

## Metaethics, Teleosemantics and the Function of Moral Judgements

Neil Sinclair (neil.sinclair@nottingham.ac.uk) – 13<sup>th</sup> March 2012

Forthcoming in *Biology and Philosophy*

The final publication is available at:

<http://www.springerlink.com/content/f5600uu426458r50/>

DOI: 10.1007/s10539-012-9316-4

**Abstract.** This paper applies the theory of teleosemantics to the issue of moral content. Two versions of teleosemantics are distinguished: input-based and output-based. It is argued that applying either to the case of moral judgements generates the conclusion that such judgements have both descriptive (belief-like) and directive (desire-like) content, intimately entwined. This conclusion directly validates neither descriptivism nor expressivism, but the application of teleosemantics to moral content does leave the descriptivist with explanatory challenges which the expressivist does not face. Since teleosemantics ties content to function, the paper also offers an account of the evolutionary function of moral judgements.

**Key words:** Gibbard; Metaethics; Millikan; Teleosemantics.

What is the content of a moral judgement, say, the judgement that murder is wrong? According to one old school of thought there are moral ways the world might be and moral judgements represent the world as being those ways. According to another, younger, school of thought moral judgements do not represent the moral way of the world, rather they represent some non-moral way the world is to be made. The old school is represented by moral realism (Moore 1903; Brink 1989; Railton 1986; Shafer-Landau 2003; Wedgwood 2007) and certain kinds of error theory (Joyce 2001a); the not-so-old school by emotivism (Stevenson 1944; 1963), expressivism and its quasi-realist and norm-expressivist developments (Blackburn 1993; 1998; Gibbard 1990; 2003). This paper attempts to shed light on this debate through the refracting prism of a particular theory of content: teleosemantics. This is the view, roughly, that the representational content of a mental state is determined by its function. Since it aspires to be a general theory of content teleosemantics has the potential to adjudicate on the issue of moral content. This paper assesses that potential.

I proceed as follows. In §1 I refine the claims of expressivism and descriptivism. In §2 I introduce teleosemantics. Since teleosemantics ties content to function, to assess moral content by reference to teleosemantics it is necessary to consider the function of moral judgements, which I do in §3. In §4, teleosemantics is applied to judgements with the type of function outlined in §3. Finally, §5 considers how this account bears on the metaethical debate between expressivism and descriptivism.

## **1. Moral Judgements, Descriptivism, Expressivism**

By ‘moral judgement’ I shall mean the state of mind conventionally expressed by sincere utterance of an indicative moral sentence. Thus the moral judgement that murder is wrong is the state of mind conventionally expressed by sincere utterance of the sentence ‘Murder is wrong’. The sentence characterises the content (simpliciter) of the judgement.

Moral judgements, like maps, wishes and instruction manuals, are about the world. They represent the way the world is (maps) or the way the world is to be made (instruction manuals) or the way the world would ideally be (wishes). The particular things that a given item is about are its ‘representational content’. Like jam, representational content comes in many flavours. An item is descriptively representational (‘descriptive’ for short) just when it represents the world, or some aspect of it, as being thus-and-so. The particular way it represents the world as being is the descriptively representational content of the state (‘descriptive content’ for short). Beliefs and maps are paradigm descriptively representational items. By contrast, an item is directive representational (‘directive’) just when it represents the world, or some aspect of it, *so as to be made* thus-and-so. The particular way it represents the world so as to be made is the directive representational (‘directive’) content of the item.<sup>1</sup> Desires and instruction manuals are paradigm directive representational items. Directive content is content most easily captured by imperatives. Note that nothing so far said precludes a single item having both descriptive and directive content. Perhaps, for example, the waggledances of bees both represent the location of nectar and instruct other bees to collect it (Millikan 1984: 99, 115; 1995: 190-1).

---

<sup>1</sup> Terminology from Millikan (1995).

These definitions allow a more perspicuous account of the debate between the old and young schools of metaethics, that is, between descriptivism and expressivism. Descriptivists accept:

**D.** Moral judgements involve a descriptively representational element that represents the world as being a certain moral way, i.e. moral content is descriptive content.

(Note that **D** is neutral on the question of whether moral judgements ever *correctly* describe the world, and if so, on the nature of the part of the world they thereby represent.)

Expressivism denies that moral judgements involve moral descriptive representation. Insofar as moral judgements are distinctively moral, expressivism claims, they involve directly representational elements of a specific non-moral kind. So, for example, a simple version holds that the judgement that murder is wrong represents the world so as to be made that (i) *no one commits murder* (ii) *everyone disapproves of murder* and possibly also (iii) *everyone disapproves of those who fail to disapprove of murder* (Blackburn 1998: 8-14; Stevenson 1963). These are distinctive non-moral directive contents: they are non-moral ways in which the world is to be made, according to the judgement. Generalising, expressivists accept the negative claim:

**E1.** Moral judgements do not involve a descriptively representational element that represents the world as being a certain moral way, i.e. Moral content is not descriptive content.

And the positive claim:

**E2.** Moral judgements involve a directly representational element that represents the world so as to be made a distinctive non-moral way.

So understood, the debate between descriptivists and expressivists concerns the crucial question: Do moral judgements have descriptive moral content? Traditionally, metaethicists have approached this question by asking whether moral content bears any of the hallmarks of descriptive or non-descriptive content. Thus, on the descriptivist side, many have argued that the fact that moral content is expressed in subject-predicate form and is capable of embedding in logical contexts such as conditionals is evidence that it is descriptive content (Brink 1989: 14-36; Shafer-Landau 2003: 13-38). On the expressivist side, many

have argued that the fact that moral contents are motivationally engaged is evidence that moral content is *not* descriptive content (Blackburn 1998: 70; Gibbard 2003: 8-13; Stevenson 1937).

One problem with all such arguments, on both sides of the debate, is that none of the proposed distinguishing or debarring marks of descriptive content are justified wholly from a prior understanding of the notion of *description* itself. In assessing the descriptivist arguments, for example, we might ask: *Why* are only descriptive mental states, as such, amenable to expression in subject-predicate form? In assessing the expressivist arguments, we might ask: Why are descriptive states, as such, necessarily motivationally inert? Both questions can be traced to a deeper, largely unaddressed, concern: What is it about the nature of descriptive states, qua descriptive states, that justifies taking these features as distinguishing or debarring of descriptive status?

A better approach to the crucial question is to work systematically from a general theory of description. On this view, we ask first ‘What are the necessary and/or sufficient conditions for any state to have descriptive content?’ and then apply our answer to the case of moral judgements to determine their descriptive content, if any. This is the approach taken here.

## **2. Teleosemantics**

### **2.1. Function for Teleosemantics**

According to teleosemantics, the descriptive content of a state is determined by the function of states of that type. Here ‘function’ is typically treated aetiologically: the function of a type is that effect for which it has been selected. ‘Selection-for’ can be understood in purely evolutionary and genetic terms (via processes of natural selection) or more broadly to encompass non-genetic inheritance (such as the products of social learning) or more broadly still to include the selection of traits within the learning processes of a single individual (Macdonald and Papineau 2006: 12-16). In the most general terms the effect for which a type has been selected is what tokens of that type have done in the past that accounts for the continued survival and proliferation of tokens of that type (Millikan 1984: 25-33). Note, therefore, that the selected-for effect need not be a common effect of items of that type, so

long as the occasions when the effect does occur are sufficient to explain the continued proliferation of the type (Millikan 1984: 29).

## **2.2. Input-based Teleosemantics**

There are two broad ways, according to teleosemantics, in which function might determine descriptive content. According to the first view the content of a descriptive representation is that worldly feature it has the function of indicating or ‘co-varying’ or ‘corresponding’ with, or carrying information about (Dretske 1986; 1988; 1995; Gibbard 1990: 108-9; Harms 2000; Joyce 2001b; Neander 1995). This view is sometimes called ‘input-based’ teleosemantics, since the content-determining function concerns the processes that produce representations (Macdonald and Papineau 2006: 4-9). More schematically, input-based teleosemantics holds:

**TS1.** State *S* has descriptive content *p* if *S* has the function of indicating *p*.

Here *p* is a type of worldly state of affairs, such as the instantiation of a property. How might a state perform the function of indicating *p*? The standard answer is when the state is, in the course of normal functioning, the product of a system whose function is to adjust some part of the world in response to the presence of (determinate values of) *p* (Dretske 1995: 2; Millikan 1984: 107-9). So, for example, if a system has the function of flooding the bloodstream with adrenaline at a particular time, in response to the presence of a predator, then the function of flowing adrenaline is indicating a co-present predator. Further distinctions within input-based teleosemantics can be made according to the strength of the notion of ‘indication’. On strong views, *X* indicates *Y* just in case if there is an *X* then there is a *Y* (Dretske 1988). On weaker views, *X* indicates *Y* just in case *X*’s are, in the relevant environments, strongly correlated with *Y*’s, perhaps through a causal connection between *Y*’s and *X*’s (Neander 1995; see Lloyd 1989 for even weaker views, based on probabilistic dependence).

## **2.3. Output-based Teleosemantics**

According to output-based teleosemantics, the descriptive content of a state depends on the way it affects the behaviour of its ‘interpreters’ or ‘consumers’. On this view, the function of a descriptive state is to adapt the behaviour of a consumer mechanism so that the consumer

mechanism can fulfil *its* function. The Normal condition under which the consumer mechanism, as adapted to the descriptive state, fulfils its function is the descriptive content of that state. Here a ‘Normal condition’ is, roughly, the condition that held in those past cases relevant to selection, that is, those past cases that explain the continued survival and proliferation of the consumer mechanism and descriptive state-type. These claims require some unpacking.

A Normal explanation of a function is an explanation of how past tokens of the mechanism have (predominantly) performed this function. The Normal conditions of the performance of this function are those that must be mentioned in such explanations (Millikan 1984: 33-4). Typically the consumer mechanism of a descriptive state has the function of bringing about some state of affairs. For example, a consumer mechanism might be a behavioural module of a toad which is responsible for bringing it about that food is deposited in the mouth of the toad. This consumer mechanism is guided by (‘consumes’) a descriptive state insofar as the behaviours that it prompts differ according to the descriptive state that prompts it. In the past, consumer mechanisms of this type that were so guided *did* perform their function, and the condition that must be mentioned in the explanation of how they achieved this is the Normal condition for the performance of this function. This condition is the descriptive content of the guiding state. So the Normal condition under which a particular behaviour, prompted by a particular descriptive state, fulfils the function of the consumer is the content of the guiding descriptive state. For example, suppose the toad behavioural module ‘consumes’ a particular descriptive state that prompts the behaviour of a ‘snap’ of a sticky tongue in a particular direction.<sup>2</sup> The Normal condition under which this behaviour fulfils the function of the consumer, viz. getting food, is the condition of there being toad food in the immediate (snap-directed) environment. Hence, as ‘consumed’ by this mechanism, the descriptive state describes the world as being such that: *food is there*.

Schematically, output-based teleosemantics holds:

**TS2.** S has descriptive content p if S has the function of guiding consumer mechanisms to fulfil *their* (i.e. consumers’) functions, and p is the Normal condition of fulfilment of those functions. (Millikan, 1984: 95-113; 1995; Papineau 1987: 67).

---

<sup>2</sup> This is a simplification: see Neander 2006.

On this view the content of descriptive states does not depend directly on their functions. Rather their content depends on the Normal condition under which they fulfil their function of guiding consumer mechanisms to successful (i.e. function-fulfilling) action.

Millikan summarises output-based teleosemantics thus:

...[descriptive] content is determined by the rules...in accordance with which representations need to correspond to world affairs (to what they represent) if the responses of [their] interpreters are going to produce the effects they are designed to produce by the way of normal mechanisms. (2006: 100)

Millikan also provides an example of more complex descriptive state, viz. a belief. She notes:

If my belief that there is an umbrella in the hall closet is to help guide my decision-making and action-guiding apparatus (i.e. the belief's "interpreter") such that it serves its function of helping fulfil my desires, for example my desire to keep off the rain, then there needs to be an umbrella in the hall closet. If it is to help me make a correct inference concerning, for example, whether or not Susan returned my umbrella, one that yields this truth not by accident but in accordance with the good design of my cognitive systems, still there needs to be an umbrella in the hall closet – and so forth. (1995: 189)

Note that, as this last example demonstrates, most descriptive states, at one time and another, will have more than one consuming mechanism (such as the decision-making mechanism and inference-making mechanism) each with different functions. But univocality of descriptive content will be secured so long as (and only so long as) there is univocality of the Normal conditions of the successful fulfilment of *all* those functions (Millikan 1989: 289-90).<sup>3</sup>

## 2.4. Directive Content

---

<sup>3</sup> Note that, although input-based and output-based teleosemanticists disagree about the precise way in which function generates representational content, they will tend to agree about the mechanisms at work in actual cases of representation. In particular, the input-based teleosemanticist will accept that which worldly feature a system has the function of indicating is determined by how the system is used by its consumers and an output-based teleosemanticist will accept that the way in which a representational system fulfils its function of successfully guiding consumer mechanisms is by indicating some worldly feature. I distinguish the views here both because it is common to do so and, more importantly, because it is not obvious that **TS1** and **TS2** generate the same content when applied to particular cases, such as the case of moral judgements.

Although the issues of directive and direct content can be treated separately, it is common for defenders of teleosemantics, in both input-based and output-based forms, to accept the following teleological account of *directive* content:

**TS3.** A state has directive content *p* if it is the function of that state to bring it about that *p* obtains. (Millikan 1984: 99-102)

Thus my desire for chocolate has the function of bringing it about that I have chocolate, and the behavioural module of the toad has the function of bringing it about that the toad has some food. Thus these states represent the world so as to be made that, respectively, I have chocolate and that the toad have food.

Given **TS3**, both input-based and output-based versions of teleosemantics can admit the possibility that some descriptively representational states are also *directively* representational states.

Consider, first, input-based teleosemantics. It is possible for a state to satisfy *both* the sufficient condition given by **TS3** and the sufficient condition given by **TS1**. Consider the waggledance of honeybees. This state is the product of a mechanism that has been selected for to produce ‘dances’ whose final stroke indicates the direction of nectar. So the dance itself has the function of indicating the direction of nectar. But the mechanism has also been selected for because of its effects, namely the motion of other (‘consumer’) bees towards the nectar. So products of that mechanism (individual ‘dances’) also have the function of producing a certain state of affairs, namely ‘consumer’ bees flying towards the nectar. So waggledances have the function of directing consumer bees to fly towards nectar, but they achieve *this* function in virtue of functioning to indicate the presence of nectar.

Consider, second, output-based teleosemantics. It is perfectly possible for a state to satisfy *both* the sufficient condition in **TS3** and the sufficient condition in **TS2**. Millikan gives an example:

Consider...a very primitive representation: the food call of a hen to its brood. A...function of this call is to cause chicks to come to the place where the food is and so to nourish them. Assume, what is reasonable, that this is the...only effect the call has been selected for. Then the call is directive, saying something like “come here now and eat!”. But it is also a *condition* on the proper performance of



the call that there be food there when the hen calls. So the call is also descriptive, saying something like “here’s food now”. (1995: 190)

As Millikan puts it, such states are ‘Pushmi-pullyu (PP) representations’, since they represent both the way the world is and the way the world is to be made.

In the Philosophical Zoo of mental states, PP representations are housed next to the cage marked ‘besires’ (although no one is really sure whether the latter is inhabited). Like PP representations, besires are supposed to have the ‘direction of fit’ associated with both descriptive and directive representations. Unlike the PP representations noted above, though, besires are supposed to be mental states of mature deliberating humans. Scepticism about the existence of besires often rests on the claim that whatever their apparent closeness, the descriptive and directive contents of such states can always be pulled apart, at least modally. That is, we can imagine that the same individual possesses a state with just the descriptive content, or just the directive content. This point is certainly true, but given the relatively weak constraints on imaginability its relevance is unclear. For example, it is certainly not true that in the actual case of the hen’s call the relevant individual first possesses a representation of the way the world is, then ‘decides’ to couple this representation with a representation of the way the world is to be made, with the eventual result in action. In the hen at least, the connection between descriptive and directive content is more direct, with no possibility, while the cognitive mechanisms of the hen remain as they are, of decoupling the two contents in inferential thought. As Millikan puts it:

...the effect of the call on the chicks is not filtered through an all-purpose cognitive mechanism that operates first by forming a purely descriptive representation (a belief that there is food over there), then retrieving a relevant directive one (the desire to eat), then performing a practical inference and, finally, acting on the conclusion. Rather, the call connects directly with action. *Its function is to mediate the production of a certain kind of behaviour such that it varies as a direct function of a certain variation in the environment, thus directly translating the shape of the environment into the shape of a certain kind of conforming action: where the hen finds food, there the chick will go* (1995: 190).

PP representations, at least in hens, are psychologically unified mental states. It remains a possibility, therefore, that such Janus-faced representational states also play a role in the motivational lives of mature deliberating humans, even though most of an agents’ descriptive

states do have contents that are psychologically separable (e.g. in inference) from any directive content. Once more, this possibility will need to be considered in the case of moral judgements.

## 2.5. Teleosemantics and the Desiderata for Theories of Descriptive Content

Why consider either version of teleosemantics as a yardstick by which to judge the issue of moral content? The theory is reasonably popular (Dretske 1986; Macdonald and Papineau 2006; Millikan 1984; Papineau 1984) and there is precedent for applying it to moral content (Gibbard 1990: 105-125; Harms 2000; Joyce 2001b). Furthermore, teleosemantics satisfies several plausible desiderata for theories of mental representation.

First, it is naturally respectable, including in its explanations of representational content only notions – such as ‘function’ – either elsewhere employed by the natural sciences or definable in terms of such notions (Macdonald and Papineau 2006: 1-3).

Second, as the examples above demonstrate, it generates plausible content attributions, that is, the representational contents it attributes to mental states generally match those we pretheoretically assign. (For this desideratum see Neander 2006: 167 and Papineau 1993: 94.)

Third, teleosemantics allows for the possibility of misrepresentation, where the descriptive content of representational states fails to obtain. Teleosemantics explains *misrepresentation* in terms of *malfunction*, or failure to perform selected for effect (Dretske 1986; Millikan 1995). So, on input-based teleosemantics, a particular representation can have the function of indicating some feature, but on a particular occasion may be present whilst failing to perform that function. The point applies, *mutatis mutandis*, on an output-based view.

Fourth, teleosemantics – at least in a suitably supplemented form – allows for a particular kind of misrepresentation that occurs when the representational content of a state involves non-existent entities, like unicorns or phlogiston. It is a feature of our representational states that we can represent to ourselves ways the world could be that are never realised. This has been called the no-reference problem. Initially, teleosemantics may seem to face problems here, for it is a theory on which our powers of representation rest on the things that we represent *being there*, if not concurrent with every representation of them, at least in the conditions relevant to the selection of representations of that kind (the Normal

conditions). A common response is to supplement the basic teleosemantic theory by admitting a type of complex representation whose content is determined by (basic) inferential ties to more simple representations to which teleosemantics directly applies. For example, a state might represent an object ( $x$ ) as a unicorn when it is inferentially tied to more basic (descriptive) representations with contents such as: *x is a quadruped*; *x is horned*; and so on. The complex descriptive state will be given its content by being associated with a dossier of simpler, teleologically explicable, representations (Joyce 2001b; Millikan 1984: 132-4; Sterelny 1990: 134-40). Sometimes the contents of the dossier are such that there is nothing in the world that exemplifies all the qualities therein represented. In that case the descriptive state will be systematically false – as in the case of representations of unicorns.

Finally, a theory of descriptive representation must allow that distinct descriptive contents can represent the same worldly state of affairs, albeit in different ways, or under different ‘modes of presentation’. For example, ‘Penelope believes that Eric is brave’ and ‘Penelope believes that Bananaman is brave’ report type-distinct descriptive states but have the same subject matter (for Eric is Bananaman). This has been called the co-reference problem. The preceding response to the no-reference problem also provides the materials to deal with this problem. Teleologists can hold that distinct complex concepts, defined by distinct inferential relations between more basic concepts, can be used to describe the same state of affairs, thus providing distinct modes of presentation of that state. For example, ‘That is Hesperus’ ties inferentially to the belief that that appears in the morning, whereas ‘That is Phosphorous’ ties inferentially to the belief that that appears in the evening. As it happens, Hesperus and Phosphorus are the same celestial body, but the distinct content-determining inferential ties provide distinct modes of presentation of the same state of affairs.

As these last two points highlight, teleosemantics is most plausible when it takes a ‘modest’ form (Sterelny 1990: 124). Such views do not attempt to apply the teleosemantic theory directly to *all* descriptive states. Rather teleosemantics is taken to describe the basic level of representational content, from which more complex representations can be constructed using distinct mechanisms of, for example, abstraction and recombination. This is why I stated **TS1** and **TS2** as providing only sufficient conditions for descriptive content, since they will not directly apply to the type of complex content constructed through inferential ties. This modesty will be important in what follows, since we will have to consider the possibility that moral concepts are of type to which the teleosemantic analysis does not directly apply (see §5.2).

For all these reasons, teleosemantics appears to be a promising a theory of (descriptive) representation. This is not to deny that teleosemantics faces serious difficulties – it faces many (see, for example, Macdonald and Papineau 2006). But I take the above points to at least provide teleosemantics with sufficient respectability to make worthwhile the task of tracing its consequences for the issue of moral content.

### **3. The Function of Moral Judgements**

#### **3.1. Assumptions**

According to teleosemantics, content is determined by function. So determining the content of moral judgements requires determining their function. Teleosemantics is thus a bridge between the expressivist-descriptivist debate in metaethics and empirical debates concerning the function of moral judgements.

For the purposes of exploring the metaethical side of this bridge, I will make several controversial assumptions about the function of moral judgements. These assumptions are both common and, I believe, defensible. However, I will attempt only a partial defence here, since my primary concern in this paper is with metaethics. Accordingly, the ensuing arguments must be understood in the same way as many of the arguments of other metaethicists who consider the evolutionary history of morality (for example, Gibbard 1990: 107-8 and Street 2006: 112), namely as conditional on accuracy of the evolutionary account sketched. It is, it seems to me, an important part of the job of philosophy to trace the philosophical consequences of empirical hypotheses that are plausible, if not (yet) conclusively established.

The first empirical assumption is that, in the case of moral judgements, the type of function at issue is evolutionary function: morality is an evolutionary adaptation. Second, I will assume that the mechanisms that produce and regulate moral judgements have a *single* type of evolutionarily selected-for effect, from which the functions of particular types of moral judgement can be derived. That is, I will assume that at some point in evolutionary history, the ability to make and be moved by moral judgements – what I will call ‘the moral habit’ – contributed to the biological fitness of individuals or groups who possessed that ability, and the single type of effect that this ability had that explains this increased fitness is the evolutionary function of the moral habit. Relatedly, I will assume that the moral habit is a

functionally unified kind, that is, that different types of moral judgements (of wrongness, goodness and so on) are not products of separate adapted mechanisms with distinct evolutionary histories. Finally, I will assume that different types of moral judgement contribute to the fulfilling of this function in different ways.

My defence of these assumptions will be partial, but sufficient, I hope, to justify tracing their consequences. First, as previously noted, these assumptions are not uncommon (see, for example, Gibbard 1990: 61, Harms 2000 and Joyce 2006). At least then, what follows will be of interest to those who share the assumptions. Second, some justification for the assumption that morality has a function is given by the fact that there is an obvious sense in which moral judgements can be defective (i.e. fail to fulfil their function) and by the fact that morality clearly plays an important role in our lives and interactions (where current use is defeasible evidence of selected-for effect – see Sturgeon 1992). Furthermore, the seeming ubiquity of the moral habit (it is present even in culturally isolated populations – see Gibbard 1990: 3) and the fact that it develops in all normally brought-up humans with little formal instruction is (defeasible) evidence that this function is evolutionary.<sup>4</sup> Regarding the unity of this function (the second and related assumption) again the current use to which moral judgements are put is some guide. We find, for example, that it is when we concerned, directly or indirectly, with the behaviour of ourselves or others that we reach for moral verdicts, be they of wrongness, blameworthiness or cowardice: ethics is other people's business (Blackburn 1998: 8-14). This unifying theme in the current use is defeasible evidence for a unified function.<sup>5</sup> Furthermore (in support of the final assumption) it is

---

<sup>4</sup> Machery and Mallon (2010) elucidate arguments for thinking that the capacity to make *normative judgements* is an adaptation, but deny that these arguments transfer to the moral case, given a 'thick' understanding of moral judgements. My position is consistent with these arguments since my understanding of moral judgements is 'thin' – that is, identical with Machery and Mallon's understanding of normative judgements. Note also that, in any case, my arguments do not require that the function at stake is evolutionary, since the notion of 'function' involved in teleosemantics is liberal, applying equally well where the selection process works through mechanisms other than natural selection, such as social learning and transmission (see Macdonald and Papineau 2006: 12-16).

<sup>5</sup> One may doubt that the following account of the function of morality does justice to a diverse phenomenon, in particular it seems to provide a more natural explanation of judgements of fairness and reciprocity than more self-directed judgements of courageousness and self-respect. There are two possible replies. (1) Accept the diversity and read what follows as applying only to that narrow range of moral judgements for which the evolutionary sketch offered here is plausible. This does not affect the impact of my arguments on the debate between expressivism and descriptivism, since, as usually offered, these aim to be accounts of all moral judgements. (2) Deny that the apparent diversity demonstrates a disunity of function, since it is perfectly possible that different types of judgement help towards the fulfilment of a common function through fulfilling

obvious, or obvious after cautious observation of current use, that although different types of moral judgement are all concerned with the behaviour of others, they are concerned with it in distinct ways: some are concerned with the motivation behind the behaviour, others with the rule or principle under which the behaviour falls, others with consequences. Finally, all these assumptions can be justified on the basis of the explanatory power of the theory of morality that emerges from them. That is, insofar as these assumptions can allow for the construction of a theory that explains all of the important phenomena that a theory of the moral habit ought to explain, the assumptions are shown to be justified (see Gibbard 1990: 35). It is to that theory that I now turn.

### **3.2. An Evolutionary Sketch**

What, then, is the evolutionary function of moral judgements? I propose it is interpersonal co-ordination. Roughly, moral judgements are the products of a mechanism that allows groups of interacting individuals to co-ordinate their actions and emotions for mutual benefit. The function of the moral habit is therefore to produce mutually beneficial co-operative patterns of action and emotion. Particular types of moral judgement derive their more particular functions from the particular role they play in this process. Below I sketch some of salient features of the view, which I take to be that defended by Blackburn (1998: 24-83), Millikan (1995: 192-4), Gibbard (1990: 55-83) and Joyce (2006: 107-143).

First, begin with the observation that groups of individuals living in the same space will face evolutionary bargaining (co-ordination) problems. These are situations where certain patterns of cooperative behaviour will lead to mutual benefit, in terms of biological fitness, as compared to a certain baseline of non-cooperative behaviour (Gibbard 1990: 67). Here a mutually beneficial pattern of co-operative behaviour is a pattern of actions on behalf of a number of individuals that will benefit each individual so long as the others also (mostly) perform their roles. An example is mutual grooming: adult members of a group engaging in mutual grooming to remove parasites will be of benefit to all members of the group, as compared to no grooming. In most situations, however, more than one pattern of mutually beneficial co-operation will be feasible, with certain individuals apt to gain more from some

---

different derived functions (in just the same way that the different members of a football team have different derived functions even though every member of the team has been selected to help win the game).

patterns of co-operation than others. For example, members of the group could spend ten minutes per day engaged in mutual grooming, or two hours. The former allows the group to stay mobile and avoid more predators but removes fewer parasites, thus benefiting individuals more resistant to parasites and less able to fight off predators. The latter allows less mobility but weeds out more parasites, thus benefiting individuals less resistant to parasites but more able to fight off predators. In such cases particular evolutionary pressures will select for a particular pattern of co-operation, known as an 'evolutionarily stable strategy':

...a strategy such that if all adopt it, then for each individual that strategy is at least as fitness enhancing as would be any other strategy easily accessible by mutation. (Gibbard 1990: 67)

For beings with desires and goals tending towards, but whose objects are not always co-extensive with, biological fitness, a related type of bargaining problem will arise. These are situations where patterns of co-operative behaviour will lead to mutual benefit, in terms of increased desire-satisfaction, as compared to a baseline of non-cooperative behaviour. Again, where agents' desires tend to be (but are not invariably) for fitness-enhancing things (such as desires for sweet food, sex) evolution will select an evolutionarily stable strategy. An example here concerns norms of conversational distance (Gibbard 1990: 70). A group of interacting agents desire both to communicate and to maintain a 'personal space', they co-operate by following (implicit) rules of conversational distance that allow individuals to satisfy both to a certain degree. Clearly having *some* such rules is mutually beneficial, compared to the baseline of having none, but again there will be many possible co-operative patterns, some that benefit some individuals more than others.

For complicated social beings, collective bargaining problems will continually arise. This is not just because token bargaining problems of the same type will continually arise but because problems of new and distinct types will arise.<sup>6</sup> For such individuals the evolutionary pressures that fix upon stable co-operative strategies will be slow to act, their effect taking generations. It would therefore be biologically advantageous for a group of individuals to have quick and flexible method of generating, testing and sustaining patterns of mutually beneficial co-operation in response to novel bargaining problems. It would be even more

---

<sup>6</sup> It is this step which plausibly distinguishes the co-ordinating needs of complex social animals, such as humans, from those of other animals, such as wolves, who need to co-ordinate their actions when hunting (for example).

advantageous to be able to generate, test and sustain such patterns that relate to hypothetical, past and future situations, for which a pattern of co-operation has yet to emerge (Gibbard 1990: 71-2).

There are, of course, many ways of influencing the actions and attitudes of others in the name of progress towards (and stability of) mutually beneficial patterns of co-operation. Physical attack, physical restraint, withdrawal of co-operation and gossip are just a few of the ‘sanctioning’ behaviours available to individuals to co-opt the behaviour and attitudes of others (and there is an equally diverse range of ‘encouraging’ behaviours).<sup>7</sup> But such sanctions themselves carry costs and take time to be effective. What would be even more advantageous, therefore, is a system that advertises a preparedness to engage in these robust sanctioning behaviours without actually doing so. The threat of sanctions can itself have a sanctioning effect. Hence such a system will have the beneficial effects of more robust sanctioning, while seldom having to incur the costs.

Moral discourse – the use of moral language – fulfils both these roles. It allows agents to negotiate, sustain and reinforce patterns of co-operation concerning new, hypothetical, past and present bargaining situations *and* allows them to do so without resorting to costly physical and social sanctions (Gibbard 1990: 71-2; Joyce 2006: 115-18). By judging that cloning humans is wrong, for example, I express a willingness to impose these sorts of sanctions on those who fail to share my disapproval of cloning. My judgement is a negotiating position, a step on the long road to co-operative agreement around the issue. It is part of a discussion that, if things go well, will facilitate progress towards a mutually beneficial co-operative strategy on the issue, and thence sustain it.<sup>8</sup> Moral discussion acts as a demilitarized zone where we can discuss adjustments of attitude and action, in response to novel and distant bargaining problems, without resorting to the military option of physical sanction. For individuals with an interest in coming to and sustaining collective agreement in response to bargaining problems, moral discussion is a pragmatically useful option.<sup>9</sup>

---

<sup>7</sup> For a formal account see Boyd and Richerson (1992).

<sup>8</sup> Of course, the issue of human cloning is not an *evolutionary* bargaining problem, but it is an issue over which agreement has a potential to lead to mutual benefit (in terms of increased desire satisfaction) and over which disagreement has a potential to lead to mutual disadvantage.

<sup>9</sup> As, of course, are systems of etiquette, such as those pertaining to the proper use of cutlery. The difference, it seems to me, is that systems of etiquette are arbitrary and open to less negotiation. We could, without criticism, quite easily have agreed to use our cutlery in a different way to the way we actually do (or are supposed to), but



The moral habit allows individuals to pursue this option. The ability to form moral judgements makes possible a particular type of shared linguistic evaluation whose ultimate goal is to produce and sustain mutually beneficial patterns of co-operation. In short, the function of the moral habit is to produce and sustain such patterns through a distinctive, linguistic, means of collective problem-solving, viz. moral discussion.

For moral discussion to have this role, the positions taken within it – moral judgements – must have two important features. First, they must reliably tie to the behavioural tendencies of the individuals whose judgements they are. Were a moral judgement not to influence one's own behaviour, then expressing that judgement in discussion would not aid coordination of one's own behaviour with others (Gibbard 1990: 72-3; Joyce 2006: 108-118). Second, they must tie to appropriate avowal in forums of shared discussion directed towards co-ordinated action (Gibbard 1990: 73-4; Joyce 2006: 117-123). Were a moral judgement never to be professed or did not have any consequences for one's input into shared evaluative discussion, it would have no chance of influencing the behaviour of others in a negotiation towards mutually beneficial patterns of action. Hence, a moral judgement is, roughly, a motivational state involving a disposition to adhere to a particular pattern of action and a preparedness to avow that state in shared discussion, with the aim of getting others to share it (Gibbard 1990: 75).

According to this speculative evolutionary story, the function of the moral habit is to produce and sustain mutually beneficial patterns of co-operative action and attitude. Moral judgements as a whole have the function of producing such patterns. A particular type of moral judgement derives its more particular function from the distinctive role it plays in this process. For example, the particular function of a judgement that  $\Phi$  is wrong is to produce a mutually beneficial (and stable) pattern of co-operation that involves (at least) the non-performance of  $\Phi$ . To do this the judgement must produce attitudes that will sustain this pattern, such as negative attitudes towards  $\Phi$ ing and further negative attitudes towards those who fail to share those negative attitudes. Hence, roughly, judgements of wrongness have the function of steering patterns of co-operation *away* from the action condemned, and thereafter sustaining them, via the mechanism of shared moral discussion. Conversely, judgements of rightness have the function, roughly, of steering patterns of co-operation *towards* the action

---

now that we have the system we do, that is sufficient justification to preserve it ('that's just the way we do things around here').

praised, and thereafter sustaining them, again via the mechanism of shared moral discussion. In general, there will be as many types of moral judgements as there are distinct roles that such judgements can play in the process of negotiating towards and sustaining shared patterns of mutually beneficial co-operation.<sup>10</sup>

## 4. Moral Content

We have at hand two teleological theories about how function determines (descriptive) content. We also have a story about the function of moral judgements. We are therefore in a position to use both to determine the issue of moral content.

### 4.1. Directive Content of Moral Judgements

Begin with what is common ground between many versions of input and output-based teleosemantics: their view of directive content (**TS3**). According to this view, a state has directive content  $p$  if it has the function of bringing it about that  $p$ . Clearly, on the above evolutionary story, moral judgements have a directive function. The function of the moral habit, and hence the general function of all moral judgements, is to produce stable mutually beneficial patterns of co-operation. The function of a particular type of moral judgement is derived from this more general function. The function of a judgement of the form ‘ $\Phi$  is wrong’ for example, is to produce a stable pattern of co-ordination where no one  $\Phi$ 's, and relatedly, where people disapprove of  $\Phi$ ing and disapprove of those who fail to disapprove of  $\Phi$ ing. In other words, judgements of the form ‘ $\Phi$  is wrong’ have the directive function of bringing it about that no one  $\Phi$ 's and that everyone disapproves of  $\Phi$ ing and everyone disapprove of those who fail to share this disapproval. Therefore, according to **TS3** a judgement that  $\Phi$  is wrong represents the world so as to be made that: *no-one does  $\Phi$* ; and also so as to be made that: *everyone disapproves of  $\Phi$ , and everyone disapproves of those*

---

<sup>10</sup> Some might worry that the account of the function of moral judgements offered here begs the question in favour of expressivism. In particular, realist versions of descriptivism will claim that the function of moral judgements is to indicate the presence of moral properties. But insofar as this claim purports to be a scientific explanation of the moral habit, it fares badly against the account offered here. First, it is less parsimonious. Second, it provides no clear account of how this ‘tracking’ of moral properties truth might have aided reproductive success, and thus be selected for. See Street (2006: 125-34) for similar arguments against related ‘tracking’ accounts and Enoch (2010) for further discussion.

*who fail to disapprove of  $\Phi$ . Very roughly, assuming that imperatives capture directive content, the directive content of ‘ $\Phi$  is wrong’ can be characterised as: *Stop it! Disapprove of it! Disapprove of those who fail to disapprove!*<sup>11</sup>*

Note, of course, that this account doesn’t require each judgement of the form ‘ $\Phi$  is wrong’ to have selectional history for every value of ‘ $\Phi$ ’. The judgement ‘Cloning is wrong’ had no selectional history before 1700. But even before it acquired a selectional history of its own, the judgement was the product of a system whose function is to produce patterns of co-ordination for mutual benefit, thus had the derived function of facilitating the establishment and sustaining of such patterns. More particularly, the first token of the judgement that ‘Cloning is wrong’ was the product of that *subsystem* of the moral habit that, in its normal functioning, has the function of directing interpersonal negotiation *away* from the action thus condemned. Hence even the very first token judgement of ‘Cloning is wrong’ had the derived particular function of producing a pattern of action that does not involve cloning.<sup>12</sup> By **TS3**, its content was therefore that the world be made so that: *no one clones, and everyone disapproves of cloning and, everyone disapproves of those who fail to disapprove of cloning.*

#### **4.2. Input-based Teleosemantics and the Descriptive Content of Moral Judgements**

The crucial issue at stake between descriptivism and expressivists, however, concerns descriptive content. Consider first input-based teleosemantics, as given by **TS1**. The moral habit, we are assuming, has the function of producing mutually beneficial patterns of co-operation. Particular types of moral judgement derive their functions from the particular role they play in fulfilling this function. At first sight, therefore, it seems as if neither the moral habit nor particular types of moral judgement have an indicator function, and therefore, if they have descriptive content, it is not accounted for by a direct application of input-based teleosemantics.

---

<sup>11</sup> Note the similarity with Stevenson’s (1944: 26) ‘working model’ of an analysis of wrongness as ‘I disapprove of this: do so as well!’. There may seem some oddity in wanting everyone to disapprove of  $\Phi$  and also everyone to disapprove of those who fail to disapprove of  $\Phi$ : if the first clause is satisfied the second is redundant. But having such a state is not odd in the common case where the world is not as one desires it to be.

<sup>12</sup> For the account of derived function deployed here, see Millikan (1984: 39-49).

This type of argument is not unprecedented. Gibbard's focus is on normative judgements in general, of which moral judgements are a specific species, but he makes a similar point. He notes:

[According to the evolutionary sketch given above]...[t]he biological function of the mechanisms underlying our normative capacities is to coordinate. Hence the psychic mechanisms that produce normative judgements are not systems of natural representation, they are coordinating systems. Their biological function is not to put something in the head in correspondence with their subject matter; it is to coordinate what is on one person's head with what is in another's. Therefore the judgements are not natural representations but something else. (1990: 110)

Of course, this argument is too quick, as Gibbard realises. For one way in which a set of judgements might help co-ordinate is *by* corresponding with states of affairs. More generally, states can fulfil the function of producing some state of affairs partly *in virtue of* fulfilling the function of indicating some environmental feature. Recall the waggledance. This has the function of producing the movement of 'consumer' bees towards nectar, in virtue of having the function of indicating the presence of nectar. This is just what PP representations do: they directly translate the shape of the environment into the shape of a certain kind of conforming action. So noting that moral judgements function to produce co-ordination doesn't rule out that they also have the function of indicating some environmental feature and hence that they are PP representations.

Gibbard professes that he knows of nothing to rule this possibility out, but he is doubtful:

I myself have found no kind of fact, correspondence to which might plausibly, among our ancestors, have done the co-ordinating job of normative judgements. (1990: 116)

But there is reason to think that this last claim premature. The evolutionary story tells us that moral judgements function to produce certain patterns of mutually beneficial co-operation. Gibbard's claim is that moral judgements need not *also* function to indicate any environmental feature in order to fulfil this co-ordinating function. Note, however, that the function of moral judgements is not *simply* co-ordination: it is co-ordination centred on *mutually beneficial* courses of action (where benefit is cashed out in terms of increased

desire-satisfaction). In simplified terms, moral judgements would not be doing their job if they produced a pattern of co-ordination that was mutually *disadvantageous* to all involved. They do their job, when they do it, only when they produce patterns of co-ordination that are mutually beneficial. And they will produce patterns of co-ordination that are mutually beneficial *only when* they have some sort of tendency to be sensitive to facts about which patterns are of that type and which actions are included in such patterns. Hence, as with PP representations, moral judgements only fulfil their directive function when they fulfil a related descriptive function.

For example, suppose a group of individuals is using moral discussion to negotiate towards a settled pattern of response to the practice of sexual intercourse. Feeling the need to quickly co-ordinate their responses they hit upon a simple strategy: to judge wrong all and only those sexual practices that Anna judges wrong. Adopting this strategy, the group quickly co-ordinate their attitudes and actions concerning sex. Unfortunately, Anna's view is that all sexual relations are wrong. Adopting the strategy therefore secures co-ordination, but is not beneficial to the group: many of their desires go frustrated, and their population plummets. In this case, have the moral judgements fulfilled their function? If the function is *simply* co-ordination, then it seems that they have. But of course the function is more fine-grained than that: it is to produce patterns of co-operative action that are *mutually beneficial*. Hence the moral judgements in this case have not fulfilled their function, since the pattern of action produced is not mutually beneficial (in terms of desire-satisfaction).

This demonstrates that moral judgements will fulfil their co-ordinating function only when they produce patterns of co-operation that are mutually beneficial. And they will produce patterns of co-operation that are mutually beneficial only when they are in some way sensitive to which facts about which patterns of action are mutually beneficial. More particularly, they will produce patterns of co-operation that are mutually beneficial *only when* judgements of wrongness tend to correspond to actions that are not part of mutually beneficial patterns, and when judgements of rightness tend to correspond to actions that *are* part of mutually beneficial patterns of action.

Return to the question: What did judgements of wrongness do in the past that explains their continued survival and proliferation? What is their function? The answer is that they produced co-operative patterns of action that were *mutually beneficial*, by pushing agents' behaviour away from actions that are part of mutually disadvantageous patterns of action.

They did this latter by indicating which actions were part of such *disadvantageous* patterns, and pushing interpersonal negotiation away from such actions. Moral judgements of wrongness thus have two functions: first, producing (in a particular way) mutually beneficial patterns of action and second, indicating which actions are not part of such patterns. In simple terms, a moral judgement that  $\Phi$  is wrong is a representation with the descriptive content:  *$\Phi$  is part of a pattern of action that is mutually disadvantageous.*<sup>13</sup> Since we have already seen that judgements of wrongness are *also* directive representational, their full representational content can be given, very roughly, as:  *$\Phi$  is part of a pattern of action that is mutually disadvantageous; Stop it!*

It seems, then, that Gibbard was premature to suggest that no indicating function is involved in the co-ordinating function of moral judgements. Moral judgements function to co-ordinate *around mutually beneficial patterns of action* partly by indicating which actions are and which are not part of such patterns. In fact, the proper performance the indicating function is a Normal condition of the proper performance of the co-ordinating function. According to input-based teleosemantics, therefore, moral judgements have both descriptive and directive representational content. Moreover, since neither component of this content is available in mediate inference (theoretical or practical), it is not the content of psychologically distinct mental states (Millikan 1995: 191-2). As in the case of the hens' call to their chicks, where a moral judger sense beneficial patterns, there she will go. Moral judgements are PP representations.

### **4.3. Output-based Teleosemantics and the Descriptive Content of Moral Judgements**

Consider, next, output-based teleosemantics (**TS2**). We are assuming that moral judgements have the function of producing mutually beneficial patterns of co-operation. Moral judgements, therefore, have the function of guiding their consumer mechanisms (in this case, mechanisms responsible for behaviour) towards the fulfilment of that same function. Moral judgements 'guide' these consumer mechanisms in a particular way. Consider judgements of wrongness. The consumer mechanism has the function of generating mutually beneficial patterns of co-operation, and as guided by a judgement of wrongness concerning  $\Phi$  it

---

<sup>13</sup> Compare Railton's 'reforming definition' of wrongness (1986: 189-90). Given footnote 11, this view is a synthesis of Stevenson and Railton.

produces actions designed to push co-operative negotiation away from a pattern involving  $\Phi$ . Such actions will include personal avoidance of  $\Phi$ , expressions of disapproval directed at  $\Phi$ , taking negative positions in moral discussion about  $\Phi$  and so on. Under what past condition did these consumer mechanisms, as so guided by judgements of wrongness, fulfil their function? These mechanisms would, by these actions, have produced beneficial patterns of co-operation only so long as judgements of wrongness pushed collective negotiation *away* from actions that are part of disadvantageous patterns of co-operation. Hence the Normal condition for the fulfilment of the consumer's function in this case would have been an appropriate type of mapping between judgements of wrongness and facts concerning which actions are part of mutually disadvantageous patterns of co-operation. Hence output-based teleosemantics assigns judgements of wrongness the descriptive content:  *$\Phi$  is part of a pattern of action that is mutually disadvantageous*. The full representational content, including the directive content generated by **TS3**, will be, very roughly:  *$\Phi$  is part of a pattern of action that is mutually disadvantageous; Stop it!* Input and output-based here agree about the representational content of moral judgements.

This point about moral content according to output-based teleosemantics can be made another way. The function of moral judgements is to cause agents' behaviours to stably coalesce around mutually beneficial patterns of co-operation. Assume, what is reasonable, that this is the only effect moral judgements have been selected for. Then, by **TS3** they are directive, saying something like: *Do this!* and *Stop that!* But it is also a Normal condition on the proper performance of this function that positive moral judgements are applied to actions that are part of mutually beneficial patterns – call these 'pro-social actions' – and negative moral judgements are applied to actions that are not so part – call these 'anti-social actions'. So the judgements are also descriptive, by **TS2**, saying something like: *This is a pro-social action* and *That is an anti-social action*. Assume further, that the effect of the judgement on the agent is not filtered through an all-purpose cognitive mechanism that operates first by forming a descriptive representation (belief that an action is pro-social), then retrieving a relevant desire (the desire to act pro-socially), then performing a practical inference and finally acting on the conclusion. Rather the judgement connects directly with action. *Its function is to mediate the production of a certain kind of behaviour such that it varies as a direct (mathematical) function of a certain variation in the environment, thus directly translating the shape of the environment into the shape of a certain kind of conforming action:*

where an agent finds pro-social actions, there he (and, hopefully, others) will go. Thus moral judgements are PP representations.<sup>14</sup>

Note that, unlike in the case of input-based teleosemantics, Gibbard's question – concerning whether or not moral judgements 'correspond' to some worldly affair – doesn't arise. On output-based teleosemantics the issue concerns not whether judgements of wrongness, say, ever indicate or correlate with the presence of anti-social actions, but rather whether the presence of anti-social actions is a Normal condition on proper (directive) functioning of such judgements. Normal conditions, remember, may be rare. Hence output-based semantics does not require a correlation between judgements of wrongness and anti-social actions. It only requires that *when* judgements of wrongness co-ordinate in the way they Normally do, they are applied to anti-social actions. So long as the Normal conditions are frequent enough to explain selection, they need not be common.

## 5. Consequences for the Debate between Descriptivism and Expressivism

### 5.1. Direct Application of Teleosemantics

How does the preceding account of moral content affect the debate between descriptivism and expressivism?

Consider first descriptivism. Given the foregoing evolutionary account, on both versions of teleosemantics, moral judgements have descriptive content. But this does not vindicate descriptivism. Descriptivism, recall, holds that:

**D.** Moral judgements involve a descriptively representational element that represents the world as being a certain moral way, i.e. moral content is descriptive content.

For example, descriptivists hold that the judgement that murder is wrong represents the world as being such that *murder is wrong*, where the italicised sentence gives the descriptive content of the state. Given the above evolutionary sketch, the judgement that murder is wrong, meets a *sufficient* condition for having a particular descriptive content, namely the content: *murder is part of a pattern of action that is mutually disadvantageous* (this applies, I have

---

<sup>14</sup> This paragraph is a deliberate adaption of Millikan's account of hens' food calls, quoted above.



argued, whether the sufficient condition is given by **TS1** or **TS2**). This content is not moral content, as tests of a Moorean nature show (see Moore 1903). An agent competent with all the relevant concepts could rationally assent to one content but not the other.<sup>15</sup> Thus a direct application of either input-based teleosemantics (**TS1**) or output-based teleosemantics (**TS2**) fails to support the view that moral content is descriptive content, as descriptivists claim (for the same conclusion, see Gibbard 1990: 113-17).

It is important to note the argumentative position here. Since **TS1** and **TS2** provide only *sufficient* conditions for a state having particular descriptive content, the argument is not that these theories show descriptivism (as characterised by **D**) to be *false*. To show *that*, **TS1** and **TS2** would need to be rephrased as providing *necessary* conditions on descriptive content and it would need to be shown that moral content does not meet this conditions. But since **TS1** and **TS2** do not provide necessary conditions on descriptive content, all this argument shows so far is that *if* moral content is descriptive content, it cannot be content that is accounted for by direct application of **TS1** or **TS2**. If descriptivism is true, teleosemantics cannot yet explain *why* it is true.

Relatedly, neither does the forgoing teleosemantic-plus-evolutionary account of the content of moral judgements vindicate expressivism. Admittedly, the application of **TS3** – the teleosemanticist account of directive content – does generate something close to the expressivists’ positive claim **E2**, namely the claim that moral judgements have distinctive non-moral *directive* content. On the foregoing view, the judgement that murder is wrong has directive content that can be characterised as: *Stop murdering! Disapprove of murder! Disapprove of those who fail to disapprove of murder!* This is content that represents the world so as to be made a distinctive non-moral way, as **E2** claims. But expressivism also involves the negative claim **E1**, namely that moral content is *not* descriptive content. And this has so far not been shown. For to show that moral content is *not* descriptive content would require showing that there is a *necessary* condition for being descriptive content that moral

---

<sup>15</sup> At this point the descriptivist might appeal to a reduction of moral properties (such as wrongness) to non-moral properties (such as being part of a pattern of action that is mutually disadvantageous) along the lines of the reduction of the property of being water to the property of being H<sub>2</sub>O. But in the present context, such a move is beside the point, for even the moral reductionist accepts that *murder is wrong* does not mean the same as *murder is part of a pattern of action that is mutually disadvantageous*. More generally, a reductivist account of moral *properties* fails to speak to the present issue, which concerns the nature of moral representational *content*.

content cannot meet, and this – to repeat – has not been shown, since **TS1** and **TS2** present only *sufficient* conditions for a state to have particular descriptive content.

The conclusion so far, then, is this. A direct application of teleosemantics to moral judgements of wrongness (given the evolutionary sketch of the function of those judgements) shows that they have particular non-moral descriptive content *and* particular non-moral directive content. But, since teleosemantics provides only sufficient conditions, it does not show that this is the *only* content these moral judgements have, hence does not pronounce on the issue of whether such judgements also have *moral* content that is descriptive.

## 5.2. Descriptivism and Modest Teleosemantics

Can teleosemantics provide any further traction on the issue of whether moral judgements possess *moral* descriptive content? The expressivist might argue that, having shown moral judgements to possess both non-moral descriptive and non-moral directive content, there is simply no explanatory need to posit any further (descriptive) moral content. Such an argument would be arduous, however, since it would require detailed consideration both of what needs to be explained in this context, and whether only this narrow attribution of contents can explain it.

Perhaps a stronger case for expressivism might be possible if we could show that, having first ruled out a direct application of **TS1** and **TS2**, there is no other way in which moral descriptive content could arise. The thought would be this: we have seen that neither **TS1** nor **TS2** help us explain how moral judgements could have *moral* descriptive content. But, on any plausible account of how **TS1** or **TS2** are extended to provide accounts of all possible forms of representational content (that is, extended so as to provide a necessary as well as sufficient conditions for descriptive content), there is also no way there can be moral descriptive content. If this is right, teleosemantics not only fails to vindicate, but actually undermines, the existence of moral descriptive content, and thus supports expressivism over descriptivism. In the rest of this section, I tentatively pursue this line of argument.

How, other than through meeting the sufficient conditions set out in **TS1** and **TS2**, might descriptive content arise? We have seen one answer to this question already in the discussion of the no- and co-reference problems. On the modest teleosemantic view described in §2.5, descriptive content that is not directly teleologically explicable can only arise by

being constructed from inferential ties to directly teleologically explicable content (that is, content possessed by states that meet a sufficient condition specified by **TS1** or **TS2**).<sup>16</sup> This suggests a possible strategy for descriptivists in the moral case: to claim that moral content is complex descriptive content constructed, though inferential ties, from the basic content generated by a direct application of the teleological theories **TS1**, **TS2** and/or **TS3**.

It seems, however, that such views face a dilemma, which arises by considering this question: What is the nature of the *basic* (i.e. directly teleologically explicable) content, inferential ties to which might generate *moral* descriptive content?

There appear to be only two reasonable answers.

First, moral descriptive content might be generated by inferential ties to basic non-moral *descriptive* content (that is, content arising from application of **TS1** or **TS2**). The problem with this approach is that it is difficult to see how any dossier of *non-moral, descriptive* content can capture the *normative force* of moral content. Moral content is normative content, that is, content that serves to recommend or discourage courses of action or emotion (it is, to borrow from Sellars, content that is ‘fraught with ought’). As Joyce (2007: 58) puts it, moral content has ‘practical clout’. However, any descriptive content accounted for by teleosemantics seems to lack such clout. Describing an action as fitness-enhancing, deadly, painful or arduous does not draw attention to a deliberative consideration that cannot be ignored, since I may not care for offspring, death, freedom from pain or hard work. The problem is that if moral content is normative content, and if any non-moral descriptive content of which teleosemantics might be true is not normative content, it is difficult to see how the former can be the product of any inferential ties to the latter. The ingredients with which to cook up moral descriptive content are of simply the wrong sort.<sup>17</sup>

Second, moral descriptive content might be generated by inferential ties to basic non-moral *directive* content (that is, content arising from application of **TS3**). On this view descriptive moral content is generated by inferential ties to *directive* non-moral content

---

<sup>16</sup> Strictly speaking, it is contentful mental states, not their contents, that stand in inferential relations. For ease of reading I here use the slightly looser way of speaking.

<sup>17</sup> Note that this argument applies however normativity is understood. In other words, whatever the difference between normative and non-normative content, it is difficult to see how the former can be concocted from (inferential ties to) the latter. In this section I assume that normative content is content that ‘cannot be ignored’, but this is not necessary for the (admittedly presumptive) argument to go through here.

(Wedgwood 2007). The problem with this approach is that it is mysterious how a tie to any *directive* content can generate new type of *descriptive* content. In the most common cases this is not how new descriptive content arises. For example, the content: *that's a unicorn* is generated by inferential ties to other descriptive content, such as *that's a quadruped* and *that's a horned creature*. It is easy to see how (and common to assume that) connections to *further* descriptive content can enrich, even create, descriptive content. But it is difficult to see how connections to further *directive* content can do the same.

Neither horn of this dilemma is fatal. Both represent challenges that descriptivists need to address rather than decisive refutations. On the first horn the challenge is showing how moral (normative) descriptive content can arise from inferential ties to non-moral (non-normative) descriptive content. On the second horn the challenge is showing how moral descriptive content (indeed, any descriptive content) can arise from inferential ties to non-moral *directive* content. Expressivists, on the other hand, face neither challenge, since their account of moral content is exhausted by the observation that moral judgements are states with distinctive non-moral contents that play a distinctive role in the interpersonal coordination of attitude and action (Blackburn 1998: 68-70).

The conclusion of this section is therefore as follows. The view I have called modest teleosemantics provides a glimpse of how we might turn the sufficient conditions for descriptive (and directive) content given by **TS1**, **TS2** and **TS3** into necessary conditions. Though promising, the modest theory is incomplete and there may be alternatives. But if it is on the right track, it spells problems for descriptivism. For it requires that moral descriptive content is constructed from inferential ties to either non-moral descriptive content or non-moral directive content. In both cases, however, it is mysterious how such ties can provide content that is both descriptive and carries normative force.

## **6. Conclusion**

I have argued that combining teleosemantics with a plausible account of the function of moral judgements generates the conclusion that such judgements have both descriptive and directive representational content. These contents are not, at least initially, inferentially distinct in the minds of moral judges. Thus moral judgements are PP representations.

This account of moral content vindicates neither descriptivism nor expressivism. It fails to vindicate descriptivism because the directly teleologically explicable descriptive content of moral judgements is *not* moral content. It fails to vindicate expressivism because there is, at least on modest versions of teleosemantics, more to descriptive content than directly teleologically explicable content. Nevertheless, there are further obstacles to the theory that moral content is descriptive content. For if moral content isn't *basic* descriptive content, it must be constructed from inferential ties to basic content. But neither inferential ties to basic *descriptive* content, nor inferential ties to basic *directive* content seem apt to generate distinctive moral descriptive content. In light of this, although the teleological and functional theories presented here do not provide conclusive support for expressivism, they should make us lean in its favour.

## References

Blackburn, S. 1993. *Essays in Quasi-Realism*. Oxford: Oxford University Press.

———. 1998. *Ruling Passions*. Oxford: Clarendon Press.

Boyd, R. 1988. 'How to be Moral Realist' in G. Sayre-McCord, ed., *Essays on Moral Realism*. Ithaca: Cornell University Press. 181-228.

Boyd, R & Richerson, P.J. 1992. 'Punishment Allows the Evolution of Co-operation (or Anything Else) in Sizable Groups' *Ethology and Sociobiology* 13: 171-95.

Brink, D. 1989. *Moral Realism and the Foundations of Ethics*. Cambridge: Cambridge University Press.

Dretske, F. 1986. 'Misrepresentation' in R.J. Bogdan (ed.), *Belief: Form, Content and Function*. Oxford: Oxford University Press. 297-314.

———. 1988. *Explaining Behaviour*. Cambridge, Ma.: MIT Press.

———. 1995. *Naturalising the Mind*. Cambridge, Ma.: MIT Press.

Enoch, D. 2010. 'The Epistemological Challenge to Metanormative Realism: How Best to Understand It, and How to Cope with It', *Philosophical Studies* 148: 413-38.

Gibbard, A. 1990. *Wise Choices, Apt Feelings: A Theory of Normative Judgement*. Oxford: Oxford University Press.

———. 2003. *Thinking How to Live*. Cambridge, Ma.: Harvard University Press.

Harms, W.F. 2000. 'Adaptation and Moral Realism', *Biology and Philosophy* 15: 699-712.

Joyce, R. 2001a. *The Myth of Morality*. Cambridge: Cambridge University Press.

———. 2001b. 'Moral Realism and Teleosemantics', *Biology and Philosophy* 16: 725-34.

———. 2006. *The Evolution of Morality*. Cambridge, Ma.: MIT Press.

Lloyd, D. 1989. *Simple Minds*. Cambridge, Ma.: MIT Press.

Macdonald, G. and Papineau, D. (eds.) 2006. *Teleosemantics*. Oxford: Clarendon Press.

Millikan, R. 1984. *Language, Thought, and other Biological Categories*. Cambridge, Ma.: MIT Press.

———. 1989. 'Biosemantics' *Journal of Philosophy*, 86: 281-97.

———. 1995. 'Pushmi-Pullyu Representations' in J. Tomberlin, ed., *Philosophical Perspectives: Volume IX, AI, Connectionism and Philosophical Psychology*. 185-200.

Moore, G.E. 1903. *Principia Ethica*. Cambridge: Cambridge University Press.

Neander, K. 1995. 'Misrepresenting and Malfunctioning' *Philosophical Studies* 79: 109-41

———. 2006. 'Content for Cognitive Science' in Macdonald, G. and Papineau, D. (eds.), *Teleosemantics*. Oxford: Clarendon Press. 167-94.

Papineau, D. 1984. 'Representation and Explanation' *Philosophy of Science* 51: 550-72.

———. 1987. *Reality and Representation*. Oxford: Basil Blackwell

———. 1993. *Philosophical Naturalism*. Oxford: Basil Blackwell

Price, C. 2001. *Functions in Mind: A Theory of Intentional Content*. Oxford: Oxford University Press.

Railton, P. 1986. 'Moral Realism', *Philosophical Review* 95: 163-207.

Shafer-Landau, R. 2005. *Moral Realism*. Oxford: Oxford University Press.

Smith, M. 1994. *The Moral Problem*. Oxford: Basil Blackwell.

Sterelny, K. 1990. *The Representational Theory of Mind*. Oxford: Basil Blackwell.

Stevenson, C.L. 1937. 'The Emotive Meaning of Ethical Terms', *Mind* 46: 14-31.

———. 1944. *Ethics and Language*. New Haven and London: Yale University Press.

———. 1963. *Facts and Values*. New Haven and London: Yale University Press.

Street, S. 2006. 'A Darwinian Dilemma for Realist Theories of Value', *Philosophical Studies* 127: 109-166.

Sturgeon, N. 1992. 'Nonmoral Explanations' in Tomerlin, J. (ed.) *Philosophical Perspectives* 6: *Ethics*. Atascadero: Ridgeview. 97-117.

Wedgwood, R. 2007. *The Nature of Normativity*. Oxford: Oxford University Press.