

Erica Roedder and Gilbert Harman

Revised October 30, 2008

## Grammar<sup>1</sup>

Analogies are often theoretically useful. Important principles of electricity are suggested by an analogy between water current flowing through a pipe and electrical current “flowing” through a wire. A basic theory of sound is suggested by an analogy between waves caused by a stone being dropped into a still lake and “sound waves” caused by a disturbance in air. At the very least, analogies suggest questions and shape the direction of research; at most, analogies may be crucial to explaining and understanding natural phenomena (Holyoak and Thagard 1995). Of course, principles suggested by such analogies need to be confirmed by relevant evidence. Even where analogies are fruitful, they are only partial. Sound waves and electricity are not wet; electricity does not spill out when a wire is cut.

In this chapter we consider what is suggested by taking seriously an analogy between language and morality. Recently there have been a number of striking claims made about such a linguistic analogy – claims that if true, have profound implications for longstanding debates in moral philosophy about the innateness of moral capacities, and the existence of moral universals. (For example, the title of Hauser (2006) is *Moral minds: how nature designed a universal sense of right and wrong*.) We will indicate what support pursuit of the linguistic analogy might provide for such claims. Perhaps more importantly, we will discuss implications the linguistic analogy might have for the methodology and structure of moral theory.

---

<sup>1</sup> We have benefited from comments on an earlier version of this chapter from members of our moral psychology research group at a workshop in Pittsburgh November 2007 and more recently from detailed comments from Stephen Stich.

## *Moral grammar*

Rules of morality have sometimes been described as analogous to rules of grammar with the occasional further suggestion that moral theory may be at least in some respects analogous to linguistic theory.

Adam Smith appeals to this analogy when he says,

The rules of justice may be compared to the rules of grammar; the rules of the other virtues, to the rules which critics lay down for the attainment of what is sublime and elegant in composition (Smith 1817, 281).<sup>2</sup>

More recently, starting in the 1950s (Chomsky 1957, 1965), a few moral theorists began to see analogies between moral theory and generative linguistic theory. According to John Rawls,

[O]ne may think of moral theory at first (and I stress the provisional nature of this view) as the attempt to describe our moral capacity; or, in the present case, one may regard a theory of justice as describing our sense of justice. ... A conception of justice characterizes our moral sensibility when the everyday judgments we do make are in accord with its principles. ... We do not understand our sense of justice until we know in some systematic way covering a wide range of cases what these principles are.

A useful comparison here is with the problem of describing the sense of grammaticalness that we have for the sentences of our native language [here

---

<sup>2</sup> We are indebted here to Richard Holton.

Rawls refers to Chomsky (1965), pp. 3-9]. In this case the aim is to characterize the ability to recognize well-formed sentences by formulating clearly expressed principles which make the same discriminations as the native speaker. This undertaking is known to require theoretical constructions that far outrun the ad hoc precepts of our explicit grammatical knowledge. A similar situation presumably holds in moral theory. There is no reason to assume that our sense of justice can be adequately characterized by familiar common sense precepts, or derived from the more obvious learning principles. A correct account of moral capacities will certainly involve principles and theoretical constructions which go much beyond the norms and standards cited in everyday life ... (Rawls, 1971, section 9).

Moreover, the appeal of the linguistic analogy has not been limited to philosophers in the Western “analytic tradition.” Pope Benedict XVI uses a similar analogy in his message for the 2007 World Day of Peace:

The transcendent “grammar”, that is to say the body of rules for individual action and the reciprocal relationships of persons in accordance with justice and solidarity, is inscribed on human consciences, in which the wise plan of God is reflected. As I recently had occasion to reaffirm: “we believe that at the beginning of everything is the eternal word, reason and not unreason (4).” Peace is thus also a task demanding of everyone a personal response consistent with God's plan. The criterion inspiring this response can only be respect for the “grammar” written on

human hearts by the divine creator (Benedict 2006).<sup>3</sup>

The authors above are explicitly concerned with analogies between principles of grammar and principles of justice. In addition, other authors invoke an analogy between rules of grammar and more general moral principles.

For instance, three years before Rawls (1971), Nozick (1968) offered an account of “moral complications and moral structures” in general, not limited to principles of justice. Nozick (1997: 4-5) reports that one “model [for the 1968 paper] was Noam Chomsky’s *Syntactic Structures*, and I meant ‘Moral Complications’ to begin to uncover the structure underlying our moral judgments.” Nozick (1968) mentions that “one needs a distinction similar to that which linguists make between linguistic competence and linguistic performance,” referring to Chomsky (1965), a distinction we discuss below.

As another example, Kamm (2007) describes the methodology she uses in that book and in Kamm (1992, 1993, and 1996) as follows:

[P]eople who have responses to cases are a natural source of data from which we can isolate the reasons and principles underlying their responses. The idea [is] that the responses come from and reveal some underlying psychologically real structure, a structure that was always (unconsciously) part of the thought processes of some people. Such people embody the reasoning and principles (which may be thought of as an internal program) that generates these responses. ... (Unlike the deep structure of the grammar of a language, at least one level of

---

<sup>3</sup> Here we are indebted to Liberman (2006).

the deep structure of moral intuitions about cases seems to be accessible upon reflection by those who have intuitive judgements ...) If the same “deep structure” is present in all persons—and there is growing psychological evidence that this is true (as in the work of Professor Marc Hauser)—this would be another reason why considering the intuitive judgements of one person would be sufficient, for each person would give the same response.

We return below to the issue whether the same moral deep structure is present in all persons. At the moment we are content to point to Kamm’s acceptance of an analogy between the grammar of a language and an individual’s moral principles.

Finally, Mikhail (2000, 2007) provides an account of “universal moral grammar,” based on legal categories. We eventually return to the idea of a *universal* moral grammar. But we want first to consider what proposals about “moral grammar” are suggested by the analogy with the grammar of a language, appealing to developments in contemporary linguistics starting with Chomsky (1957, 1965).

## **I-Grammar**

The primary object of study of linguistics so conceived is not language in the ordinary sense in which English, German, Mohawk, and Chinese are languages. Any ordinary language in this sense has different dialects that blend into one another in ways that do not correspond to national or geographical boundaries. There is a well known saying that “a language (in the ordinary sense) is a dialect with an army and a navy.” What counts as a dialect of French rather than Flemish is a social or political issue, not an issue in the science of language. It may be that any

two speakers have at least somewhat different dialects, with at least somewhat different vocabularies. Chomsky (2000) and other linguists (e.g., Freidin 1992, Isac and Reiss 2008) are concerned with language as a property of a particular person, assumed to be abstractly specifiable by an internal grammar, or I-grammar.

This is one of a number of significant idealizations or abstractions linguistics makes in the course of arriving at a theoretically useful account of an otherwise unwieldy phenomenon. Others include the distinction, mentioned below, between “competence” and “performance” and a more controversial distinction some linguists make between “core” and other aspects of grammar.

In pursuing possible analogies between linguistics and moral theory, we might consider whether similar idealizations or abstractions are theoretically useful to moral theory. Should moral theory be in the first instance concerned with an individual’s I-morality—that *particular* person’s moral standards? This might be captured by some sort of (possibly unconscious) abstract representation, an I-moral-grammar. I-morality in this sense would be distinguished from the principles or conventions of a social group; it would also be distinguished from the correct moral principles, if there are any.

### **“Generative” grammar**

A generative grammar of a language (a generative I-grammar) would fully and explicitly characterize the relevant linguistic properties of expressions of the language (Chomsky 1965, Freidin 1992). It is very important to recognize that the word “generative” is here used as a technical term. There is no implication that the principles of a generative grammar are followed

by speakers in generating (in a more ordinary sense) what they say. To suppose that there is a passive “transformation”—a function that associates the structure underlying an active sentence like *Bill gave Mabel a book* with the structure underlying a passive sentence like *Mabel was given a book by Bill*—is not necessarily to suppose that a speaker first forms the active structure and then, by some causal process, converts it into the passive structure.

A generative grammar of a particular person’s language would specify the linguistic *structure* of expressions of that language, indicating the nouns, verbs, adjectives, prepositions, etc., as well as the phrases of which these are the “heads,” as when a noun is the head of a noun phrase and a preposition the head of a prepositional phrase. A generative grammar might specify important aspects of the meanings of expressions, depending on how the expressions are structured and possibly indicating the scope of quantifiers and other such operators. The grammar would also specify aspects of the pronunciation of expressions (for spoken languages—something similar would be true of the grammar of a sign language). So, a generative grammar would relate pronunciations of expressions to possible interpretations, in this way indicating certain sound-meaning (phonetic-semantic) relationships.

An analogous moral grammar would attempt explicitly to characterize an individual’s moral standards. Just as a grammar specifies the structure of a well-formed linguistic sentence by using a specialized linguistic vocabulary, an I-moral grammar might specify the structure of impermissible actions, using a specialized moral vocabulary<sup>4</sup>. That is, just as English grammar might specify how a noun phrase and verb phrase combine to form a grammatical sentence, an I-

---

<sup>4</sup>Mikhail (2000, see also 2007) suggests that relevant aspects of structure may be arrived at through using Goldman’s (1970) theory of action and legal concepts of battery, end, means, side effect, intention, etc. Abstractly considered, the resulting structures are somewhat similar to the phrase structures that figure in linguistic theory.

moral grammar might specify that that certain actions, in virtue of their structure, are impermissible, e.g. intentional harm directed towards an innocent person without reason. In this way, an I-moral grammar might specify certain action-assessment relationships—or, indeed, character-assessment or situation-assessment relationships. Thus the rules of an I-moral grammar would generate a set of impermissible actions, just as a linguistic grammar generates a set of grammatical sentences. An I-moral grammar would have to specify the relevant structures of the objects of moral assessment (e.g., acts, character traits, situations, etc.) using a specialized moral vocabulary, perhaps including terms like ought, obligation, duty, justice, fairness, right, wrong, responsibility, excuse, justification, etc.

Relatedly, principles of generative grammar specify how the structure of a sentence might be transformed, grammatically, into another. For instance, the structure underlying a declarative sentence *The book is on the chair* is transformable into the structure underlying a question, *What is on the chair?* Characterized abstractly, such principles allow the theoretician to move from the known grammaticality of one sentence structure to the grammaticality of a transformed sentence structure. Similarly, there might be found certain transformative principles within I-morality; we suggest an example of such a principle below.

Given such a picture, it may be tempting to assume certain causal structures. However, this would be a mistake. For instance, let us assume that the moral grammar specifies action-assessment relationships. If this is right, one might assume the following causal structure: first, some non-moral brain system assigns a structural description of the action. Second, the structural description is provided as input to the moral system, and the moral system generates a moral assessment. As in the linguistic case, however, there is no guarantee that the brain's

causal pathways parallels that of the grammatical principles. While the existence of such a moral grammar guarantees that there is, in a mathematical sense, a mapping of action to assessments, the brain's method of processing information may be quite different. (Compare also Marr's (1982) distinction between computational and algorithmic descriptions of mental processing.)

This point about the non-causal character of moral grammars is important for at least two reasons, which we discuss immediately below. The first reason is a theoretical point about recursive embedding; the second reason draws on some recent empirical studies.

### **Recursive embedding**

The rules of a generative grammar specify recursive embedding of grammatical structures. For example, a larger sentence can contain a smaller sentence within it, which can contain another even smaller sentence, and so on without limit:

Jack told Betty that Sally believes that Albert no longer thinks Arnold wants to leave Mabel.

Similarly, the rules of grammar imply that a noun phrase can contain a noun phrase within it, which can contain another noun phrase, and so on:

the man next to the girl near the door with the broken hinge.

So, a particular generative grammar might imply rules that could be represented as follows:

$S \Rightarrow NP + (V + S)$

NP => Art + N + (P + NP)

This point about linguistic recursive embedding is a point about grammatical rules. It is not just that a sentence can occur within a larger sentence or a noun phrase within a larger noun phrase. This sort of “recursion” is common enough: many artifacts can contain smaller things of the same kind, e.g., a large raft can be made of smaller rafts which have been tied together<sup>5</sup>. Thus there is a sense in which recursion is not particularly uncommon. However, there is a substantial difference between recursion found in linguistic grammars and this sort of everyday “recursion.” In the case of linguistics, recursion is an explicit feature of rules which have proved theoretically useful. In contrast, recursion is not explicitly implied by theoretically fruitful rules either for artifacts in general, or for rafts in particular. Were we to construct rules to answer the question, “What objects count as rafts?” the answer would presumably focus on how the object functions (it bears weight as it floats on water), its similarity to certain prototypes, etc. without making use of recursion.

The analogy between morality and language suggests that a “generative moral grammar” might also imply recursive embedding. It might, for example, imply that *it is wrong to encourage someone to do something that is wrong*, which is recursive in the sense that it implies that it is wrong to encourage someone to encourage someone to do something that is wrong, etc. Similarly, a generative moral grammar might imply that it is wrong to promise to do something that it would be wrong to do.<sup>6</sup>

---

<sup>5</sup> We thank Jesse Prinz for bringing this point to our attention and John Doris for the specific example.

<sup>6</sup> Such principles are famously problematic for expressivist theories that try to explain the meaning of *wrong* simply by saying the word is used to express disapproval (Geach 1965).

There are two points to be made about recursive embedding. First, true recursion—as opposed to the natural embedding that occurs with rafts—is an important feature of grammatical rules. This recursive character allows a relatively small number of rules to generate a potentially infinite set of sentences. Recursion is an important, and perhaps distinctive, feature of generative grammars; thus if moral judgments can only be modeled using a recursive theory, that is a point in favor of the linguistic analogy.

Second, recursive embedding emphasizes that, in a generative I-moral grammar, the normative evaluation assigned to structural descriptions can depend on normative evaluations embedded within those structural descriptions. The point does not yet by itself rule out a processing model that first constructs nonnormative action and situation descriptions in terms of causes and effects, ends and means, distinguishing what is done intentionally or not, etc., and then, when that is done, adds normative labels to various parts of the structure, perhaps starting with smaller parts, and then labeling larger parts to handle the recursive aspects of the I-moral grammar. But such a processing model faces the further worry that seemingly nonnormative aspects of relevant descriptive structure themselves seem to depend on normative considerations.

### **Structural dependence on normative assessments**

There is considerable evidence (summarized in Knobe 2008) that whether people accept certain seemingly descriptive claims can depend on whether they accept certain normative claims about side effects of actions. For example, suppose Tom increases his profits by doing something that, as a side effect, either (a) harms or (b) helps the environment, where Tom does not care either way about the effect on the environment. People who think that it is good to help the

environment and bad to harm it tend to say (a) that in so acting Tom intentionally harmed the environment but will tend not to say (b) that in so acting Tom intentionally helped the environment (Knobe 2003; for further discussion, see Knobe & Doris, this volume).

They also tend to say (a) that Tom increased profits *by* harming the environment but tend not to say (b) that he increased profits *by* helping the environment. They tend to say (a) that Tom harmed the environment *in order to* increase profits but tend not to say (b) that he helped the environment *in order to* increase profits (Knobe 2007). And they are more inclined to say (a) that the harming of the environment *was the same thing* as the implementing of the policy than they are to say (b) that the helping of the environment *was the same thing* as the implementing of the policy (Ulatowski 2008).

Furthermore, sometimes whether one will judge that a particular person *caused* a certain result can depend on whether one judges that the person is morally at fault for doing what he or she did (Mackie 1955, Dray 1957, Alicke 1992, Thomson 2003). For example, whether one judges that Tom's omitting to water the plants caused their death can depend in part on whether one thinks Tom was obligated to water the plants.

So, it seems that relevant structural descriptions of the situations or actions must themselves sometimes involve normative assessments.

## **Categorization versus grammar**

Many moral judgments offer a moral categorization of an action, person, or situation. So, instead of drawing an analogy between moral theory and generative grammar, some theorists (e.g. Stich 1993) propose instead to draw an analogy between moral theory and psychological

theories of categorization, especially theories in which an instance to be categorized is compared with a number of standard prototypes (or exemplars) of various categories and is then assigned the category of the prototype to which it is most similar (theories of this sort are described in section 6 in Harman, Sinnott-Armstrong, and Mason, this volume).

We can offer two preliminary comments on the categorization proposal: first, it is not clear that a categorization approach competes with a moral grammar approach. The two approaches may be apples and oranges. Categorization is a mental process, i.e. the term refers to a form of cognitive processing. In contrast, the proposal that there is a mental grammar is not, first and foremost, about mental processes at all. Rather, it is a proposal about the structures of the contents of moral assessments, just as a linguistic grammar is a proposal about the structures of linguistic expressions. It might both be true that there is a moral grammar, and that we make moral judgments by applying certain prototypes. This might be the case if, for instance, all of a person's "murder" prototypes shared certain basic features, e.g. intentional harm of an innocent.

Second, modern theories of concepts and categorization often emphasize the multi-faceted nature of concepts: they are not just prototypes, they also embedded in theories (Murphy 2004, Machery forthcoming). Prototypes and exemplars alone seem insufficient to explain certain features of categorization. If moral concepts also have a theoretical component, as seems likely to be demonstrated empirically, then that would leave the door open to a moral grammar approach as a theory of mental processing.

### **“Competence Performance Distinction”**

In developing a theory of linguistic grammar, Chomsky (1965) distinguishes between

“competence” and “performance”—where these are technical terms that are not used in their ordinary senses. By “competence” Chomsky means an individual’s internalized grammar. He uses the term “performance” to refer to all other factors that may affect the use of language. In Chomsky’s framework an individual’s linguistic intuitions may or may not reflect the individual’s competence, in Chomsky’s sense.

For example, compare the following sentences:

*This is the cheese that the mouse that the cat that the dog that the man owns bothered chased ate.*

*This is the man that owns the dog that bothered the cat that chased the mouse that ate the cheese.*

Intuitively, the first sentence is hard to understand—deviant in some way—while the second is much easier to understand. The relative deviance of the first sentence appears to be the result of “center embedding”. When this center embedding is made explicit, it can appear that the semantic content of the two sentences is similar:

*This is the cheese that the mouse ate.*

*This is the cheese that the mouse {that the cat chased} ate.*

*This is the cheese that the mouse {that the cat [that the dog bothered ] chased} ate.*

*This is the cheese that the mouse {that the cat [that the dog (that the man owns)] bothered chased} ate.*

The intuitive deviance of the first sentence may fail to reflect something about one's linguistic competence in the sense of an internalized grammar. It may instead be due, for example, to limitations of one's parsing strategies, in which case it may be purely the result of performance limitations. The sentence may fit one's I-grammar, but performance factors affect one's ability to comprehend the sentence as fully grammatical.

The linguistic analogy suggests that it might be fruitful to suppose that there is a distinction between moral competence and performance in something like Chomsky's technical sense of the terms "competence" and "performance". As in the linguistic case, we might suppose that other considerations beyond one's assumed I-morality affect moral intuitions. In addition to memory and other processing limitations, moral intuitions might be affected by "heuristics and biases" (Gilovich, Griffin, and Kahneman 2002; see Sinnott Armstrong et al., this volume; also Sunstein 2004, Horowitz 1998, Doris and Stich 2005; Sinnott-Armstrong 2007), prejudices, emotional responses, self-interest, etc<sup>7</sup>. For example, utilitarians such as Greene (2008) have argued that intuitions that appear to conflict with utilitarianism are due to morally irrelevant emotional factors (see also Baron 1993).

Of course, in most cases, performance factors are not distorting; they are an integral part of proper functioning. One needs an internal parsing strategy in order to comprehend what others say, so that strategy is not necessarily distorting, as it may be in the above example. Similarly, *pace* Greene, emotions like disgust may play an important role in getting one to appreciate the moral character of a situation--even though, in other cases, they can lead one astray. Someone with full moral "competence" (that is, with an internalized moral grammar) who is unable to

---

<sup>7</sup> Of course, there is an ongoing debate whether such "biases" should or should not be thought of as bad reasoning. See, for instance, Gigerenzer (2007).

experience relevant emotions may have trouble with moral evaluations in the same way that someone with full linguistic “competence” but with a damaged parser might be unable to understand other speakers. Here we have in mind work on so-called ‘psychopaths’ or other brain-damaged patients who do not exhibit normal emotive responses. Research has shown that such individuals make abnormal moral judgments (Blair 1995, Blair 1997, Koenigs et al). Some of these patients might possess moral ‘competence’ (an internalized moral grammar) but lack certain critical performance capabilities. Of course, whether or not this is a useful way to understand such patients depends on future empirical work. We offer this only as an example of how the distinction between performance and competence might be put into play.

More generally, our point is that in moral theory as in linguistics, we can consider whether it makes sense to postulate a distinction between an individual’s internal moral grammar or moral “competence” in contrast with various other factors that determine the person’s intuitions and actions. For linguistics, the corresponding assumption has been theoretically fruitful and illuminating. With respect to moral theory, the assumption amounts to assuming that there is a distinctive moral competence or I-moral grammar. A contrary view is that one’s morality is determined entirely by performance factors, so that for example one’s emotional responses are actually constitutive of one’s moral view rather than merely enabling or distorting performance factors.

It may be that a distinction between competence and performance only applies to certain parts of morality and not other parts. We come back to this possibility below in considering whether to distinguish aspects of morality we share with other animals and aspects that are distinctively human.

## **Summary of suggested issues about generative moral grammar**

We have so far discussed one idea from linguistics that might be relevant to the theory of morality, namely, generative grammar—an explicit specification of the grammatical properties of expressions of a language. The conception of grammar as generative or explicit in this way has proved quite productive for the study of language and some moral theorists try to see an analogous conception of moral grammar might be productive for the study of morality.

Among the issues such an analogy raises for moral theory are (1) whether the useful unit of analysis for moral theory is an individual's I-grammar in contrast, for example, with the moral conventions of a group; (2) whether and how such a moral grammar might associate structural descriptions of actions, situations, etc., with normative assessments; (3) whether and how the rules of such a moral grammar might involve recursive embedding of normative assessments; and (4) whether it is useful to distinguish moral "competence" from moral "performance," using these terms in the technical senses employed in linguistic theory.

### ***Universal Morality***

We now turn to a related issue about linguistic and moral universals. Features of language are universal if they occur in all natural human languages that children can acquire as first languages. Universal grammar is concerned in part with such linguistic universals and also with limits on ways in which natural human languages can differ. To what extent might it be fruitful to develop an analogous universal moral theory which would seek to describe common features of any moralities children can acquire as first moralities—along with an account of how moralities can differ?

Some people believe that various familiar moral principles are universal in the sense that they are cross-culturally accepted “You should not steal.” “It is wrong to murder or harm others.” “You should not tell lies.” “You should do your part in useful group activities.” “You should help those worse off than yourself.” “You should not treat others as you would not want to be treated.” “You should do unto others what you would want them to do unto you.”

Instead of directly discussing whether such *familiar principles* are universally accepted, we will focus on two other sorts of universality suggested by the analogy with linguistics. (That is, we will not discuss whether everyone, cross-culturally, believe that it is wrong to murder; instead, we will consider universality of the sort that linguists find.) The first sort of universality is the existence of universal *constraints* on possible I-moral grammars; the second is the notion of universal *parameters* (with local variance in parameter settings).

### **Constraints on linguistic grammars**

Universal linguistic grammar is concerned in part with constraints on the rules of particular I-grammars. Such grammars might contain two sorts of rules—*phrase structure* rules for forming phrases out of their parts and *transformations* or *movement* rules for reorganizing these parts (Chomsky 1957, 1965; Freidin 1992; van Riemsdijk and Williams 1986).

For example, an adjective like *green* can combine or “merge” with a noun like *apple* to form a noun phrase *green apple*; a determiner like *a* or *the* can merge with a noun phrase to form a determiner phrase like *a green apple*; a preposition like *from* can merge with a determiner phrase to form a prepositional phrase like *from a green apple*; and so on. These are instances of phrase structure rules. Universal grammar places constraints on the phrase structure rules for a given I-

language. One such constraint implies that, if there are prepositions and so prepositional phrases in which prepositions appear before their objects, then the same must be true of verb phrases and noun phrases: verbs must appear before their objects and nouns before their objects. On the other hand, if instead of prepositions there are postpositions and postpositional phrases in which postpositions appear after their objects, then the same must be true of verb phrases and noun phrases: verbs and nouns must appear after their objects.

Transformations allow items in phrase structures to be rearranged. For example, a transformation (*wh*-movement) might allow the derivation of the structure underlying a relative clause like (2) from something like the structure underlying (1).

(1) *Bob gave which to Bill.*

(2) *which Bob gave to Bill*

A different transformation (topicalization) might allow the derivation of the structure underlying (4) from something like the structure underlying (3).

(3) *Bob gave a book to Bill.*

(4) *A book, Bob gave to Bill.*

Chomsky (1964) observed that there must be constraints of various sorts on such transformations. He proposed an “*A* over *A*” constraint that would rule out moving an item of a given type *A* from a larger phrase of type *A*. That turned out to be overly simple and Ross (1967) proposed to replace it with a number of more specific “island constraints”, where “islands” are structures that items cannot be moved out of. Ross’ constraints include a “coordinate structure

constraint” and a “complex *NP* constraint,” among others; here we describe only his coordinate structure constraint.

Coordinate structures are special cases of the *A* over *A* principle in which an item of type *A* immediately contains two or more coordinate items of type *A*, as in the noun phrase: *a man, a woman, and a child*; or the verb phrase: *kissed Jack or hugged Harry*. These larger phrases themselves immediately contain smaller coordinate phrases of the same type.

Ross’ coordinate structure constraint prevents transformations from moving any coordinate item out of a coordinate structure and also prevents transformations from moving anything contained in a coordinate item out of that item. Suppose, for example, there is a transformation of topicalization that starts with (5) and moves *a book* to the front to yield (6):

(5) *Bob gave a book to Bill.*

(6) *A book, Bob gave to Bill.*

Ross’ constraint would prevent topicalization from moving *a book* from a position in a coordinate structure like (7) to yield (8).

(7) *Bob gave a record and a book to Bill.*

(8) \* *A book, Bob gave a record and to Bill.*

Observe that it is unclear how a coordinate structure constraint of this sort could be learned by children either from explicit instruction or by trial and error. No one had heard about this constraint before Ross (1967), so no one could have told children about it before that time.

Furthermore, children never seem to make errors that consist in violating the constraint, so they don't seem to acquire it by trial and error learning.

One way to explain the above facts about acquisition is to postulate that children are predisposed towards certain linguistic constraints. If there is such a predisposition, however, it would be present in all normally developing children

Of course, it is possible to invent and use an artificial language that is not subject to the constraint. But such a language might not be easily acquired as a first language by children. We might expect them to acquire a variant that is subject to the constraint. Furthermore, we might expect that children not exposed to human language (e.g., because they were deaf) who developed a language among themselves (presumably a sign language) would develop a language that was subject to the constraint. There is evidence that this expectation is correct. Padden (1988) claims that coordinate clauses in American Sign Language (ASL) adhere to the coordinate structure constraint.

### **Constraints on moralities?**

A number of moral philosophers (Foot 1978, Quinn 1993) have suggested that ordinary moral intuitions obey certain non-obvious rules. For instance, it may be that moral intuitions reflect some version of the Principle of Double Effect. Here is one version of that principle:

*Double Effect:* It is worse knowingly to harm one person X in saving another Y when the harm to X is intended as part of one's means to saving Y as opposed to when the harm to X is merely a foreseen unintended side-effect of one's attempt to save Y.

For another example, Thomson (1978) suggests that some ordinary moral intuitions might reflect a principle of the following sort:

*Deflection:* It is better to save a person X by deflecting a harmful process onto another person Y than by originating a process that harms Y.

As with Ross' coordinate structure constraint, these principles are not generally recognized and it is therefore unlikely that they are explicitly taught to children. If children do acquire moralities containing such principles, it would seem that they must be somehow predisposed to do so. If so, we might expect such principles to be found in all moralities that children naturally acquire.<sup>8</sup>

As in the linguistic case, even if there are such constraints, it may be possible to construct more or less artificial moral systems not subject to them—utilitarian moralities for example. But we might then expect that children of utilitarians would not easily and naturally acquire their parents' morality but would acquire a version subject to those constraints. John Stuart Mill (1859, 1863, 1873) might be an example. His version of utilitarianism appears to incorporate various principles that do not fit with the version taught him by his father, James Mill (1828). In particular, John Stuart Mill's views about the importance of personal creativity, self-development, and personal liberty led him to adopt a distinction between higher and lower qualities of pleasure that conflicts with the purely quantitative conception of amount of pleasure characteristic accepted by Bentham and James Mill.

---

<sup>8</sup> The linguistic analogy suggests that this remark is too strong. As we will explain, a principles and parameters approach of the sort we are about to discuss allows for other possibilities.

## **Linguistic principles and parameters**

We now turn to a discussion of linguistic principles and parameters and possible analogies for moral theory.

Baker (2001) begins a highly readable account of current linguistic theory with the story of eleven Navajo Code Talkers. During World War II Allied forces in the Pacific were in difficulty because Japanese cryptographers were able to decode all their messages. The US Marine Corps solved the problem by using the Code Talkers to transmit and receive military communications in Navajo. Japanese cryptographers were unable to “decode” these messages.

Baker notes that this illustrates two important points about languages. On the one hand, Navajo is sufficiently different from English and Japanese that skilled cryptographers were unable to decode the Navajo messages. On the other hand, Navajo is sufficiently similar to English that the Code Talkers could immediately translate messages from English into Navajo and from Navajo back into English. How can we account for both the similarities and the differences? Indeed, how can we account for the fact that a child exposed to a language when young will easily acquire that language, when the best cryptographers are unable to make heads or tails of it?

Part of the answer why a child can so easily learn a language may be that the child is somehow prepared to acquire a language whose syntax is largely determined by the setting of a small number of parameters—a parameter that indicates whether a sentence must have a subject as in English, or not as in Italian; a parameter that indicates whether the head of a phrase precedes its complements, as in English, or follows, as in Japanese; and so on. Baker suggests that it may be possible to give something like a structured periodical table of languages, much as

there is a structured periodical table of chemical elements. (Thus his title, *The Atoms of Language*.)

A child picks up language from its interactions with others; there is no need for explicit teaching. According to the principles and parameters theory, the child uses its limited experience to set a relatively small number of parameters and acquire the “core” syntax of a language. Other “non-core” aspects of syntax, perhaps involving stylistic variation, are learned as exceptions.

Universal grammar includes an account of principles and parameters. To repeat a point mentioned above in connection with universal constraints on rules, it is possible to devise and use a language that is not in accord with the principles and parameters of universal grammar, and people can learn to use that language. But children will not naturally acquire that language in the way they acquire languages that are in accord with universal grammar.

Another aspect of the principles and parameters universal grammar approach in linguistics is that a theorist’s conception of the generative grammar of a given language can be affected by considerations involving other languages and ease of language acquisition. The task of the child learner is easier to the extent that as much as possible of grammar is part of universal grammar and so is built into the child’s innate language faculty. So, for example, instead of detailed rules for noun phrases, verb phrases, adjective phrases, and prepositional or postpositional phrases, the basic theory of phrase structure might become part of universal grammar, perhaps with a single phrase structure rule (*merge X with something to form a phrase*), so that the grammar of a particular language like English would be limited to such matters as whether the head of a phrase (noun, verb, adjective, preposition or postposition) comes before or after its complement. The grammar of a particular language like English might contain no special rearrangement or

transformational rules; instead there might be a single rule of universal grammar (*move X* somewhere) that says something can be moved anywhere, subject to universal constraints (Chomsky 1981, van Riemsdijk and Williams 1986).

One possible parameter might be whether or not the grammar allows such movement or rearrangement. At least one currently existing language appears not to allow this, the Amazonian Pirahã (Everett 2005). Bolender (2007) argues that all or most languages existing more than 40,000 years ago would probably not have allowed such movement.

The linguistic analogy suggests considering whether it is possible to develop a universal theory of *morality* involving relevant principles and parameters.

### **How moralities differ: principles and parameters?**

Moralities differ about abortion and infanticide, about euthanasia, about slavery, about the moral status of women, about the importance of chastity in women and in men, about caste systems, about cannibalism, about eating meat, about how many wives a man can have at the same time, about the relative importance of equality versus liberty, the individual versus the group, about the extent to which one is morally required to help others, about duties to those outside one or another protected group, about the importance of religion (and which religion is important), about the importance of etiquette (and which standards of etiquette), about the relative importance of personal virtues, and so on.

What are the prospects of a principles and parameters approach to explaining these and other moral differences? Perhaps, all moralities accept principles of the form: avoid harm to members of group *G*, share your resources with members of group *F*, etc., where *G* and *F* are parameters

that vary from one morality to another. However, Prinz (2008) points out that such parameters would differ from those envisioned in linguistics by having many possible values, whereas the parameters in linguistics have a small number of values, typically two.

Dworkin (1993) argues that all intuitive moralities take human life to have what he calls a “sacred value” that increases during pregnancy and in childhood, reaching a maximum at the beginning of adulthood, and slowly declining thereafter, so that the death of twenty year old person is worse (“more tragic”) than the death of an infant or someone who has reached old age. However, according to Dworkin, individual moralities differ concerning whether an early embryo or fetus has significant sacred value. Dworkin argues that this difference is what lies behind ordinary (as opposed to theoretical) disagreements about abortion. Perhaps there are different settings of a “sacred value” parameter. In this case there might be only two relevant settings, indicating whether there is significant sacred value in the embryo or fetus immediately after conception.

A somewhat different appeal to parameters can perhaps be found in the work of certain anthropologists who see a small number of ways of organizing societies, with differences in which ways organize which aspects. Fiske (1991, 1992) distinguishes four general types of social organization: communal sharing, authority ranking, equality matching, and market pricing. Shweder et al. (1997) distinguish the “big three” of morality: autonomy, community, and divinity. In these cases, the relevant “parameters” might determine which models fit which aspects of organization in a given society.

However, we will need to have more explicit generative accounts of a variety of moralities before being able to think clearly about moral principles and parameters.

## **Does morality have to be taught?**

Acquisition of language does not require explicit instruction. Children pick it up from the locals even if parents do not know the local language. The linguistic analogy suggests we consider whether morality is like that.

How does a child acquire an I-morality? Does the child pick up morality from others in something like the way in which language is picked up? Does morality have to be taught? Can it be taught?

Perhaps all normal children pick up the local morality much as they pick up the local dialect, whether or not they receive explicit instruction in right and wrong. Children with certain brain defects might have trouble acquiring morality, just as children with certain other brain defects have trouble acquiring language.

Prinz (2008) suggests that moral conventions can be picked up in the way other conventions are, with no need for moral universals, principles and parameters, and other aspects of moral grammar. Similar suggestions were raised as objections against universal grammar in linguistics in the 1960s (e.g. Harman 1967, later retracted in Harman 1973). But today universal grammar is a central part of contemporary linguistics (Baker 2001).

## **Core morality?**

As mentioned above, some linguists distinguish “core” aspects of grammar from other “peripheral” aspects. In their view, the “core” grammar of an I-language is constituted by those aspects of grammar acquired just by setting parameters. Peripheral aspects of the I-language are

special language specific rules and exceptions that are acquired in other ways. Examples might include irregular verbs and special "constructions" as in the sentence *John sneezed the tissue across the table*, where the normally intransitive verb *sneeze* is used in a special way. However, the distinction between core and peripheral aspects of grammar is controversial, with some linguists (Newmeyer, 2005) treating it as a matter of degree and others rejecting it altogether (Culicover and Jackendoff, 2005; Goldberg, 1995).

The linguistic analogy suggests there might be a similar controversy within moral theory about the usefulness of a distinction between core and peripheral aspects of morality, where the core aspects are universal or are anyway acquired via parameter setting and other aspects are acquired in other ways. For example, it might be that people explicitly accept certain moral principles on the basis of authority, much as they might accept certain grammatical principles on the basis of authority. Where some people accept purely on the basis of authority that it is ungrammatical to end an English sentence with a preposition, others might accept purely on the basis of authority that euthanasia is always wrong even to prevent prolonged terrible suffering when someone is dying. This is just an example; we do not know whether a distinction between core and peripheral moral principles will prove to be theoretically useful in thinking about morality.

## **Humans and Other Animals**

Certain aspects of language are similar to aspects of communication systems in other animals—the use of certain sounds and gestures, for example. Other aspects are particular to human language, arguably the existence of recursive embedding rules of grammar of the sort

discussed above. An analogy between language and morality suggests that it might be useful to consider what aspects of morality are shared with animals and what aspects are specific to human moralities.

Recall Fiske's four models of social relations: communal sharing, authority ranking, equality ranking, and market pricing. Bolender (2003) notes that the first three of these models, or something like them, can be found among non-human social animals. Furthermore, only two of Fiske's models, communal sharing and egalitarian equality matching, seem to have existed in most groups of gatherer-hunters, a mode of life "which represents over ninety-nine percent of human existence" (Bolender 2003, p. 242). Bolender offers "evidence that humans do not need hierarchy in groups of 150 or less" (245) so that this form of social relation was little used before the advent of plowing ... [which] forced people to allocate meager resources in large groups" (245). Similarly, there was no previous need for market pricing.

Some questions arise here. Are the moralities of humans who use only these two models (communal sharing and egalitarian equality matching) significantly different from the moralities of social animals who use these models? Do such humans have moralities with anything like the recursive properties mentioned above, so that they accept such principles as that it is wrong to encourage someone to do something that it is wrong to do? (Do nonhuman animals accept such principles?)

One way to test speculation in this area might be to look at the morality of the Pirahã mentioned above, a small group of people with limited contact with others until recently. The Pirahã have a very limited sense of absent items, their language does not appear to use any sort

of quantification or other constructions that involve variable binding operators (Everett 2005; Bolender 2003, 2007). Presumably they cannot even express in their language the principle that it is wrong to encourage someone to do what is wrong. A Pirahã infant who was removed from that culture and brought up by adoptive parents in New Jersey, would acquire normal human language, an ability to use quantification, and similar moral principles to those accepted by others in New Jersey. On the other hand, nothing similar is true of a non-human social animal.

### ***Conclusion***

In this chapter, we have described various issues suggested by the analogy between morality and linguistics. In the first part of the chapter, we discussed what might be involved in a generative "moral grammar." In the second part we discussed what might be involved in a *universal* moral grammar that parallels universal linguistic grammar.

We believe that the issues suggested by such analogies are worth pursuing. In particular, we think that looking at the developed science of linguistics suggests new sorts of research that might be useful in thinking about the complicated structure of morality. First, if moral theory is modeled on linguistics, its primary focus would be on an individual's I-morality. Second, it may be useful to distinguish performance and competence. Third, as in linguistics, moral theory may benefit by distinguishing between the principles characterizing a moral grammar and the algorithms actually used in the mental production of grammatical sentences. Fourth, the linguistic analogy suggests ways to think systematically about how moralities differ from each other and whether there are moral universals. Finally, the analogy frames fruitful questions

regarding innateness of morality, the existence of a core morality, and the comparison of human and animal morality.

In summary, linguistics has successfully developed into a mature science of human language. In considering the linguistic analogy, then, we have been exploring one possible way to develop a mature science of human morality.

### *References*

Alicke, M. D. (1992). "Culpable Causation," *Journal of Personality and Social Psychology* 63: 368-378.

Benedict XVI, Pope. (2006), "The Human Person, the Heart of Peace," Vatican City. Available online at <[http://www.catholic.org/international/international\\_story.php?id=22342](http://www.catholic.org/international/international_story.php?id=22342)>.

Baker, M. C. (2001). *The Atoms of Language: The Mind's Hidden Rules of Grammar*. New York: Basic Books.

Baron, J. (1993). Heuristics and biases in equity judgments: a utilitarian approach. In B. A. Mellers and J. Baron (Eds.), *Psychological perspectives on justice: Theory and applications*, pp. 109-137. New York: Cambridge University Press.

Blair, Robert James (1995) "A Cognitive Developmental Approach to Morality: Investigating the Psychopath," *Cognition*, 57, 1-29.

Blair, Robert James (1997) "Moral Reasoning and the Child with Psychopathic Tendencies,"  
*Personality and Individual Differences*, 26, 731-739.

Bolender, J. (2003). "The Genealogy of the Moral Modules," *Minds and Machines* 13: 233-255.

Bolender, J. (2007). "Prehistoric Cognition by Description: A Russellian Approach to the Upper  
Paleolithic," *Biology and Philosophy* 22: 383-399.

Chomsky, N. (1957). *Syntactic Structures*. The Hague: Mouton.

Chomsky, N. (1964). *Current Issues in Linguistic Theory*. The Hague: Mouton.

Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press

Chomsky, N. (1981). *Lectures on Government and Binding*. Dordrecht: Foris.

Chomsky, N. (2000). *New Horizons in the Study of Language and the Mind*. Cambridge,  
England: Cambridge University Press.

Culicover, P. W. and Jackendoff, R. (2005). *Simpler Syntax*. New York: Oxford University  
Press.

Dray, W. (1957) *Laws and Explanation in History*. London: Oxford University Press.

Dworkin, R. (1993). *Life's Dominion: An Argument about Abortion, Euthanasia, and Individual  
Freedom*. New York: Knopf.

- Everett, D. (2005). "Cultural Constraints on Grammar and Cognition in Pirahã," *Current Anthropology* 46: 621-634.
- Fiske, A. P. (1991). *Structures of Social Life: The Four Elementary Forms of Human Relations*. New York: Free Press.
- Fiske, A. P. (1992). "The Four Elementary Forms of Sociality: Framework for a Unified Theory of Social Relations," *Psychological Review* 99: 689-723.
- Foot, P., (1967). "The Problem of Abortion and the Doctrine of Double Effect." *Oxford Review* 5, 5-15.
- Freidin, R., (1992). *Foundations of Generative Syntax*. Cambridge, MA: MIT Press.
- Geach, P. T. (1965). "Assertion," *Philosophical Review* 69: 449-465.
- Gilovich, T., Griffin, D., and Kahneman, D., editors, (2002). *Heuristics and Biases*. Cambridge, England: Cambridge University Press.
- Gigerenzer, Gerd. (2007). *Gut Feelings*. New York: Viking Penguin.
- Goldberg, A. (1995). *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago: University of Chicago Press.
- Goldman, A. I. (1970). *A Theory of Human Action*. Englewood Cliffs, N.J.: Prentice-Hall.
- Greene, J. (2008). "The Secret Joke of Kant's Soul," *Moral Psychology, Vol. 3: The Neuroscience of Morality: Emotion, Disease, and Development*, W. Sinnott-Armstrong, Ed.,

MIT Press, Cambridge, MA.

Harman, G. (1967). "Psychological Aspects of the theory of syntax," *Journal of Philosophy* 64: 75-87.

Harman, G. (1973). Review of Noam Chomsky, *Language and Mind, Revised Edition*, in *Language* 49: 453-464.

Hauser, M. (2006). *Moral Minds: How Nature Designed a Universal Sense of Right and Wrong*. New York: Ecco Press/HarperCollins.

Holyoak, K. J., and Thagard, P. (1995). Cambridge, MA: MIT Press.

Isac, D., and Reiss, C. (2008). *I-Language: An Introduction to Linguistics as a Cognitive Science*. New York: Oxford University Press.

Kamm, F. (1992). *Creation and Abortion*. New York: Oxford University Press.

Kamm, F. (1993). *Morality, Mortality, Vol. 1: Death and Who to Save from It*. New York: Oxford University Press.

Kamm, F. (1996). *Morality, Mortality, Vol. 2: Rights, Duties, and Status*. New York: Oxford University Press.

Kamm, F. (2007). *Intricate Ethics: Rights, Responsibilities, and Permissible Harm*. New York: Oxford University Press.

Knobe, J. (2003). "Intentional Action and Side Effects in Ordinary Language," *Analysis* 63: 190-

193.

Knobe, J. (2007). "Reason Explanation in Folk Psychology." *Midwest Studies in Philosophy*. 31: 90-106.

Knobe, J. (2008). "Action Trees and Moral Judgment." Unpublished manuscript. University of North Carolina - Chapel Hill.

Koenigs M, Young L, Adolphs R, Tranel D, Hauser M, Cushman F, Damasio A (2007). "Damage to the prefrontal cortex increases utilitarian moral judgments." *Nature* April 19, 2007

Liberman, M. (2006) "Chomsky and the Pope: Separated at Birth?" Posting on *Language Log* December 20, 2006 at <<http://itre.cis.upenn.edu/~myl/languagelog/archives/003940.html>>.

Machery, E., forthcoming, *Doing Without Concepts*, New York: Oxford University Press

Mackie, J. L. (1955) "Responsibility and Language." *Australasian Journal of Philosophy* 33: 143-59.

Marr, D. (1982). *Vision: A computational Approach*. San Francisco: Freeman.

Mikhail, J. (2000). *Rawls' Linguistic Analogy: A Study of the "Generative Grammar" Model of Moral Theory Discussed by John Rawls in A Theory of Justice*. PHD dissertation in Philosophy, Cornell University.

- Mikhail, J. (2007). "Universal Moral Grammar: Theory, Evidence, and the Future," *Trends in Cognitive Sciences*, 11: 143-152.
- Mill, J. (1828). *Essays on Government*. J. Innes: London.
- Mill, J. S. (1859). *On Liberty*. London: J. W. Parker.
- Mill, J. S. (1863). *Utilitarianism*. London: Parker, Son and Bourn.
- Mill, J. S. (1873). *Autobiography*. London: Longmans, Green, Reader, and Dyer.
- Murphy, Gregory M. (2004). *The Big Book of Concepts*. Cambridge: MIT Press.
- Nagel, T. (1973). "Rawls on Justice," *Philosophical Review* 82, 220-234.
- Newmeyer, F. J. (2005). *Possible and Probable Languages*. Oxford: Oxford University Press.
- Nozick, R. (1968). "Moral Complications and Moral Structures," *Natural Law Forum* 13 : 1-50.
- Nozick, R. (1997). *Socratic Puzzles*. Cambridge MA: Harvard University Press.
- Padden, C. A. (1988). "Grammatical Theory and Signed Languages." In F. Newmeyer (Ed.), *Linguistics: The Cambridge Survey* (Vol. II, ). Cambridge, MA: Cambridge University Press.
- Prinz, J. (2008). "Is Morality Innate?" *Moral Psychology, Volume 1: The Evolution of Morality: Adaptations and Innateness*, W. Sinnott-Armstrong, ed. Cambridge, MA: MIT Press: 367-406.
- Roedder, E. and Harman, G. (forthcoming). "Moral Grammar."

- Quinn, W. (1993). *Morality and Action*. Cambridge: Cambridge University Press.
- Rawls, J. (1971). *A Theory of Justice*. Cambridge, MA: Harvard University Press.
- Ross, J. R. (1967). *Constraints on Variables in Syntax*. Cambridge, MA: MIT Ph.D. Dissertation.  
(Published as *Infinite Syntax*, Ablex, 1986.)
- Shweder, R. A.; Much, N. C.; Mahapatra, M. and Park, L. (1997). "The 'Big Three' of Morality (Autonymy, Community, and Divinity) and the 'Big Three' Explanations of Suffering as Well." *Morality and Health*, ed. by A. Brandt and P. Rozen. New York: Routledge.
- Smith, A. (1817). *The Theory of Moral Sentiments*. Philadelphia: Anthony Finley.
- Stich, S. (1993). "Moral Philosophy and Mental Representation," in M. Hechter, L. Nadel & R. Michod, eds., *The Origin of Values*. New York: Aldine de Gruyter: 215-228.
- Thomson, J. J. (1978). "Killing, Letting Die, and the Trolley Problem," *Monist* 59: 204-217.
- Thomson, J. J. (2003). "Causation: Omissions," *Philosophy and Phenomenological Research* 66: 81-103.
- Ulatowski, J. (2008). "Action Under a Description." Unpublished manuscript. University of Wyoming.
- van Riemsdijk, H., and Williams, E., (1986) *Introduction to the Theory of Grammar*. Cambridge, MA: MIT Press.

