanisms to favor the social group (BLM) that intrinsically functions to implement strategies that are more instrumentally functional against the social group (the alt-right) that intrinsically functions to implement strategies that are less instrumentally functional for it. This favoring results in group selection (Wilson, 1975; Sober & Wilson, 1994; Nowak et al., 2010).

For example, the broadest recruiting system uses several mechanisms (especially economic ones) to favor BLM’s cooperation scheme over the alt-right’s cooperation scheme. For one, the groups that adopt BLM’s strategies recruit their black and other racialized members more effectively, which increases their competitiveness relative to groups that practice white supremacy or practice racial equality in less effective ways (e.g., by aiming to be “color-blind”). For another, these groups also form stronger and more rewarding alliances with other groups in their environment. For example, corporations have discovered that explicitly allying with BLM (more or less genuinely and effectively) improves customer relations with people of color and white allies, who outnumber and/or out-influence members of the alt-right. So, economic mechanisms can drive group selection and so serve the critical instrumental function of rewarding and selecting groups that find better ways to approach instrumental functionality for the broadest recruiting system.21

Obviously, this is only a preliminary explanation, but it should suffice to illustrate the explanatory power of the FCAMR framework. It also suffices to address the worries that I mentioned earlier. The first worry is that functional frameworks for morality are committed to “evo-conservatism”: the implausible claim that moral judgment-making is pre-determined by our evolutionary past (Buchanan & Powell, 2018; Buchanan, 2020; see also: Smyth, 2017). To be clear, many functional frameworks are committed to this idea. But our preliminary explanation shows that the FCAMR framework isn’t. On the contrary, it’s consistent with the social practice of moral judgment-making being a radically open-ended process that has left our evolutionary past behind, is approaching the foothills of the cooperation landscape, and will eventually ascend to global peaks, where it can achieve ideal forms of cooperation that exceed anything that we can currently imagine. For now, all that we could know about these ideal forms of cooperation is that they are such that explaining them is necessary and sufficient to achieve a general understanding of the relations between cooperation and the broadest recruiting systems.

The second worry is that functional frameworks for morality entail moral relativism. Horgan & Timmons (1991, 2000, 2015) have raised this objection against similar theories that use reference to individuate moral properties (Boyd, 1988; Brink, 1989; Copp, 1995).22 Strictly speaking, the FCAMR framework does entail moral relativism (in the context of cooperation): what counts as

---

21 Obviously, the reality is much more complicated than the idea that every economic process unilaterally supports moral progress by rewarding the search for the cooperative behaviours that are instrumentally functional for the broadest recruiting system. This complexity opens an entire domain of inquiry: moral economics.

22 Horgan & Timmons (2000) argue that using reference to individuate moral properties (per Boyd, 1988) can also entail conceptual relativism or indeterminacy. However, they argue that moral relativism is the implication for any approach that uses a theory of reference (including informational teleosemantics) that claims that relations between terms and the external world are sufficient to determine reference (without intentions).
morally right in the context of cooperation is relative to the instrumental functions of cooperative behaviour for the broadest recruiting systems. However, this kind of moral relativism doesn’t seem to be problematic. After all, it rules out that moral rightness is relative to any immoral cooperative schemes: e.g., Nazi Germany’s, the conservative establishment’s, and the alt-right’s, because these are small, dysfunctional components recruited by much broader recruiting systems. In fact, it even rules out that moral rightness is relative to any moral cooperative schemes: e.g., BLM’s or even the global society’s for the same reason that these are components of recruited broader systems like Earth’s socio-ecosystem or even the Solar System’s emerging socio-ecosystem. So, we have no reason to believe that the kind of moral relativism in the FCAMR framework is problematic.

§6. Conclusion

In conclusion, the morality-as-cooperation framework is implicitly committed to the claim that morality is grounded in the normative notion of cooperative functionality. Moreover, the function of cooperation is an instrumental relation between the cooperative behaviour and the system that recruits it for its own functioning. So, there are as many cooperative functions for any behaviour as there are systems that recruit it. Which of these are moral? Well, the evidence suggests that the intrinsic function of moral judgment-making is to encode whether behaviours are cooperatively functional for the broadest system that recruits it. If the referent of a term is just the thing that the term functions to encode information about (per informational teleosemantics), then the referent of the term ‘right’ is just the property of being cooperatively functional for the broadest recruiting system. From this, we should conclude that moral rightness consists in this property within the context of morality. We’ve found that this functional-cooperation-as-morally-right (FCAMR) framework seems to earn its keep. It explains how the social practice of moral judgment-making is an open-ended process that breaks free from our evolutionary history: it searches through the cooperation landscape for the set of all instrumental functions for the largest recruiting system.

The explanatory power of the FCAMR framework is mostly due to the explanatory power of the functional inferences that we used. After all, I quickly rejected the need to appeal to evolutionary evidence to make functional inferences—following Smyth’s (2017) domain-specific concern that our current moral situation is discontinuous from our historical moral situation and the domain-general observation that scientists often draw functional inferences without evolutionary evidence (Godfrey-Smith, 1993; Kitcher, 1993; Amundson & Lauder, 1994; Walsh, 1996; Neander, 2016). Instead, following Neander (2017), I used a new rule for functional inferences that assumes that they are optimized to solve the problem of generalization: the activities that are functional are the activities that are such that explaining them is necessary to achieve a general understanding of the acting system. This rule supported several functional inferences in my argument that I couldn’t have supported by appealing to evolutionary evidence.23 Since this rule for functional inferences is indispensable to the FCAMR framework, future work should develop it, find its limitations, and explore its other implications.

As we saw in the introduction, the success of the FCAMR framework should be restricted to the context of cooperation, to avoid the cooperation dogma. This provides some relief for traditional moral philosophy: if a behaviour seems morally right despite being cooperatively dysfunctional or

23 In fact, this non-evolutionary notion of function is the reason why the FCAMR framework can explain how moral judgment-making has broken free from its evolutionary history without losing its function.
non-functional for the broadest recruiting system, we can insist that it counts as morally right outside of the context of cooperation. But the FCAMR framework still poses a threat to traditional moral philosophy, because it provides a proof-of-concept for Boyd’s (1988) strategy. We can then appropriate this strategy to individuate moral rightness in non-cooperative contexts. For example, we can (a) identify the instrumental function of moral judgment-making in our relations with non-human animals, (b) infer its intrinsic function, (c) infer the referent of ‘right’ that is predicated in those moral judgments, and then (d) individuate moral rightness as that referred property within the context of our interactions with non-human animals. So, we can follow the FCAMR framework in appropriating Boyd’s (1988) strategy to progressively individuate moral rightness across as many contexts as it could instantiate in. Future work should extend this strategy and consider the implications for traditional moral philosophy.

Ultimately, the FCAMR framework (and any other framework that it inspires) must provide the moral sciences with epistemic access to what it aims to explain: the natural kinds that compose the domain of morality. Future work should drive this framework from here in philosophy out into the moral sciences—to see whether the FCAMR framework can displace the MAC framework and earn its keep. One promising strategy for doing this is to show that the FCAMR framework offers the resources to support a functional decomposition of moral cognition, which would involve clearly defining what counts as correct performance on moral judgment-making tasks and using experimental manipulations to improve and impair performance on those tasks to determine which components contribute in which ways performances on the tasks. This is the standard methodology in cognitive psychology (e.g., Anderson, 1990; Piccinini & Craver, 2011; Griffiths et al., 2020). Unfortunately, though, moral psychology hasn’t been able to use this standard methodology due to lacking clearly defined metrics for what counts as correct performance on a moral judgment-making task. So, one excellent way for the FCAMR framework to show that it earns its explanatory keep is to show that it supports a functional decomposition of moral cognition. Future work should use this strategy and look for others in order to drive this framework into the moral sciences.

References


