To appear in Ray Jackendoff, *Language, Culture, Consciousness: Essays on Mental Structures* (MIT Press). Comments more than welcome

Chapter 8
The Peculiar Logic of Value[1]

1. Overview

The end of chapter 4 alluded to the fundamental issue of how a system of values can be grounded. As a preliminary step toward being able to address this issue, it is of interest to ask what a system of values *is*. This chapter takes up the challenge, as an exploratory effort.

As in previous chapters, the basic approach will not be to ask what values "are in the real world." They are not *anything* independently of the people who conceptualize them. Rather, the question is how humans conceptualize values (especially unconsciously[2]) and how values play a role in governing people's judgments and behavior. As in previous chapters, this question will be approached in part by using linguistic expressions of value as clues for the structures we are seeking. The inquiry will be validated by the extent that, by positing a relatively constrained set of conceptual building blocks, we can describe a rich variety of linguistic expressions and commonplace intuitions involving value. Through understanding the conceptual structures in which values are embedded, it is to be hoped that we are in a better position to inquire into the evolutionary and cultural roots of systems of value.

One thing I will not aspire to here is to decide what value system we should adopt, that is, to make value judgments over value systems. Nevertheless, I concur with Greene 2003 that investigation of this sort can be useful in addressing such issues.

My overall hypothesis is that value is an abstract concept/property connected to (conceptualized) objects, persons, and actions. It is abstract because it is not directly perceptible. The value of an entity plays a role in various rules of inference which affect the way one reasons about the entity. Thus value serves as an intermediary in a system of logic – not logic in any standard propositional/formal sense, but in the sense of a conceptual/heuristic logic. The rules of this logic are a sort of internal accounting system that helps connect many sorts of disparate objects, actions, and persons. Cultures vary in what value is conventionally assigned to what sort of entity and in what context, but, according to my hypothesis, the basic parameters of the logic

---

[1]I am grateful to Hildy Dvorak, Beth Jackendoff, Steve Umans, Janet McIntosh, and Marion Smiley for useful discussions of this material.

[2]See Barth 1993 for discussion of how (in the present terms) conscious, verbalizable values may be quite different from implicit (unconscious) values, which can be detected through regularities of behavior.

of value – that is, the internal accounting system – are to some degree universal.

In order for objects, persons, and actions to have values, there must be rules of inference that assign values to entities on various grounds, i.e. rules in which value appears in the consequent of the rule ("If such-and-such takes place, then such-and-such a value is assigned."). These rules are the entryways into the value system. But just assigning values is of little use unless values have some effect on behavior. So there must also be rules of inference that favor performing certain actions on the basis of values, i.e. rules in which value appears in the antecedent of the rule – the "outputs" of the value system ("If such-and-such an action has such-and-such a value, then do it"). In between entry and output there may be many inferences that involve values in both the antecedent and consequent, that is, reasoning internal to the value system.

An example of such an internal accounting system may make the idea clearer. Consider the conceptual status of points in a game (a case discussed by Searle 1995). Such-and-such a physical action in the game leads to the assignment of so-and-so many points to a player: this is the entry into the system. Within the system, points are totaled by addition as they are assigned, an operation that has no physical counterpart. The output of the system is the rule that says that the winner at the end is the player with the most points (or, depending on the game, the player with the least points, or the player to reach a prescribed total first). The points therefore serve as an inferential intermediary between the actions in the game and the outcome; without this intermediary, the outcome could not be determined. The points have no significance except within the context of the game (or the *frame* in the sense of chapter 4): it is senseless to say, out of the blue, "I have 35 points," as if it were like "I have a new car."

Values are more complex than points for a variety of reasons. First, a value has two dimensions: a *valence* (positive or negative) and a *magnitude*. Values are not just *good* and *bad* (valence), but also *better* and *worse* (magnitude). Except in the case of monetary value, the magnitude is not a numerical quantity, but a relative quantity, perhaps measured by the basic mammalian magnitude system (Dehaene 1997, Hauser 2000). Thus values can be compared and combined approximately, but there is a lot of room for slop (Weber's Law error).

Far more problematic is that there turn out to be several different sorts of value, each of which pertains to different entities and plays its own role in rules of inference. My impression is that crossdisciplinary discussions of value have often foundered because psychology deals primarily with one sort, economics with another, moral philosophy with yet another, and ordinary language conflates them. In addition, most of these sorts of value appear in two different versions, which I will call the *objective* and the *subjective* versions. In the objective version, the judger is attributing value to something in his/her conceptualized world, e.g. *X is of value*. In the subjective version, the judger is attributing value to something in his/her conceptualized world, relative to some observer (which may be the judger him/herself): *X is of value to Y/to me*. This distinction appeared already in the discussion of psychological predicates in chapter 7, and we will build on it here. A final difference between values and points is that the rules of inference internal

to the value system admit many more possibilities than simple addition. In working these rules out, I hope to establish that they are intuitively plausible, since after all they are meant to capture basic generalizations about how we reason with values.

2. Six kinds of value

This section will outline the dimensions of the value system, so as to set a context for the more detailed discussion to follow.

2.1. Affective value. The first type of value might be called *affective value* (or *A-value*). An event or situation has affective value for someone if it has a positive or negative effect on them, if it yields pleasure or suffering, a benefit or a cost. A simple expression of A-value is *good/bad for so-and-so*, as in (1).

(1) a. Eating your dinner will be good for you.
b. Being overweight is bad for Max.

Obviously, the same event may simultaneously be beneficial to one person and harmful to another; for example the action of *revealing the name of the thief* might be good for the snitch and bad for the thief. Thus affective value arises not from the event per se but from its effect on individual participants in it.

There are expressions of affective value which do not name a participant, for instance (2).

(2) a. Drinking milk is good.
b. Being overweight is bad.

Here the affected participant is an implicit generic individual identified with the subject of the generic subordinate clause: a close paraphrase is (3).

(3) a. Drinking milk is good for one/for people/for ya.
b. Being overweight is bad for one/for people/for ya
(where *ya* is taken in the generic personal sense discussed in chapter 7)

When the implicit generic individual is implicit, as in (2), the assertion of value presents itself as an objective property of the generic event or situation. The contrast precisely parallels that with the evaluative predicates such as (4), discussed in chapter 7.

(4) a. That problem interests me/is interesting to me. ["subjective"]
b. That problem is interesting. ["objective"]

I'll call examples like (1) expressions of *subjective* A-value, and those like (2) expressions

of *objective* A-value. Now comes an absolutely crucial point. Following the overall mentalist tenets of the present approach, *objective A-value is still not value in the world, independent of observers*. You may think that it's good to drink milk, and I may not. But each of us conceptualizes this value as a property of drinking milk. What makes the value objective in my sense is that we are in disagreement about the value of drinking milk independent of any particular person. By contrast, if you think drinking milk is good for Bill, and I think it's not good for Harry, i.e. if we're talking in terms of subjective A-value, we have no disagreement.

Affective value is the type of value most connected to biological issues. Questions of preference, likes and dislikes, and approach and avoidance can be couched in terms of the A-value of the situations and actions involved.[3] One's judgments of A-value of a situation (from one's own point of view) are linked to the character of one's own experience. In order to develop this point, recall the discussion in chapter 2 of the "valuation features" of consciousness: the perceptual systems give the contents of consciousness their *form*, but the valuation features give them their *feel* – the sense of familiarity or novelty, the sense of reality versus imagination, the sense of self- versus world-controlled, and so on. Among the valuation features proposed there was one called *affective*, which comes with a valence. If the cognitive structure of an entity contains the feature [+affective], this entity is experienced as something that *matters*, either positively or negatively.

However, I want to make a distinction between *experiencing* something as attractive or aversive (i.e. having an experience that involves the feature [+affective]) and *judging* its A-value. There are two reasons. First, one can make judgments of A-value that distance one from experience: "That may look attractive, but I know better – it's really dangerous." Second, the affective valuation feature at best connects only to one's own experience. It cannot account for the ability to attribute likes and dislikes to others, and to compare one's own with theirs: "That problem is interesting to her, but not to me." That is, judgments of "subjective" A-value go cognitively way beyond the experiential system, clearly involving Theory of Mind.

2.2. Resource value. A second type of value is what I will call *resource value* (or R-value). An object has a resource value if it is *good for someone to have*; a simpler expression is just that the object is *valuable*. One reason something may be good to have is that it offers the potential (or affordance) for an event with affective value. For a simple case, food has R-value because it offers the potential of being eaten, which is in turn an action of A-value to the eater. Similarly, a house has R-value because it offers the potential of being lived in. Another reason something may be good to have is that it adds to the esteem of its owner, as in the R-value of a famous painting. For another prominent case, money has R-value to its holder because it offers the potential of being exchanged either for other objects with R-value or for the performance of actions with A-value to the holder. A full discussion of the sources of R-value and their interactions are well

---

[3]Thus this is the kind of value apparently of greatest concern to psychologists, e.g. Herrnstein 1993 and Mandler 1993.

beyond the scope of the discussion here.[4] However, sections 5 and 7.2 offer entries into the issues.

Resource value, like affective value, comes in subjective and objective varieties. All the examples so far are objective, in the sense that the object has simply has R-value rather than R-value *to so-and-so*. The subjective/objective contrast is hard to express using the word *good*, but it turns up in expressions like (5).

(5) a. This piece of land is very valuable/worth a lot. [objective]
b. This piece of land is very valuable/worth a lot to Harry. [subjective]

In (5a) the sense is that anyone will value the land highly; (5b) leaves the question open of whether it means anything to anyone else.

Resource value is defined partly in terms of its affordance for A-valued actions. But there is a secondary interaction between R-value and A-value. To the extent that having available resources reduces anxiety, the situation of having things with resource value can itself be of affective value: *having stuff is good for you; lacking stuff is bad for you.* Of course, the strength of this interaction varies from person to person.

2.3. Quality. An object or event can be valued in terms of its quality relative to other objects or events of the same type. Just to keep the terminology consistent, I'll call this *Q-value*.

(6) a. This is a good/terrible computer. [object]
b. That was an excellent/miserable back dive. [action]

Typically, when objects are rated for quality, it is in terms of a particular action for which the object is to be used.

(7) a. This spatula is good for frosting cakes with.
b. This book is good for learning about deconstructionist phonetics.

When the *for*-phrase is absent (*a good spatula, a good book*), there is still an implicit purpose, as observed by Katz 1966 and Pustejovsky 1995. The default interpretation is that the object in question has quality with respect to performing its *proper function* – what the object is *for* (in the sense of Millikan 1984, or the *telic quale* in the sense of Pustejovsky 1995). A good spatula is one that is good for scraping and spreading viscous materials (usually food-related), and a good book is one that is good for reading.

---

[4]Economists, whose basic data are exchanges, are therefore most concerned with resource values and those A-valued actions for which exchanges for R-valued objects are possible, e.g. labor and services. See, for example, the contributions of Scitovsky 1993 and Akerlof and Yellen 1993 in Hechter et al. 1993.

The adjective *excellent* is natural in expressions of Q-value; by constrast, it is somewhat awkward in expressions of A-value:

(8) a. This spatula is excellent for frosting cakes with.
b. ?? Drinking milk is excellent for you.

An extension of Q-value concerns the use of some object for the function normally played by something else (Aronoff 1980). Expressions like (9), in particular the *...makes a good X* construction, are characteristic of this reading.

(9) a. This rock is/makes a good table.
b. This table is/*makes a good table.

Alternatively, the purpose may be inferred from conversational context: one might say *THAT cloud's good* in the context of comparing clouds for their resemblance to barnyard animals.

2.4. Normative value. Another kind of value is *normative value* (or N-value), which concerns conformance to social norms of the sorts discussed in chapter 4. Among the subvarieties of N-value are moral/ethical value,[5] religious value, and valuation according to standards of etiquette (manners, politeness, etc.). Unlike the previous sorts of value, this is strongly situated in the social domain: it not only has to do with people, but with people in the context of social interaction.

As observed in chapter 4, the subvarieties of normative value share a great deal of their linguistic expression and often apply in similar ways to similar situations, but they can still be teased apart into separate subdomains. For example, it is possible for a highly moral person to have bad manners; and conversely, a person with exemplary manners may well still be deeply immoral. Nevertheless, for the sake of keeping the exposition relatively compact, the present chapter will not distinguish these subdomains.

Characteristic expressions of normative value are shown in (10). Those in (10a,b) might be called expressions of *action-focused* N-value; that in (10c) is *person-focused* N-value.

(10) a. It is good of Harry to wash the dishes without being asked. } [action-focused]
b. Washing the dishes without being asked is good of Harry. }
c. Harry is good to wash the dishes without being asked. [person-focused]

The attribution of N-value, as with A-value, is again focused on a relation between a person and an event. However, this time the event *must* be something that the person does, as in (11a) – it cannot be something that happens to the person, as in (11b,c). By contrast, A-value can be ascribed to such situations (11d,e).

---

[5]The main sort of value of interest to philosophers, e.g. Stich 1993, Harman 2000.

(11) a. Washing the dishes is good of Harry.
b. *Being overweight isn't good of Harry.
c. *Being elected chairman is good of Harry.
[except if this means he is good to *allow* himself to be elected chairman, which is an action on his part]
d. Being overweight isn't good for Harry.
e. Being elected chairman is good for Harry.

There are two ways that attributions of N-value can be made more general. One is to omit the person from an action-focused attribution of N-value, as in (12). This implies that it would be good of *anyone* to perform the action.

(12) a. It is good to wash the dishes without being asked.
b. It is bad to kill people.

The difference between (10) and (12) is that (10) is a relation between an actor and an act, while (12) is simply a valuation of a generic act.

The second way to bleach out the relational character of N-value is to can omit the action from a person-focused expression of N-value, so that N-value is ascribed simply to a person, as in (13a). The sense is then that the person's generic actions, whatever they may be, are of N-value. This manipulation is impossible with A-value: a paraphrase like (13b) in terms of A-value makes no sense.[6]

(13) a. Harry is good. [= 'Harry does things of positive N-value']
b. Harry is good. [≠ 'things happen of positive A-value for Harry']

(13) shows that *good* can be used for both A-value and N-value (as well as Q-value). Chapter 4 mentioned a parallel ambiguity with *should*. The "predictive" sense (14a) does not have to do with values. But there are two readings that do involve values: there is a "prudential" sense (14b), which expresses affective value, and a "normative" sense (14c), which expresses normative value.

---

[6]Note the parallel between the four ways of expressing of N-value and the four ways of expressing evaluative predicates pointed out in chapter 7:

(i)	Washing the dishes is good of Harry.	This story bores me.
(ii)	Washing the dishes is good.	This story is boring.
(iii)	Harry is good to wash the dishes.	I'm bored with this story.
(iv)	Harry is good.	I'm bored.

(14) a. The bus should arrive soon. [predictive]
b. You should take an umbrella in case it rains.
[prudential: 'the A-value of taking an umbrella is positive'; 'it would be good for you to take an umbrella']
c. You should offer to wash dishes.
[normative: 'the N-value of offering to wash the dishes is positive'; 'offering to wash the dishes would be good of you']

There are interactions between N-value and A-value. For example, it is often the case that actions that are good *of* you (N-value) are good *for* someone else (A-value), for example acts of charity; conversely, gratuitous violence is bad *of* the perpetrator and bad *for* the victim. Another interaction is that it may *feel* good/bad (A-value) to *do an act* that's good/bad (N-value): that is, performing an act with N-value may result an accompanying secondary A-value. In common language, we say that someone who experiences this interaction "has a conscience." In particular, when the value is negative, I think the secondary A-value is called guilt.

However, as pointed out in chapter 4, such synergy between N-value and A-value is not invariable. Consider for example the performance of religious rituals, which are of positive N-value but don't really benefit anyone directly. Similarly, illicit sex may be of positive A-value but is definitely of negative N-value (that's why we call it illicit).

All the description of N-value so far has been in "objective" terms: such-and-such an action is normatively good, such-and-such a person is normatively good. There is also a subjective version, where other people's opinions are being compared with one's own.

(15) To Joe, Harry's a good guy (–but not to me).

This is tricky to express in English, requiring the prepositional phrase *to Joe* at the beginning of the sentence, because the most expedient grammatical expression, *Harry is good to Joe*, is taken up by another meaning: 'Harry acts in a manner that's good for Joe (of positive A-value to Joe)'.

2.5. Prowess. In addition to the normative reading of *good* attributed to people, illustrated in (13a), there is another reading that rates quality of performance, e.g. (16a). The latter has another syntactic form (16b), which in turn shows affinities with the notion of quality (Q-value) attributed to artifacts, as in (6)-(7) and (16c). We might call the sense in (16a,b) *prowess* or P-value. Note that like quality, prowess can be expressed by using the adjective *excellent*.

(16) a. Harry is good/excellent (at singing). [Prowess]
b. Harry is a good/excellent singer. [Prowess]
c. This is a good/excellent knife. [Quality]

And, with proper contextual support, there is yet another reading, in which Harry is construed as a resource: 'Harry is a good choice for the job.' Like Q-value, prowess is in effect an affordance

for some task.

2.6. Esteem. A final notion of value also pertains specifically to persons and might be characterized as *esteem* (for consistency, I'll call it E-value). Esteem seems to be a composite of person-focused normative value, prowess, status in the dominance hierarchy, wealth (accumulation of R-value) and perhaps other factors such as simple personal attractiveness. Group membership also plays a role: members of groups other than one's own are by default accorded less respect, as are members of generally low-status groups. However, particular individuals of low-status groups who have other highly respected qualities (wealth, prowess, leadership, etc.) may be accorded greater respect than the default. This shows the interactive, composite character of E-value.

Objective and subjective versions of E-value appear in (17a,b) respectively.

(17) a. Harry is prestigious/well-respected. Harry has a good reputation. [objective]
b. Joe respects Harry. [subjective]

Unlike prowess, esteem is clearly a socially rooted value. In fact, esteem or reputation is an important point where the values system outputs to behavior. As Fehr and Fischbacher 2004 suggest, if one is esteemed or respected, others seeke to engage in cooperative interactions with one. Performing normatively good actions is one way to enhance esteem, which in turn leads to opportunities for A-valued interactions. Hence esteem might be amount to sort of personal resource value.

Tables 1 and 2 are an effort to sum this all up. The rest of the chapter deals mostly with A-value, R-value, N-value, and E-value. Section 7.4 will introduce a further type, "moral debt" or MD-value.

Type of value	Applies to ontological type:	Subjective version	Objective version
Affective (A-)value	events, situations	Situation *X* is good for Y	Situation *X* is good
Resource (R-)value	objects	Object *X* is valuable to Y	Object *X* is valuable
Quality (Q-value)	events objects		Event *X* was a good one Object *Y* is good for doing X
Normative (N-)value	action-focused, relational: action-focused, absolute: person-focused, relational: person-focused, absolute:	To Z, doing *X* is good of Y To Z, doing *X* is good To Z, *Y* is good to do X To Z, *Y* is good	Doing *X* is good of Y Doing *X* is good Person *Y* is good to do X Person *Y* is good
Prowess (P-)value	persons		*Y* is good at doing X
Esteem (E-)value	persons	X respects *Y*	*Y* is prestigious

Table 1. Varieties of value. Entity to which value is ascribed is in italics.

Table 2. Interactions of varieties of value (so far)

A. Having more R-valued objects ==> Having more security (A-value)
B. Actor performs actions of good/bad A-value to beneficiary/victim ==> Actions are of good/bad N-value for actor
C. Actor performs actions of N-value ==> Actions may be of A-value to actor (conscience)
D. N-value of actor contributes to E-value of actor
E. E-value of person contributes to desirability to others (A-value) of interacting with this person
F. Opportunity for interaction with others ==> R-value to self

3. Formalization and some preliminary inference rules

3.1. Notation. It is now useful to introduce some formalization. I will use a notation that is relatively easy to read and manipulate, of the general form shown in (18).[7]

(18) a. Objective form:
X-VAL (ENTITY) = valence • mag
'the x-value of entity is valence times magnitude'
(where X ranges over A, R, N, E and valence ranges over + and -)

b. Subjective form:
X-VAL (ENTITY, PERSON) = valence • magnitude
'the x-value of entity for person is valence times magnitude'

Whether the objective or subjective form is intended is usually clear from the number of variables.

Using this format, some of the examples from the previous section can be notated as in (19). (*YA* is the generic person introduced in chapter 7.)

(19) a. Being overweight is bad.
A-VAL (YA BE OVERWEIGHT) = -

b. Drinking milk is good for Harry. Harry should drink milk.
A-VAL ((HARRY DRINK MILK), HARRY) = +

c. That land is valuable.
R-VAL (LAND) = +

d. That land is valuable to Harry.
R-VAL (LAND, HARRY) = +

e. Joe respects Harry.
E-VAL (HARRY, JOE) = +

The problem case for the notation is N-value. Here there are potentially three variables: the action, the actor, and the subjective evaluator. (20) suggests a notation in which the evaluator, if present, is marked with *TO*.

---

[7]A more traditional notation might use a function X-VAL' of two or three variables that yields a truth value, along the lines of (i)-(ii).
(i) Objective: [X-VAL' (ENTITY, valence • magnitude)]
(ii) Subjective: [X-VAL' (ENTITY, PERSON, valence • magnitude)]
The function X-VAL in (18) can be derived from these by lambda-extraction on the last variable. However, since we eventually want to be able to compare and add values, the form in (18) is more direct.

(20) a. Harry is good to wash the dishes. Harry should wash the dishes.
N-VAL ((HARRY WASH DISHES), HARRY) = +

b. It's bad to kill people.
N-VAL (YA KILL PEOPLE) = -

c. Harry is good. [in normative sense]
N-VAL (HARRY) = +

d. To Joe, Harry is good to wash the dishes.
N-VAL ((HARRY WASH DISHES), HARRY, TO JOE)

e. To Joe, it's bad to kill people.
N-VAL ((YA KILL PEOPLE), TO JOE)

3.2. A route through the system. With just this much notation, we can begin to formalize the intuitive logic of values. For a first step, we need an entry into the system, an inference rule whose antecedent is not about values but whose consequent is. Chapter 7 suggested one such, repeated here with slight alteration as (21).

(21) *Tuning of value to valence: "Nice things are good for you"*

a. $\left[\begin{array}{l}\textit{[any thematic tier]} \\ \text{(X) AFF}^{\alpha}\ \text{Y}\end{array}\right]^{\beta} \Rightarrow_{\text{default}} \text{A-VAL}\ (\beta, \text{Y}) = \alpha$ (where α ranges over + and -)

b. $\left[\begin{array}{l}\textit{[any thematic tier]} \\ \text{X EXP}^{\alpha}\ \text{(Y)}\end{array}\right]^{\beta} \Rightarrow_{\text{default}} \text{A-VAL}\ (\beta, \text{X}) = \alpha$

Recall what this says. The left-hand side of the rule is an event or situation; the right-hand side is a value judgment, as desired. (21a) involves the macrorole function X AFF$^{\pm}$ Y, 'X affects Y positively or negatively.' In the former case, Y is a *beneficiary* of the action; in the latter, Y is a *patient*. Thus (21a) says basically 'an action in which Y is the beneficiary is of positive A-value to Y; an action in which Y is the patient is of negative A-value to Y'. In short, it's nice to be helped and yucky to be victimized.

Turning to (21b), the macrorole function on the left-hand side is X EXP$^{\pm}$ Y, 'X experiences Y as positive or negative'. Thus (21b) says that 'an event or situation which X experiences positively is of positive A-value to X; an event or situation which X experiences negatively is of negative A-value to X'. The upshot of these two rules, then, is that the character of an event in which one is involved can lead to a judgment of that event's value.

Having gained entry into the value system, let's next look at some rules that manipulate values within the system – the parallels to adding points in a game. We will start by formalizing interaction B in Table 2. The informal name, "It's good to be nice to people," suggests the gist of the rule. However it should be borne in mind that in the rule itself, the value is a variable which may be positive or negative, so the rule actually encompasses the negative counterpart as well – "It's bad to be mean to people" – and similarly for subsequent rules.

(22) *"It's good to be nice to people"*

$$\begin{bmatrix} \text{ACT (X)} \\ \text{A-VAL } (\beta, \text{Y}) = \alpha \end{bmatrix}^{\beta} \Rightarrow_{\text{default}} \text{N-VAL } (\beta, \text{X}) = \alpha$$

This says that an action β by X that has an affective value for someone else has a corresponding normative value for X. That is, it's good *of* X to do something of positive A-value *for* Y; it's bad *of* X to do something of negative A-value *for* Y.

Next comes an interesting step. The basic intuition is that if you do something good, you're a good person; if you do something bad, you're a bad person. (To keep things from getting overwhelming, I ignore the distinctions among different sorts of N-value.) This is a rather peculiar principle, but, strikingly, it conforms to intuition.

(23) *"Doing good things makes you good"*

$$\text{N-VAL } ((\text{X ACT, X}) = \alpha \Rightarrow_{\text{default}} \text{N-VAL (X)} = \alpha$$

A slightly more sophisticated version of this rule might say that good acts add to your total "goodness", and bad acts subtract from it. Such a rule cannot be stated within a static logic, which presumes a timeless database. Rather, it is necessary to introduce a dynamic or procedural logic, one that allows values to be updated. Such a system will be necessary in any event in order to allow for belief revision (i.e. deciding one was wrong about something). Without being very specific about how such a system works, we might state the rule in question as something like (24), where the material in **special type** is meant as a procedural instruction.

(24) *"Doing good things makes you better"*

$$\begin{bmatrix} \text{ACT (X)} \\ \text{N-VAL } (\beta, \text{X}) = \alpha \end{bmatrix}^{\beta} \quad \textbf{==> add } f \bullet \alpha \textbf{ to } \text{N-VAL (X)}$$

(where α = valence • magnitude, and f is a multiplier on the magnitude)

As a result of (24), a person's normative value at any moment is related to his or her history of performing normatively valued actions. The multiplier f is sort of a wild card, of an intuitively justified sort: "You have now performed so-and-so many good actions. Is that enough to make up for all the bad things you've done?" "That one horrible thing you just did has wiped out my whole good opinion of you." "Even though what you just did was horrible, I'm not going to hold it against you." Such common statements show that rule (24) can be applied in highly subjective fashion. The multiplier f is meant to encode the locus of this subjectivity; exactly what determines it under different circumstances is a question far beyond the scope of the present exploration.

Just another reminder: rule (24) is not meant to treat X's normative value as a free-standing thing unto itself. We are not describing how X acquires "real" normative value, but rather how the judger conceptualizes X acquiring "objective" normative value. Lest this should seem a problematic stance on normative value, it should be recalled that this is the very same

stance taken in studying vision, where perception is described in terms of the perceiver developing a conceptualization of the objective "world out there", in response to certain inputs to the perceptual system. In both cases we are concerned with the individual's sense of what is real. And from this point of view, goodness is as real as size.

The contribution of normative value to esteem (Interaction D in Table 2) might be encoded as follows:

(25) *"Your value as a person depends in part on how good you are"*
E-VAL (X) = c_1 • N-VAL (X) + c_2 • P-VAL (X) + ... other factors
(where c_i is a normalizing constant, possibly context-dependent)

As observed in section 2, the other factors beyond normative value might include X's prowess (P-value) in relevant activities, X's status in the dominance hierarchy, X's attractiveness, X's wealth (accumulated R-value), and perhaps others. The ratio of importance among these factors may be highly variable and context-dependent. (Being *notorious* is having a high E-value despite a negative N-value.) Moreover, since we are dealing with the analog magnitude system in measuring values, the idea of a normalizing constant in the usual mathematical sense is far too specific.

As pointed out in section 2, being esteemed isn't enough: it has to do you some good. Here is a somewhat more formal statement of Interaction E in Table 2.

(26) *"It's good for you to associate with esteemed people"*
Subjective: E-VAL (X, Y) = α $\Rightarrow_{\text{default}}$ A-VAL ((Y ACT WITH X), Y) = α
Objective: E-VAL (X) = α $\Rightarrow_{\text{default}}$ A-VAL (YA ACT WITH X) = α

Returning to rule (25) for a moment: recall from section 2 that group membership has a strong effect on esteem: by default one accords members of one's own group higher esteem than members of other groups, and one accords members of high-status groups (whatever their other qualities) default higher status. The consequence, following rule (26), is that one prefers to associate with and do business with members of one's own group and with members of high-status groups, all things being equal. This seems just the right result, in light of the discussion in chapter 4.

Rules (22)-(26) are all inside the value system; they are the counterpart of adding up points in a game. In order for all this to impact on behavior, we also need an "output" rule in which the value system affects action. About the simplest possibility is that one should prefer to do actions that are better for one; that is, the A-value of a potential action affects its preferability. In order to affect one's action, this rule has to be stated in terms of the relative value of *EGO*'s contemplated actions. The outcome of the rule has to be procedural, the making of a choice. (27) is an attempt.

(27) *"Do what's better for you"*
A-VAL (EGO ACT_1) > A-VAL (EGO ACT_2) ==>$_{default}$ **Execute** EGO ACT_1

This is perhaps too blatant, in that it appears to lead directly to the execution of an action. Perhaps better is a version in which one creates an intention, or adds a new intention to one's working memory. (Following the discussion of chapter 6, *INTEND* could be replaced with *COM*.)

(28) *"Decide to do what's better for you"*
A-VAL (EGO ACT_1) > A-VAL (EGO ACT_2) ==>$_{default}$ **ADD** INTEND (EGO ACT_1)

(27)-(28) both compare objective A-values of actions and lead to a choice. The subjective version is more reflective. You can't decide on someone else's action. The best you can do is infer what they'll choose. So instead of the procedural function in the right-hand side of the rule, there is a conceptual function which we'll call *CHOOSE*. This is the Theory of Mind's description of what's going on in other people (and oneself) when their action is determined by (27/28).

(29) *"People decide to do what's better for them"*
A-VAL ([X ACT_1], X) > A-VAL ([X ACT_2], X) =>$_{default}$ CHOOSE (X, [X ACT_1])

This gives us a whole loop in and out of the system: you observe X doing something nice for somebody (21), you infer that this is good of X (22), and therefore that X is good (23/24); from this you infer that it is good for you to associate with X (26), and therefore you choose to associate with X (27/28). There are many other routes through the system. For instance, in addition to rule (22) ("It's good to be nice to people"), there are many culture-specific attributions of normative values to actions: "It's normatively bad to dress such-and-such a way", "It's normatively bad to eat such-and-such", "It's normatively bad to show attraction to people of the same sex", and so on. These are other entries into the value system, and they likewise contribute to normative judgments of people who commit these acts. On the output end, the system eventually has to ground out in action, perhaps always through a judgment of A-value that leads to rule (27/28), "(Decide to) do what's better for you."

It's also worth mentioning that judgments of A-value are highly context-dependent. For instance, "giving in to temptation" is valuing an immediate gratification (an action with positive A-value) in preference to a potentially greater A-value to be realized over a longer time-span. (See Stevens and Hauser 2004 for discussion of this issue of "temporal discounting."

4. Reciprocity

Next let us consider the relation between two actions expressed by a certain use of the preposition *for* in English.

(30) a. Susan praised her son Sam **for** behaving nicely.
b. Fred cooked Lois dinner **for** fixing his computer.
c. Susan insulted Sam **for** behaving badly toward her.
d. Lois slashed Fred's tires **for** insulting her sister.

These sentences describe situations in which someone does something "in return" for someone else's action. Those in (30a,b) describe actions with positive values; those in (30c,d) describe actions with negative values. Such acts of reciprocity can felicitously take place only with another person, an entity that can be regarded as having values and responsibility. One cannot sanely punish one's car for getting a flat tire.

If we switch around the actions among the examples in (30) we get sentences that sound odd or perhaps ironic.

(31) a. #Susan insulted Sam for behaving nicely.
b. #Lois slashed Fred's tires for fixing her computer.

This shows that we expect a positively valued action in return for a positively valued action, and a negatively valued action in return for a negatively valued one.

A further look shows that reciprocity is sensitive not only to valence, but also to the (analog) magnitude of values as well: we find it odd if the two actions related by *for* do not match in quantity. The sentences in (32) convey some of this oddness:

(32) a. #Fred cooked Lois dinner for saying hello to him.
b. #Fred cooked Lois dinner for rescuing all his relatives from certain death.
c. #Fred slashed Lois's tires for eating too little at dinner.
d. #Fred slashed Lois's tires for murdering his entire family.

In (32a) and (32c), we sense Fred as overreacting, as doing something unwarranted in return for Lois's action; in (32b) and (32d), we sense him as underreacting, as doing something that is not nearly enough to recognize the importance of Lois's action.

The intuition, then, is that a reciprocal action calls for rough equivalence of value between the two actions. Crucially, a particular action may be of different value to the participants: you may not even know that your action helped or harmed me. Thus the principle of reciprocation must be stated in terms of the particular person the action affects, i.e. subjective A-value. (33) is a good first approximation; the function *RECIP* is a modifier to be read as 'in return for'. For the moment, we take the relevant notion of value to be affective value. Like many inferences with value, this one is defeasible.

(33) One act in return for another matches the original act in valence and magnitude

$$\left[\begin{array}{l} Y\ ACT_2 \\ RECIP\ [X\ ACT_1]^{\alpha} \end{array}\right]^{\beta} \Longrightarrow_{default} \text{A-VAL}\ (\beta, X) = \text{A-VAL}\ (\alpha, Y)$$

The logic of reciprocity expressed by (33) is a cognitive elaboration of a behavioral strategy well-documented in the ethological literature, *reciprocal altruism*: "I'll scratch your back because you scratched mine."[8] I leave open how much of its detail can be attributed to nonhuman primates (not to mention elephants and bats).[9] What strikes me as particularly human, though, is the broad generality of the actions available for reciprocation.

Because (33) is neutral as to whether the value in question is positive or negative, it serves not only to express reciprocal altruism but also retaliation (or retribution). In this case, the equivalence of values amounts to a more or less formal statement of "the punishment fits the crime": this helps guide what responses are appropriate in retaliation for harmful actions.

There is another kind of reciprocation for negative actions, illustrated in (34).

(34) a. Fred cooked Lois dinner (to make up) for having embarrassed her in public.
b. Fred brought Lois flowers (to make up) for forgetting her birthday.

Here the perpetrator of the negative action is performing a positive action in *restitution*, righting the balance. Again there has to be a rough equivalence of value: notice the weirdness of (35).

(35) a. #Fred gave Lois his vast fortune (to make up) for forgetting her birthday.
b. #Fred brought Lois flowers (to make up) for killing her whole family.

Thus the rule for restitution might be stated as (36). Unlike reciprocity (33), it is not neutral to the valence of the original action: there is no counterpart that sanctions doing someone ill because you have been nice to them. Thus the principle must explicitly encode the valence of the original action.

(36) An good act that makes up for a bad one matches the original in magnitude

$$\left[\begin{array}{l} X\ ACT_2 \\ REST\ \left[\begin{array}{l} X\ ACT_1 \\ \text{A-VAL}\ (\alpha, Y) = - \end{array}\right]^{\alpha} \end{array}\right]^{\beta} \Longrightarrow_{default} \text{A-VAL}\ (\beta, Y) = -\text{A-VAL}\ (\alpha, Y)$$

---

[8]I use this slightly nontraditional phrasing deliberately. "I'll scratch your back and you scratch mine" suggests an explicit agreement, a different situation which we address in the next section.

[9]See Stevens and Hauser 2004 for arguments that reciprocal altruism is much less common among nonhumans than usually thought.

Notice that if X does something harmful to Y, one sort of retribution for Y is to force X to perform restitution, perhaps through the intervention of the authority of the group. However, this sort of retribution is not equivalent to retaliation. For one extreme case, "Nothing you can make the murderer do will bring my son back," that is, restitution is impossible, even though retaliation might be (e.g. killing the murderer's son).[10]

Finally, (33) and (36) are stated as inferences from reciprocation to the value of the actions in question. However, this is not enough: one *should* reciprocate actions that benefit one and one *should* perform restitution for having harmed others, that is, there is a normative value attached to these actions. How this value plays out – what actions should be reciprocated and restituted, and what counts as appropriate reciprocation and restitution – is variable among cultures and subcultures. But the overall principle seems universal. (37) is one way to state these principles. They are special cases of rule (22), "It's good to be nice to people."

(37) a. *"It's good to reciprocate nice things"*

$$\left[\begin{array}{l} X\ ACT_1 \\ A\text{-}VAL\ (\beta, Y) = + \end{array}\right]^{\beta} \Longrightarrow \text{N-VAL}\left(\left[\begin{array}{l} Y\ ACT_2 \\ RECIP\ (\beta) \end{array}\right], Y\right) = +$$

b. *"It's good to make up for harming someone"*

$$\left[\begin{array}{l} X\ ACT_1 \\ A\text{-}VAL\ (\beta, Y) = - \end{array}\right]^{\beta} \Longrightarrow \text{N-VAL}\left(\left[\begin{array}{l} X\ ACT_2 \\ REST\ (\beta) \end{array}\right], X\right) = +$$

Depending on the circumstance, there are two possible negative counterparts of (37a). One is "you should retailiate", generalizing the left-hand side of (37a) to negative valence, as in (38a). The other is "you should turn the other cheek", in which case the right-hand side places a negative N-value on reciprocation, as in (38b).

(38) a. *"It's good to retaliate for bad things"*

$$\left[\begin{array}{l} X\ ACT_1 \\ A\text{-}VAL\ (\beta, Y) = - \end{array}\right]^{\beta} \Longrightarrow \text{N-VAL}\left(\left[\begin{array}{l} Y\ ACT_2 \\ RECIP\ (\beta) \end{array}\right], Y\right) = +$$

b. *"Turn the other cheek"*

$$\left[\begin{array}{l} X\ ACT_1 \\ A\text{-}VAL\ (\beta, Y) = - \end{array}\right]^{\beta} \Longrightarrow \text{N-VAL}\left(\left[\begin{array}{l} Y\ ACT_2 \\ RECIP\ (\beta) \end{array}\right], Y\right) = -$$

Not only do these two principles conflict with each other, but in addition (38a) conflicts with rule (22), while (38b) does not. Again, a lot of cultural variation arises from how these rules are understood to apply in practice. I suspect that it's partly in the service of negotiating such conflicts that more explicit moral and legal codes arise.

There are a number of ways for slippage to be introduced into reciprocation. Perhaps the

---

[10]The distinction between these two in conceptual development is noted by Piaget 1932; he claims the notion of restitution is established later than that of retribution.

most pernicious arises from a general cognitive bias toward overestimating harm to oneself and underestimating harm to others. Hence, if I retaliate against you, you judge the harm done to you to be greater than the harm you originally did to me. You are therefore motivated to even the scores by retaliating further, leading to escalating cycles of violence.

The rules of reciprocation so far have been stated in terms of the A-value of actions to their participants. Another application of reciprocal *for*, involving normative value and prowess, appears in the examples in (39).

(39) a. Joe praised Sue **for** saving the drowning child.
b. The club honored Sue **for** her service to the community.
c. Sue was awarded a prize **for** winning the race.

In these cases, Sue has done nothing to benefit the individual or organization that is acting reciprocally. Rather, she has done something that has raised her normative value (39a,b) or that has demonstrated prowess (39c), both of which contribute to the esteem in which she is to be held (by rule (25)). The reciprocal actions in (39) are performative actions of esteem-creation (honoring, thanking), which have the effect of making the esteem public (and ostensibly objective). Such statements of esteem are naturally of positive A-value to Sue as well; and the resource value of increased esteem is a further benefit. The expected negative counterpart of (39) also exists: public humiliation and community-sanctioned punishment for normative transgressions – even for moral transgressions that do not directly harm anyone else. A simple example of this counterpart is *Sue scolded Bill for losing his wallet.* A different negative counterpart is apology: restitution for harm done by offering expressions of self-humiliation before the injured party.

(40) is an attempt at stating these sorts of reciprocation, involving a very particular type of reciprocal action, expressing (or otherwise demonstrating) esteem.

(40) a. *"It's good to honor people who have done estimable things"*

$$\begin{bmatrix} X\ ACT_1 \\ N/P\text{-}VAL\ (\beta) = + \end{bmatrix}^{\beta} \Longrightarrow\ N\text{-}VAL\ \left(\begin{bmatrix} Y\ EXPRESS\ (E\text{-}VAL\ (X) = +) \\ RECIP\ (\beta) \end{bmatrix}, Y \right) = +$$

b. *"It's good to apologize to people you've hurt"*

$$\begin{bmatrix} X\ ACT_1 \\ A\text{-}VAL\ (\beta, Y) = - \end{bmatrix}^{\beta} \Longrightarrow\ N\text{-}VAL\ \left(\begin{bmatrix} X\ EXPRESS\ (E\text{-}VAL\ (X) = -) \\ REST\ (\beta) \end{bmatrix}, X \right) = +$$

In these cases of reciprocation, it is hard to know what counts as equivalence in value between the original act and the reciprocal act. Perhaps the best one can do is context-dependent proportionality: first prize ought to be more valuable than second prize; greater praise should be accorded to someone who saves 80 lives than to someone who stops on the highway to help you fix your car; a bigger faux pas calls for more fervent contrition.

Note that etiquette (i.e. N-value) demands that the recipient of one of the reciprocal

actions in (40) respond "I don't deserve it; what I did was nothing." Why is this the case? The immediate cause seems to be a principle to the effect that you shouldn't think too highly of yourself. (41) gives two possible versions of this principle.

(41) a. *"You shouldn't rate your esteem very high"; "Don't have too high an opinion of yourself"*
N-VAL (E-VAL (X, X) >> 0) = -

b. *"You shouldn't rate your esteem higher than it really is"*
N-VAL (E-VAL (X, X) > E-VAL (X)) = -

I leave it an open question whether this principle can be derived from more basic principles.

5. Exchange

5.1. Exchange of actions. In the cases of reciprocation discussed so far, reciprocation is a freely chosen act in response to a freely chosen act. Another scenario for reciprocation is *exchange*, in which the actors *agree* each to do something for the other's benefit, so that the two actions are conceptually yoked more closely. In simple reciprocation, you've scratched my back, so I volunteer to scratch yours in return. In an exchange, an agreement is made: "I'll scratch your back if you'll scratch mine." "OK." Exchanges are basic to every sort of contract in every human society.

A good way to describe this linkage is to treat the exchange as a joint task in the sense of chapters 4 and 6: the actors are co-actors, each performing his or her part in the task. Using the notation of section 6.9, the overall frame for an exchange might be notated as (42).

(42) Exchange:

$$\begin{bmatrix} \{X, Y\}\ ACT \\ COMPOSED\text{-}OF\ \{ \begin{bmatrix} X\ ACT_x \\ A\text{-}VAL\ (\beta, Y) = + \end{bmatrix}^{\beta}, \begin{bmatrix} Y\ ACT_y \\ A\text{-}VAL\ (\gamma, X) = + \end{bmatrix}^{\gamma} \} \\ [FROM\ [\{X, Y\}\ INTEND\ \alpha]] \end{bmatrix}^{\alpha}$$

Decoding this: The first line says that X and Y are performing a joint action. The second line says that the parts of this action include X doing something positive for Y and Y doing something positive for X. The third line says that this whole action is done intentionally as a joint intention – X and Y are jointly committed to performing it. This is possible only if they have come to an agreement.

Following the discussion in section 6.9, two pairs of entailments come from the fact that an exchange is jointly intended. The first pair is that X intends to do ACT_x and that Y intends to do ACT_y. The second pair is that X is obligated to Y to do ACT_x and that Y is obligated to X to do ACT_y. Because of the reciprocal obligations, neither actor is free to opt out; this is the sense

in which the actions are yoked.

Before plunging in further, it is worth noting that every part of this quite complex schema and its entailments has been motivated and related to other forms of action and value. Yet the schema is universally part of human understanding – every human society engages in exchange, and with little apparent effort at learning. The notion of exchange presents itself as a unified gestalt.

In order to go on, it is useful to introduce a notational abbreviation for (42) (in effect simulating its gestalt-like character in experience). X's and Y's actions vary from one exchange to the next, but the rest of the frame is fixed. Thus we can treat X's and Y's actions as free variables and abbreviate the rest as a constant function *EXCH*.

(43) [X ACT_x] EXCH [Y ACT_y]

Relativizing the entailments of joint intention to (43), we get (44), where the notation for obligation anticipates the treatment in chapter 9:

(44) [X ACT_x] EXCH [Y ACT_y] ==>

a. X INTEND [X ACT_x]	'X intends to do ACT_x'
b. Y INTEND [Y ACT_y]	'Y intends to do ACT_y'
c. X HAVE [OB ([X ACT_x, TO Y])]	'X is obligated to Y to do ACT_x'
d. Y HAVE [OB ([Y ACT_y, TO X])]	'Y is obligated to X to do ACT_y'

Like free reciprocation, exchange comes with the presumption (or default inference) that the values of the two actions are related. In the objective version, the values are equal (45a), paralleling (33). The subjective version is more nuanced: each actor comes out ahead in terms of benefits vs. costs (45b).

(45) a. Objective version:
[X ACT_x] EXCH [Y ACT_y] $\Longrightarrow_{default}$ A-VAL (X ACT_x) = A-VAL (Y ACT_y)

b. Subjective version:
[X ACT_x] EXCH [Y ACT_y] $\Longrightarrow_{default}$
A-VAL ([Y ACT_y], X) + A-VAL ([X ACT_x], X) > 0
A-VAL ([X ACT_x], Y) + A-VAL ([Y ACT_y], Y) > 0

The entailments in (45b) say 'the value of Y's action to X outweighs the cost (negative value) of X's action to X, and similarly for Y'.

(45b) omits some additional benefits to the participants: the A-value of conducting a favorable social interaction (elaborated in many cultures through, say, drinking together to seal a transaction), and the R-value of having a trusted trading partner with whom future transactions may be anticipated. In order to gain these benefits, a participant may sometimes agree to a

transaction in which the value of the actions exchanged is otherwise unfavorable (Nathaniel Jackendoff, p.c.). These factors can be incorporated into the equation by adding a third term into the sums in (45b) that includes the value of the entire transaction:

(46) Subjective version (including transaction values):
$[[X\ ACT_x]\ EXCH\ [Y\ ACT_y]]^{\alpha} \Longrightarrow_{default}$
$A\text{-}VAL\ ([Y\ ACT_y], X) + A\text{-}VAL\ ([X\ ACT_x], X) + A\text{-}VAL\ (\alpha, X) > 0$
$A\text{-}VAL\ ([X\ ACT_x], Y) + A\text{-}VAL\ ([Y\ ACT_y], Y) + A\text{-}VAL\ (\alpha, Y) > 0$

It could be the case that these entailments follow simply from the conditions on agreeing to a joint activity: actors will not come to agreement unless each of them believes it is in his or her interest to do so. If so, perhaps (45/46) is redundant. Nevertheless, it's worth stating for the sake of explicitness.

In agreeing to an exchange, each actor is naturally trying to maximize the A-value derived from the transaction – following rule (27/28), "Decide to do what's best for you." Thus it often requires negotiation to achieve a "fair" exchange, where both participants judge the exchanged acts to be of sufficient net value to themselves. Here, in the process of bargaining, is the place where microeconomics enters the picture. It is also an important point where Theory of Mind and Cosmides' (1989) "cheater detection" enter: one is more inclined to agree to an exchange if one believes the other's assertions of value are made in good faith.

Recall that freely chosen reciprocation has a normative value attached to it: "It's good to reciprocate nice things" (rule (37)). Exchange has no such normative principle, since both actors are acting in tandem. However, exchange transactions involve a commitment to a joint task, which creates mutual obligations. And failing to fulfill an obligation is bad (i.e. of negative N-value). Thus in freely chosen reciprocation the normative principle is "It's good to reciprocate nice things", and in exchange it amounts to "It's bad to defect on an exchange."

5.2. Exchange of objects. The paradigm case of exchange, of course, is *trade*, or exchange of objects. And a special case of trade is monetary transactions such as buying and selling. It is now easy to build these up from exchange. Consider a sentence like *Bob traded Sue his goat for a coat.* Notice that this contains the telltale *for* of reciprocity, though this time the phrase following *for* denotes an object rather than an action. Nevertheless, an action is implicit: not only is Bob giving Sue his goat, but Sue is giving Bob a coat. Plugging these actions into schema (43), we get (47).

(47) Bob traded his goat to Sue for a coat.
[BOB GIVE GOAT TO SUE] EXCH [SUE GIVE COAT TO BOB]

We want the understanding of this transaction to involve the R-values of the goat and the coat. Recall the intuitive definition of R-value: something has R-value to the degree that it's good to have it – i.e. to the degree that having it has A-value. So we can relate R-value to A-value by the

equations in (48).

(48) a. Objective: R-VAL (OBJECT) = A-VAL (YA HAVE OBJECT)
'The R-value of an object is the A-value of having it.'

b. Subjective: R-VAL (OBJECT, PERSON) =
A-VAL ([PERSON HAVE OBJECT], PERSON)
'The R-value of an object to a particular person is the A-value to that person of having it.'

Now, since giving an object away changes who has it, the act of giving is of negative A-value to the giver and of positive A-value to the recipient.[11] Thus for the exchange of objects, we can couch the inference rules for exchange in terms of R-value as follows:

(49) [[X GIVE Z TO Y] EXCH [Y GIVE W TO X]]$^{\alpha}$ $\Longrightarrow_{\text{default}}$
a. Objective:
R-VAL (Z) = R-VAL (W)
b. Subjective:
R-VAL (W, X) - R-VAL (Z, X) + A-VAL (α, X) > 0
R-VAL (Z, Y) - R-VAL (W,Y) + A-VAL (α, Y) > 0

5.3. Linguistic expression of exchange of objects and actions. We now come to a topic of hoary antiquity in linguistics (e.g. Gruber 1965): the semantics of trading, buying, and selling. The notation for exchange so far is entirely symmetrical between the two participants. However, the linguistic expression of exchanges is asymmetrical, focusing on one side and backgrounding the other. For instance, in (47), Bob's giving of a goat to Sue is foregrounded and Sue's reciprocal giving of a coat to Bob is represented only by the *for*-phrase. In order to reflect this difference in prominence, the mapping of the exchange structure into syntax has to mark one of the actions specially. (50) indicates the difference by underlining (much as we marked relative prominence of stimulus and experiencer in chapter 7).[12]

---

[11]Notice that this entailment is not true if the entity given away is *information.* If I give you my goat I don't have it any more; but if I tell you (give you information) that my party is on Saturday, I haven't thereby forgotten it. On the other hand, passing on information may reduce its R-value to me, for instance if I tell you where the treasure is hidden, or if I let you copy my manuscript which I hope some day to publish profitably. These considerations lie behind the need for spying and for copyright and patent law.

[12]As brought to my attention by (I believe) Kara Hawthorne, this account does not work for examples like (i)-(ii).
(i) The kids/Bob and Sue traded coats.
(ii) Bob traded coats with Sue.
A proper treatment of these examples calls for a more sophisticated analysis of how joint tasks are mapped into linguistic form. The syntactic form of (i)-(ii) parallels other expressions of joint tasks

(50) Syntax/phonology: X trade Z (to Y) (for W)
Conceptual structure: [[X GIVE Z TO Y] EXCH [Y GIVE W TO X]]

For the specific case in which one of the exchanged objects is money, we get the specialized verbs *sell* and *pay*.

(51) a. Syntax/phonology: X sell Z (to Y) (for [amount of money])
Conceptual structure: [[X GIVE Z TO Y] EXCH [Y GIVE MONEY TO X]]
b. Syntax/phonology: X pay [amount of money] (to Y) (for W)
Conceptual structure: [[X GIVE MONEY TO Y] EXCH [Y GIVE W TO X]]

The verb *buy* is parallel, except that its subject is recipient of the transfer that is being foregrounded – if you *buy* a book, someone else gives the book to you. However, in addition you are understood as acting intentionally, i.e. instigating the transaction and not just being a passive recipient. Where does this come from in the meaning of the verb? The answer comes from considering the full expression (42) of which *EXCH* is an abbreviation: X is a participant in an intentional joint action, hence acting intentionally. Thus the only remaining question is how to notate the foregrounding in the formal representation. For the moment, I'll just double underline the character in question, indicating that it is the foregrounded character in the foregrounded part of the transaction. There are other possible solutions,[13] but they take us beyond the scope of this discussion, which is focused for the moment on exchanges of R-value.

(52) Syntax/phonology: Y buy Z (from X) (for [amount of money])
Conceptual structure: [[X GIVE Z TO Y] EXCH [Y GIVE MONEY TO X]]

---

such as (iii)-(iv).
(iii) The kids/Bob and Sue played a duet. The kids/Bob and Sue baked a cake together.
(iv) Bob played a duet with Sue. Bob baked a cake with Sue.
It looks as if in (i) and (iii), the conjoined or plural subject maps into the set of joint actors; in (ii) and (iv), the subject maps into a foregrounded member of the set of joint actors, and the object of *with* maps into the other members of the set. Examples (i)-(ii) are trickier, though, because of the bare plural *coats*. I leave this fascinating problem for future research.

[13]What is missing from the present account is an explanation of why the preposition *from* shows up in (52). A proper answer analyzes the semantic structure of *X GIVE Z TO Y* into *Z* GO_{poss} *FROM X TO Y*, 'Y changes possession from X to Z'. If X is foregrounded as subject, then Y remains after the verb, marked either by the preposition *to* or by being placed in indirect object position. If Y is foregrounded as subject, then X remains after the verb, marked by the preposition *from*. Whichever one is subject comes to be understood as the foregrounded instigator of the action, because of the principle that subjects are interpreted as intentional whenever possible (chapter 6). Some of these details are spelled out in Jackendoff 1990 (189-194); however the interaction with joint action is new here. Even with all this, the *to* with *hire oneself out to* in (54b) is problematic and suggests there is still more going on.

There are also verbs that mix money and actions in exchanges, as in (53).

(53) a. Bob hired Sue (for $500) to paint the house.
b. Bob paid Sue ($500) to paint the house.
c. Sue hired herself out to Bob (for $500) to paint the house.

All of these are exchanges of money for an action, but each verb requires the characters to be realized differently in syntax.

(54) a. Syntax/phonology: X hire Y (to ACT_y) (for [amount of money])
X pay Y ([amount of money]) to ACT_y
Conceptual structure: [[X GIVE MONEY TO Y] EXCH [Y ACT_y]][14]
b. Syntax/phonology: Y hire self out (to X) (for [amount of money]) (to ACT_y)
[[X GIVE MONEY TO Y] EXCH [Y ACT_y]]

Next, notice that *Sue painted the house* contains no intimation that this is for anyone's benefit (i.e. that the action is of positive A-value to anyone). However, in the context of the constructions in (53), it is understood that Bob benefits from Sue's painting the house. This follows directly from the semantics of exchange, in particular from inference rules (46) and (49).[15]

Next consider the noun *price*. The *price of X* is 'the amount of money for which X can be exchanged', that is, it foregrounds the money in a monetary transaction, relating it to the other object or action being exchanged, while leaving the actors implicit. From *price* we can build the meaning of *expensive* and *cheap*: 'having a high/low price'. An even more complex case is the verb *owe*. Consider *Bob owes Sue $500 for painting the house*. This expresses Bob's obligation to Sue, where this obligation arises from a transaction in which Sue has carried out her side of the deal and Bob has not. In other words, this verb pulls in the entailments of EXCH as part of its meaning.

The larger point here is that all the words *trade*, *buy*, *sell*, *pay*, *hire*, *hire oneself out*, *price*, *expensive*, and *owe* avail themselves of the very same conceptual structure of exchange, while foregrounding different parts and leaving implicit other different parts (such as giving as a specific action, money as a specific object transferred). This leads toward a "frame-based" theory

---

[14]The verb *bribe* is exactly the same as *hire*, except that it adds the presupposition either that the normative value of ACT_y is negative or that Y is under obligation not to do ACT_y.

[15]In addition to verbs of exchange, there is another way in English to express A-value of an action to a non-participant in the action:
(i) Sue painted the house **for Bob**. [positive A-value for Bob]
(ii) Bob's car broke down **on him**. [negative A-value for Bob]
See Jackendoff 1990 (185-187) for one possible treatment (not necessarily exactly what I would adopt now, but close enough).

of lexical meaning, along lines suggested by Fillmore and Atkins 1992: the notion of a transaction is a common conceptual frame that can be evoked from different perspectives and with different specializations by using different lexical items. It is worth pointing out that this frame can be evoked in some further ways.

(55) a. Sue painted the house for $500.
b. If you give me your goat, I'll give you my coat.

In (55a), just the addition of *for $500*, using the *for* of reciprocation, brings into play the entire conceptual machinery of a transaction (Jackendoff 1990, 191-194). And although (55b) has the form of a conditional, it is understood (presumably by implicature) as an offer to engage in a transaction. In both cases all the entailments about value emerge clearly.

6. Fairness and selfishness; Fiske's Four Elementary Forms of Human Relations

The notion of A-value makes it straightforward to state a version of *fairness*: an action is fair if it is equally good or bad for everyone, i.e. if its A-value (positive or negative) to each of the individuals it affects is the same.

(56) Y acts fairly toward $X_1,...,X_n$:
For all X_i, X_j: [A-VAL ((Y ACT), X_i) = A-VAL ((Y ACT), X_j)]
'Y's act is as good for X_i as it is for X_j'

This can be played out in various ways: Y's act may be a single action that impinges on everyone at once. Or it may be a composite of multiple sub-actions at different times, each impinging on a different individual: if you do such-and-such to X_1 this time, you'd better do the same thing to X_2 the next time. Principle (56) undergirds the notion of "equality under the law", where what's often at issue is who counts as an X: everybody, members of one's own group, only men, only white men, only white men who own property, and so on.

Another prevalent pattern is distribution by rank: an action fits this pattern if the higher your rank, the better treatment you get.

(57) Y acts according to distribution by rank among $X_1,...,X_n$:
For all X_i, X_j: [X_i OUTRANKS X_j ==> A-VAL ((YACT), X_i) > A-VAL ((Y ACT), X_j)]
'If X_i outranks X_j, then Y's act is better for X_i than it is for X_j'

In turn, OUTRANK stands in for an inequality of value – either prowess (P-value), dominance, honor (N-value), or general esteem (E-value). This mode of distribution is appropriate for awarding honors and prizes. But it's of far broader social application: the top dog receives the best chair, the best mate, the most food, and the most lenient punishment; those on the lowest rungs get the fewest resources, the most unpleasant work, the most severe punishment, and so on.

And of course this principle is amply attested in social animals.

Two other important patterns of behavior are acting selfishly and acting altruistically, which come out like this:

(58) a. Y acts selfishly toward $X_1,...,X_n$:
For all X_i: [A-VAL ((Y ACT), Y) > A-VAL ((Y ACT), X_i)]
'Y's action is better for himself than it is for anyone else'

b. Y acts altruistically toward $X_1,...,X_n$:
For all X_i: [A-VAL ((Y ACT), X_i) > A-VAL ((Y ACT), Y)]
'Y's action is better for everyone else than it is for him'

With these pieces in hand, we turn to the central thesis of Alan Fiske's *Structures of Social Life* (1991): the hypothesis that human societies have exactly four ways to distribute goods, labor, and responsibility, and that these four ways are universal innate structures within the human social cognitive capacity. Cultures differ, not in having one of these structures rather than another, but rather in how they distribute the use of the four structures over different contexts. Fiske arrives at this hypothesis through painstaking analysis of a vast number of social institutions in disparate cultures.

Fiske's four "elementary forms of human relations" are the following:

- In *Communal Sharing*, each member of a group shares equally in benefits and responsibilities, or, within limits, "from each according to his abilities, to each according to his need." The prototype case is distribution of food at a family meal (in our culture, at any rate).
- In *Authority Ranking*, benefits and responsibilities are distributed according to rank.
- In *Equality Matching*, equality among group members is guaranteed by each member doing or receiving exactly the same thing. Prototype cases are turn-taking and voting.
- In *Market Pricing*, participants exchange resources and/or labor according to negotiated agreement.

The basic pieces of these frames are evident in what we have already done. There are various ways the frames can be construed within our formalism; I'll lay out versions that seem plausible to me.[16] Each of the frames can be considered a norm: "One should distribute benefits, costs, and responsibilities in such-and-such a way." That is, each frame assigns a positive N-value to a particular sort of distribution. Since the four norms often conflict with each other, the

---

[16]Fiske himself formalizes the differences among the frames in terms of different and incommensurable mathematical systems, an approach that I find questionable in cognitive terms. The present approach, grounded in independently necessary notions of value and joint action, strikes me as closer to the right approach. It is not yet clear to me how to show this is more than a matter of taste.

cultural element of the picture enters by contextualizing the frames: "In the following sorts of activities [culturally specific list here], one should distribute benefits, costs, and responsibilities in such-and-such a way." Thus an important part of learning a culture is learning the list of activities that goes with each frame.

Viewed this way, Communal Sharing might be encoded as in (59). It is clearly related to the notion of "acting fairly" worked out in (56); however, it is not saying how to act fairly, rather it is giving fairness a normative value. To spare us all the notational complexity added by quantifying over groups of arbitrary size, I'll state (59) and subsequent principles in terms of groups containing only two members. (I leave the fully quantified version as an exercise for masochistic readers.)

(59) Communal Sharing:
For actions in the category ACT_{CS}, and a group {X, Y}:

$$\text{N-VAL}\left(\begin{bmatrix}\text{YA } ACT_{CS} \\ \text{A-VAL}(\alpha, X) = \text{A-VAL}(\alpha, Y)\end{bmatrix}^{\alpha}\right) = +$$

"It's good to act in a way that is equally good for X and Y."

If the action in (59) is distributing resources to the group, the principle has two possible construals. In the simpler construal, each person in the group receives the same amount of resources. In a more sensitive, subjective construal, the differing needs of individuals are taken into account, that is, the subjectivity of A-value is taken seriously. Similarly when the action is collecting resources from the group: the simpler construal is a uniform tax, and the more sensitive construal is a progressive tax.

Next consider the situation when the actor Y is him/herself a member of the group. Given the disposition to act altruistically toward members of one's own group, (59) creates a pressure for better-off members of the group to share with less well-off members. In principle, it also creates pressure for less well-off members to take from better-off members, but this is perhaps inhibited by the opposing disposition to behave kindly toward group members.

A final special case is when the actor Y is the group as a whole, perhaps undertaking a task for the benefit of the group – preparing a communal meal, raising money for a playground and so on. Here (59) creates pressure for each individual to pitch in equally and not shirk, so the costs in labor or money are spread evenly. Of course, it's up to each individual how much to yield to this pressure. Moreover, the "sensitive" construal of the rule may license one to say, "Well, so-and-so has more time to give than I do" or "Well, so-and-so cares more about this cause than I do", justifying giving less effort.[17]

---

[17]Fiske discusses other important manifestations of Communal Sharing, such as collective ethnic identity and sense of group unity. I am inclined to see these as symptoms of group membership per se. They are only connected with Communal Sharing in the sense discussed here because the typical domain for Communal Sharing is the social group.

Next consider Authority Ranking, which is clearly related to the notion of rank-based distribution worked out in (57).

(60) Authority Ranking:
For actions in the category ACT_{AR}, and a group {X, Y}:

$$\text{N-VAL}\left(\left[\begin{array}{l}\text{YA } ACT_{AR} \\ \text{X OUTRANKS Y} \Rightarrow \text{A-VAL}(\alpha, X) > \text{A-VAL}(\alpha, Y)\end{array}\right]^{\alpha}\right) = +$$

"It's good to act in a way that reflects individuals' relative status/merit."

Authority Ranking is what justifies bosses getting paid more than workers, rewarding or honoring individuals for merit, giving first prizes that are bigger than second prizes, and so on. What's important here is that, because Authority Ranking is the operative norm in these situations, such disparities are acceptable even to low-ranked individuals.

On the other hand, Authority Ranking does not sanction dominant individuals taking resources from subordinates by force; that is, it is not synonymous with oppression. Oppression instead falls under the negative version of rule (22): "It's bad to be mean to people." It is possible that much of the delicate dynamic between dominants and subordinates is a consequence of the interplay of (60) and (22). A dominant is entitled to more resources and respect, but, not wanting to present the appearance of oppression, is motivated to show a bit more respect and generosity to subordinates, not to show off too much. (Recall also principle (41), "You shouldn't have too high an opinion of yourself.")

Let's next turn to Equality Matching. The present context enables us to see a distinction between Equality Matching and the previous two norms, not (to my knowledge) noticed by Fiske: it always takes place in the context of joint tasks, such as voting to elect an official, taking turns helping to harvest a field or carpool the children, and participating in a rotating credit association. Thus, unlike Communal Sharing and Authority Ranking, the actor is necessarily identical with the group being affected by the action. Here is how it comes out:

(61) Equality Matching:
For actions in the category ACT_{EM}, and a group {X, Y}:

$$\text{N-VAL}\left(\left[\begin{array}{l}\{X, Y\}\ ACT_{EM} \\ [\text{FROM } [\{X, Y\}\ \text{COM } \alpha]] \\ X\ ACT_x = Y\ ACT_y \\ \text{A-VAL}(\alpha, X) = \text{A-VAL}(\alpha, Y)\end{array}\right]^{\alpha}\right) = +$$

"It's good for everyone in a jointly intended task to do exactly the same thing and benefit equally."

Equality Matching thus differs from Communal Sharing in two respects: it applies specifically to joint tasks, and it specifies not only equal effect on the members of the group, but also equal participation. In this respect it is the most rigid of Fiske's four frames – but its rigidity is an excellent way to coordinate certain sorts of joint actions. In the usual cases where it involves

turn-taking, in effect it is institutionalized reciprocation.

The fourth frame, Market Pricing, is of course derivative of the dynamics of exchange transactions, and therefore, like Equality Matching, it can only be applied to joint tasks. It's crucial to Market Pricing that, although the participants are cooperating in the transaction, they're also competing, each trying to get the better of the deal. Thus it has to be conducted in the context of a selfish stance: it's okay to get more than the other guy. The selfish stance is easily derived from (58a); the normative value in (62) is allowed to be either neutral or positive (my interpretation of "it's okay to ...").

(62) Selfish stance:

$$\text{N-VAL}\left(\begin{bmatrix}\text{X ACT}\\ \text{A-VAL}(\alpha, \text{X}) > \text{A-VAL}(\alpha, \text{Y})\end{bmatrix}^{\alpha}, \text{TO X}\right) \geq 0$$

"To X, it's okay for him to act in a way that is better for himself than for others."

Let's abbreviate this as *SELFISH (X ACT)*. Now Market Pricing amounts to condoning being selfish in conducting exchanges. With the abbreviations for selfishness and for exchanges ((substituting (43) for (42)), the norm can be stated as follows:

(63) Market Pricing:

$$\text{N-VAL}\left(\begin{bmatrix}\text{X ACT}_x\\ \text{SELFISH}(\alpha)\end{bmatrix}^{\alpha} \text{EXCH} \begin{bmatrix}\text{Y ACT}_y\\ \text{SELFISH}(\beta)\end{bmatrix}^{\beta}\right) = +$$

"It's good to engage in exchanges selfishly"

Notice that since Market Pricing condones acting selfishly, it is best conducted between individuals for whom there is no conflicting norm to act selflessly. Thus, although one may conduct exchanges with members of one's family, it is less likely that one will conduct strict Market Pricing transactions with them than with members of another group to whom one owes no allegiance. In fact, following Jacobs's (1994) conjecture, it is plausible that Market Pricing arose in human society as a way to productively inhibit natural intergroup aggression, for mutual profit.

7. Slipping and sliding

This section is about situations where the value system is a bit slippery and leads to peculiar results – which we all recognize.

7.1. Reasons for work. A staple of Marxist analysis is the alienation of workers from their work because of the need to work for wages. Here is how this situation looks in the present framework.

This issue concerns the purposes or motivations behind doing particular actions. Recall the discussion of purposes in chapter 6. There we analyzed purposes as (64).

(64) X acts in order for Y to happen

$$\left[\begin{array}{l}\text{X ACT} \\ \text{[FROM } \left[\begin{array}{l}\text{X INTEND } [_{\text{Situation, +Action}}\ \beta] \\ \text{[FROM [X WANT Y]]}\end{array}\right]\end{array}\right]^{\beta}$$

Decoding, this says that X is performing ACT intentionally, and that this intention is caused by (or grows out of) X's desire for Y to happen.

With this formalization in place, we can plug in different purposes for X acting. First consider doing a task for the pleasure of the task. "The pleasure of the task" can be rephrased as "having a positive experience of doing the task", which can be coded in terms of the macrorole function EXP^+ of chapter 7. It looks like (65).

(65) Working for the pleasure of the task

$$\left[\begin{array}{l}\text{X ACT} \\ \text{[FROM } \left[\begin{array}{l}\text{X COM } \beta \\ \text{[FROM [X WANT [X } EXP^+ \beta]]]\end{array}\right]\end{array}\right]^{\beta}$$

'X is acting intentionally out of a desire for the action to be pleasurable'

In turn, rule (21b) says that *X EXP⁺ Y* ('Y is a pleasurable experience for X') usually entails that Y has positive A-value for X. Thus, the work is being motivated by its potential for A-value.

(66) X EXP^+ [X ACT] $\Longrightarrow_{\text{default}}$ A-VAL ([X ACT], X) = + [= one case of (21b)]

Of course, since not every task with a purpose succeeds in fulfilling the purpose, it's certainly possible that one can work and get no pleasure out of it.

Working for wages comes out as (66).

(66) Working for wages

$$\left[\begin{array}{l}\text{X ACT} \\ \text{[FROM } \left[\begin{array}{l}\text{X COM } \beta \\ \text{[FROM [X WANT [[}\beta\text{] EXCH [Y GIVE MONEY TO X]]]]}\end{array}\right]\end{array}\right]^{\beta}$$

This is all right: working in exchange for money is per se not a problem. However, another perspective on this situation is that X is "giving Y his work/labor." Such a reconceptualization comes out as (67).

(67) X exchanges labor for money

[[X GIVE [X ACT] TO Y] EXCH [Y GIVE MONEY TO X]]

Now the inference rule (49b) for reciprocal giving kicks in, and we derive (68).

(68) R-VAL (X ACT) = R-VAL (MONEY)

But now the action is treated as something that has R-value, i.e. as a commodity – precisely as Marxist theory treats it.

A third motivation ought to be mentioned: working to gain esteem (or fame). This looks like (69).

(69) Working to increase one's esteem

$$\left[\begin{array}{l} \text{X ACT} \\ \text{[FROM } \left[\begin{array}{l} \text{X COM } \beta \\ \text{[FROM [X WANT [(E-VAL (X)) INCREASE]]]} \end{array}\right] \end{array}\right]^{\beta}$$

The routes to gaining esteem from work have already been explored in section 3: through gaining in virtue (N-value), in wealth, in dominance (power), or in prowess. This all sounds very familiar.

Nothing precludes these three motivations being conjoined in differing proportions – after all, one's actions need not have a single unitary purpose. And in practice, most of us (at least most of us who will be reading this book) work for some mixture of these motivations, some doing it more for the fame, some more for the money, and some more for the pleasure of the work itself. On the other hand, in situations of scarcity, presumably the situation Marx was addressing, the economic motivation may be the only option open.[18]

7.2. Moral monsters: Can't buy me love. Sections 4 and 5 explored two different kinds of reciprocal action: freely chosen reciprocation and agreed-upon exchange. An instance of the former is when we exchange gifts; an instance of the latter is when we trade commodities. The outward appearances may be exactly the same. But conceptually and affectively they are quite different, and they have different entailments, in particular with respect to the normative value accorded the reciprocator.

When the actions in question consist in giving objects to the other actor, both free reciprocation and agreed-upon exchange are perfectly fine kinds of actions. However, consider the second sort of reciprocation we discussed in section 4: reciprocating an estimable act by bestowing esteem on the actor. (70) repeats the relevant principle.

(70) *"It's good to honor people who have done estimable things"*

$$\left[\begin{array}{l} \text{X ACT}_1 \\ \text{N/P-VAL } (\beta) = + \end{array}\right]^{\beta} \Longrightarrow \text{N-VAL } \left(\left[\begin{array}{l} \text{Y EXPRESS (E-VAL (X)} = +) \\ \text{RECIP } (\beta) \end{array}\right] \right) = +$$

---

[18]Schwartz 1993 describes an experimental situation in which subjects can be manipulated into doing a task for its own pleasure or for monetary reward, and shows that they perform worse in the latter case.

What happens if an expression of esteem is plugged into an exchange formula?

(71) [X GIVE MONEY TO Y] EXCH [Y EXPRESS (E-VAL (X) = +)]

Recall that *EXCH* abbreviates a jointly intended action by X and Y. Thus (71) says that X and Y agree that X will pay Y to praise X. This is pernicious in several respects. In the usual case, Y's action is presented in such a way that it appears to originate from equation (70) instead: when X is buying esteem. A clear instance of this situation is product endorsement: "Hello, I'm so-and-so, and I'm here to tell you about how wonderful product Y/candidate Z is." In such a situation, the motivation for the actions involve a mixture of two normative principles, what Jacobs 1994 calls a "moral monster."

With a little patience, it's possible to express the deviance of this situation in terms of the present formalism. Chapter 7 provided a treatment of 'it looks like such-and-such is the case'. (72) is a stripped-down version.

(72) *"It looks (to people) like Z is the case"*
[YA SENSE Z]

This can be incorporated into a modifier of Y's action in (71). For a first step, we leave *Z* unspecified.

(73) X pays Y to praise X in a way that looks like Z

$$[\text{X GIVE MONEY TO Y}] \ \text{EXCH} \begin{bmatrix} \text{Y EXPRESS (E-VAL (X) = +)} \\ \text{YA SENSE Z} \end{bmatrix}$$

Now we plug in for Z the formula for reciprocal praise from (70).

(74) X pays Y to praise X in a way that looks like freely chosen praise for X's deed

$$[\text{X GIVE MONEY TO Y}] \ \text{EXCH} \begin{bmatrix} \text{Y EXPRESS (E-VAL (X) = +)} \\ \text{[YA SENSE} \begin{bmatrix} \alpha \\ \text{RECIP} \ \begin{bmatrix} \text{[X ACT]}^{\beta} \\ \text{N-VAL } (\beta) = + \end{bmatrix} \end{bmatrix} \end{bmatrix}^{\alpha}$$

Everybody knows this is Bad; (75a) notates this intuition. In turn, (74) is a jointly intended action by X and Y, and the N-value of an action reflects on its actors, Thus we derive the intuition that X and Y are bad too, as shown in (75b).

(75) a. *"Can't buy me love"*
N-VAL ((74)) = –
b. N-VAL ((74)) = – $\Longrightarrow_{\text{default}}$ N-VAL ({X, Y}) = –
$\Longrightarrow$ {N-VAL (X) = – ; N-VAL (Y) = –}

If we fill out all the abbreviations, (75a) is a very complex expression indeed, every piece of which is motivated. There seems no way to eradicate the complexity of the relations that it encompasses. Yet everyone has clear and immediate intuitions about this case. So the question might be: Are there simpler principles from which the normative judgment in (75a) can be derived as a theorem? One is the principle that deception is normatively negative: X and Y are colluding to conceal Y's real motivation. But this principle is of course defeasible: "white lies" are okay to help someone save face, and deceiving the enemy is often desirable.

But a more specific principle relevant to this case comes from considering one's intuitions about (71) *even when* one knows that Y is being paid: there is something "dirty" about praise uttered for pay – it somehow does not count as praise, or, we might say, the praise is not sincere. Moreover, there is also something wrong if Y is being paid by X to praise (or denounce!) some other person Z. Perhaps the operative principle is that expressions of praise should be altruistic, or that they should not be made with the goal of personal gain. Since an exchange transaction is by definition undertaken with the goal of personal gain (rule (45b)), an exchange like (71) is normatively bad. (76a,b) are two possible ways of expressing this principle. They subsume not only exchanges like (71) but also such activities as toadying (or what we used to call brown-nosing): expressing praise in the hope of eliciting reciprocal favors.

(76) a. *"Praise should not be expressed for personal gain"*

$$\text{N-VAL}\left(\left[\begin{array}{l}\text{Y EXPRESS (E-VAL (X)} = +) \\ \text{[FROM (Y WANT (A-VAL (}\left[\begin{array}{l}\text{EVENT} \\ \text{COMPOSED OF } \alpha,\ldots\end{array}\right], \text{Y)} = +))]\end{array}\right]^{\alpha}\right) = -$$

'It's bad to express praise whose purpose is to be part of an event that is good for you'

b. *"Praise should not be expressed selfishly"*

$$\text{N-VAL}\left(\left[\begin{array}{l}\text{Y EXPRESS (E-VAL (X)} = +) \\ \text{FROM (SELFISH } (\alpha))\end{array}\right]^{\alpha}\right) = -$$

And perhaps there are still simpler principles from which these follow as a theorem.

7.3. Moving between subjective and objective.[19] From the outset, the exposition here has distinguished between "objective" and "subjective" notions of value, paralleling the "objective" and "subjective" notions of evaluation in chapter 7. In an important sense, "subjective" value is more true to life: being of value, like being of interest, or being boring, is fundamentally a relation between an object and a perceiver. Yet experientially, "objective" value is every bit as valid: certain actions are morally repulsive and certain people are of high esteem, and this is not a fact

---

[19]The arguments in this section and the next originated in my talk for a 1996 public symposium called "Apples and Oranges", sponsored by the New Hampshire Humanities Council. In an important sense this entire chapter arose out of an attempt to make the intuitive argument formally more rigorous.

about my own perception of them. In fact, as remarked in chapter 4, an important fact about moral systems is that they are conceived of as objective, universal, and timeless, which is why the term "moral relativism" is taken by many to be self-contradictory or tantamount to "amoral." Moreover, as pointed out by Berger and Luckmann 1966, values are usually *taught* in their objective form: "This is what we (the group) do; this is what's good; this is what's bad."

Two obvious questions are: Why have distinct subjective and objective systems in cognition? And what does one have to do with the other? Part of the answer came up already in section 2.1. On one hand, value is experienced as a property of an action, just as objectively as its duration or loudness. One's own contribution to the judgment of value is completely transparent, just like one's own contribution to the judgment of duration or loudness. On the other hand, it is important to be able to account for individual differences in values, and this is what the subjective system allows us to do: this object is worthless to you and valuable to me, this action is good for you and of no import to me. To recognize these differences, however, requires Theory of Mind (including my own) – always a cognitive stretch. Still, we may be rightly suspicious of the seeming redundancy, with identical predicates in the two systems, differing only in whether they have an experiencer argument.

In practical reasoning, we jump readily between the two systems. Something like (77) seems to be the appropriate rule of inference.

(77) Objectification and Subjectification
X-VAL (Y, PERSON) $<==>_{default}$ X-VAL (Y)

Reading from left to right: I like Y, or Y is good for me or valuable to me, therefore Y is good or valuable. That is, my own judgments by default warrant a judgment of objective value. Alternatively, if I don't know anything about Y and you tell me you like it, or it's good for you or valuable to you, then your judgments by default warrant a judgment of objective value. That is, from left to right, (77) represents the objectification of value.

Why should it be important to arrive at an objective value? The reason is that then the rule can be read from right to left to predict someone else's reactions to Y. If it's good or valuable, then it's reasonable to believe it will be good for you and me or valuable for you and me. By contrast, subjective values are useful when I want to predict that you and I will react differently to an object. However, I can't predict the object's subjective value to you without evidence about your reactions. My own reactions are of no relevance on this plane. Without such evidence, I fall back on (77) to predict your reaction.

So in practice, we slip between the two systems as it is convenient. We strongly prefer the more predictive objective system when possible, but we can easily drop into the subjective system when we have evidence of difference. This is hardly logical reasoning. But it's what we do.

However, there is a value judgment that goes with being wrong:

(78) *"It's bad to be wrong"*

a. [X-VAL (Y) = α] AND [X-VAL (Y, PERSON) = -α]
$\Longrightarrow_{default}$ E-VAL (PERSON) = -

b. [$_{Situation}$ Y] AND [PERSON COM [NOT Y]] $\Longrightarrow_{default}$ E-VAL (PERSON) = -

That is, someone whose judgment of value (78a) or of the truth of a proposition (78b) does not coincide with objective value or actual truth is to be negatively esteemed. This raises a difficult practical problem: when I disagree with you, the question arises as to who has control of objective value and truth. If I trust your judgment, say because you're an authority figure, then there must be something wrong with *me*, and my self-esteem goes down.[20] Thus (78) is another piece in the puzzle raised in chapter 7, of how *That problem's not interesting* can trigger self-doubt.

The more standard situation, though, is when *I* am in possession of objective truth and values, and I thereby think less of you. This is typically the case when two cultures encounter one another and each denounces the other as uncultured, savage and lacking in values. As in chapter 4, there is no need to recount the unpleasant consequences: between generations, between religions, between religion and science, between the sciences and the humanities, even between subcultures of a discipline. The only way for dialogue to take place is if both protagonists are capable of switching into the subjective system for their own judgments as well as the other's.

Another variant on (78) pertains to actions. (This is also a variant on rule (23), "Doing good things makes you good.")

(79) *"You're bad to do something wrong"*

[N-VAL (YA ACT_1) = -] AND [X ACT_1] $\Longrightarrow_{default}$ N-VAL (X) = -

In turn, lowering someone's normative value lowers their esteem-value, ending with the same conclusion as (78a,b). Again, this reasoning relies on operating in the objective value system, which takes normative value as absolute (or "what is done").

7.4. Deserving – from whom? I'd like to end with an especially peculiar line of heuristic reasoning that results from objectification. Let's start by repeating the inference rules for reciprocation (with only the variables denoting the actors altered):

(80) a. *"It's good to reciprocate nice things"* (=37a)

$$\begin{bmatrix} Z\ ACT_1 \\ A\text{-}VAL\ (\beta, Y) = + \end{bmatrix}^{\beta} \Longrightarrow \text{N-VAL}\ \left(\begin{bmatrix} Y\ ACT_2 \\ RECIP\ (\beta) \end{bmatrix}, Y \right) = +$$

---

[20]In the latter case, it's an interesting and important question how I "repair" my knowledge base, replacing "It's good/true" with "It's bad/false but I used to believe the opposite." But this goes beyond the scope of the present enterprise.

b. *"It's good to retaliate for bad things"* (=38a)

$$\left[\begin{array}{l}\text{Z ACT}_1 \\ \text{A-VAL}\,(\beta, \text{Y}) = -\end{array}\right]^{\beta} \Longrightarrow \text{N-VAL}\,(\left[\begin{array}{l}\text{Y ACT}_2 \\ \text{RECIP}\,(\beta)\end{array}\right], \text{Y}) = +$$

c. *"It's good to make up for harming someone"* (=37b)

$$\left[\begin{array}{l}\text{Y ACT}_1 \\ \text{A-VAL}\,(\beta, \text{Z}) = -\end{array}\right]^{\beta} \Longrightarrow \text{N-VAL}\,(\left[\begin{array}{l}\text{Y ACT}_2 \\ \text{REST}\,(\beta)\end{array}\right], \text{Y}) = +$$

Suppose we look at these from the point of view of the individual being helped or harmed by the second action. In all the expressions in (80) this individual is Z. In ordinary English, these inferences then might be phrased as (81).

(81) a. If Z does something good for Y, then Z should be rewarded by Y for his (Z's) action.
b. If Z does something bad to Y, then Z should be punished by Y for his (Z's) action.
c. If Y does something bad to Z, then Z should be compensated by Y for his (Y's) action.

This can be rephrased in a way that further emphasizes Z's point of view.

(82) a. If Z does something good for Y, then Z *deserves* to be rewarded by Y for Z's action.
b. If Z does something bad to Y, then Z *deserves* to be punished by Y for Z's action.
c. If Y does something bad to Z, then Z *deserves* to be compensated by Y for Y's action.

However, this shift of perspective calls for an innovation in the notation. The N-values in (80) pertain to actions by Y. Thus if there is anyone whose personal N-value is affected by these actions, it is Y, not Z. Yet (82) (and to some extent (81)) express the situation in terms of a sort of "moral debt" owed to Z. As observed in section 2.4, one can accrue N-value only by performing actions, not by having actions done to one. By contrast, one's "moral debts" involve having good or bad things done to one. In order to distinguish this sort of value from the usual type of normative value, let us call it MD-value. Introducing a notation (83) for MD-value, we can add inferences (84) to those in (80).

(83) Y owes a moral debt to Z to perform ACT_1; Z deserves to have Y do ACT_1 on his (Z's) behalf
$\text{MD-VAL}\,([\text{Y ACT}_1], \text{Y}, \text{TO Z}) = +$

(84) a. *"Someone who does good things for others deserves to be rewarded by them"; "One good turn deserves another"*

$$\left[\begin{array}{l}\text{Z ACT}_1 \\ \text{A-VAL}\,(\beta, \text{Y}) = +\end{array}\right]^{\beta} \Longrightarrow \text{MD-VAL}\,(\left[\begin{array}{l}\text{Y ACT}_2 \\ \text{RECIP}\,(\beta)\end{array}\right], \text{Y}, \text{TO Z}) = +$$

b. *"Someone who does bad things to others deserves to be punished by them"*

$$\left[\begin{array}{l} \text{Z ACT}_1 \\ \text{A-VAL}\,(\beta, \text{Y}) = - \end{array}\right]^{\beta} \Longrightarrow \text{MD-VAL}\,\left(\left[\begin{array}{l} \text{Y ACT}_2 \\ \text{RECIP}\,(\beta) \end{array}\right], \text{Y, TO Z}\right) = +$$

c. *"Someone who has bad things done to them by others deserves to be compensated by them"*

$$\left[\begin{array}{l} \text{Y ACT}_1 \\ \text{A-VAL}\,(\beta, \text{Z}) = - \end{array}\right]^{\beta} \Longrightarrow \text{MD-VAL}\,\left(\left[\begin{array}{l} \text{Y ACT}_2 \\ \text{REST}\,(\beta) \end{array}\right], \text{Y, TO Z}\right) = +$$

Now it has to be admitted that the notion of moral debt is altogether suspect from a logical point of view. Nevertheless it's altogether intuitive, particular if Z in (84a,c) is me and in (84b) is someone else:

(85) a. If I do something good for you, I deserve to be rewarded by you.
b. If you do something bad to me, you deserve to be punished by me.
c. If you do something bad to me, I deserve to be compensated by you.

But further steps are possible along this line. Having taken Z's point of view, it is natural to drop Y out of the picture: it does not matter any more exactly who *owes* the moral debt. (86) is a partially objectivized version of (84).

(86) a. *"Someone who does good things for others deserves to be rewarded"; "One good turn deserves another"*

$$\left[\begin{array}{l} \text{Z ACT}_1 \\ \text{A-VAL}\,(\beta, \text{Y}) = + \end{array}\right]^{\beta} \Longrightarrow \text{MD-VAL}\,\left(\left[\begin{array}{l} \text{YA ACT}_2 \\ \text{RECIP}\,(\beta) \end{array}\right], \text{TO Z}\right) = +$$

b. *"Someone who does bad things to others deserves to be punished"*

$$\left[\begin{array}{l} \text{Z ACT}_1 \\ \text{A-VAL}\,(\beta, \text{Y}) = - \end{array}\right]^{\beta} \Longrightarrow \text{MD-VAL}\,\left(\left[\begin{array}{l} \text{YA ACT}_2 \\ \text{RECIP}\,(\beta) \end{array}\right], \text{TO Z}\right) = +$$

c. *"Someone who has bad things done to him by others deserves to be compensated"*

$$\left[\begin{array}{l} \text{Y ACT}_1 \\ \text{A-VAL}\,(\beta, \text{Z}) = - \end{array}\right]^{\beta} \Longrightarrow \text{MD-VAL}\,\left(\left[\begin{array}{l} \text{YA ACT}_2 \\ \text{REST}\,(\beta) \end{array}\right], \text{TO Z}\right) = +$$

In the right-hand side of (86), the reciprocal act (ACT_2) is no longer performed by Y, the other character. Rather it is performed by the generic actor *YA*. But since a generic actor can't owe Z anything, the function *MD-VAL* is reduced to the more objective form with only two arguments.

We can go still further. Consider first (86c). The fact that it is specifically Y that is doing something bad to Z is now actually irrelevant, since Y plays no role in the moral restitution. So (86c) can be generalized to (87), where no other characters need be involved.

(87) *"Someone who has bad things happen to him deserves to be compensated"*

$$\begin{bmatrix}\text{EVENT}\\ \text{A-VAL}\ (\beta, Z) = -\end{bmatrix}^{\beta} \Longrightarrow \text{MD-VAL}\ (\begin{bmatrix}\text{YA}\ \text{ACT}_2\\ \text{REST}\ (\beta)\end{bmatrix}, \text{TO Z}) = +$$

Counterparts to (86a,b) take one extra step. Recall rule (22):

(88) *"It's good to be nice to people"* (= 22)

$$\begin{bmatrix}\text{ACT}\ (Z)\\ \text{A-VAL}\ (\beta, \text{to Y}) = \alpha\end{bmatrix}^{\beta} \Rightarrow_{\text{default}} \text{N-VAL}\ (\beta, Z) = \alpha$$

Applying (88) to the left-hand side of (86a), we get "If Z does something good for Y, it is good of Z to do it"; applied to the left-hand side of (86b), we get "If Z does something bad to Y, it is bad of Z to do it". These inferences eliminate Y from the equation. Making a rather shady move, we use these inferences to replace the left-hand side of (86a,b). This yields (89a,b).

(89) a. *"Someone who does (N-)good things deserves to be rewarded"*

$$\begin{bmatrix}Z\ \text{ACT}_1\\ \text{N-VAL}\ (\beta, Z) = +\end{bmatrix}^{\beta} \Longrightarrow \text{MD-VAL}\ (\begin{bmatrix}\text{YA}\ \text{ACT}_2\\ \text{RECIP}\ (\beta)\end{bmatrix}, \text{TO Z}) = +$$

b. *"Someone who does (N-)bad things deserves to be punished"*

$$\begin{bmatrix}Z\ \text{ACT}_1\\ \text{N-VAL}\ (\beta, Z) = -\end{bmatrix}^{\beta} \Longrightarrow \text{MD-VAL}\ (\begin{bmatrix}\text{YA}\ \text{ACT}_2\\ \text{RECIP}\ (\beta)\end{bmatrix}, \text{TO Z}) = +$$

One further shifty step takes us to (90).

(90) a. *"Good people deserve to be rewarded"*

$$\text{N-VAL}\ (Z) = + \Longrightarrow \text{MD-VAL}\ (\begin{bmatrix}\text{YA}\ \text{ACT}_2\\ \text{RECIP}\ (\beta)\end{bmatrix}, \text{TO Z}) = +$$

b. *"Bad people deserve to be punished"*

$$\text{N-VAL}\ (Z) = - \Longrightarrow \text{MD-VAL}\ (\begin{bmatrix}\text{YA}\ \text{ACT}_2\\ \text{RECIP}\ (\beta)\end{bmatrix}, \text{TO Z}) = +$$

Of course (87) and (90) are massively counterexemplified in the world: bad things happen to people all the time with no hope of compensation, wicked people frequently do very well indeed, and all too often "Nice guys finish last." Here is the existential "problem of evil." How is it to be resolved? Different traditions have different ways. Christianity's solution is to put off reward and punishment until the afterlife, conveniently linking up with the strongly held belief in the survival of the soul after death (see chapter 4). Judaism tends to take the view that if something bad is happening to me now, it must be punishment for something bad I (or even my ancestors) did in the past. Hence the comedians' version of Jewish guilt: I must have done

something wrong, and I'm sorry – but what was it?[21] Yet another solution is "virtue is its own reward," which gives up on reward coming from *outside*, and so in a way negates the spirit that leads to (90).

But who is going to carry out the acts of reward and punishment? The anticipated reciprocal acts can't depend on people, since they are intended precisely as the way of circumventing people's injustice in the real world. Enter gods: animate moral beings who lie outside the human sphere and who take care of righting the moral scales. This puts gods in the role of protectors, beings whom one can plead for justice and to whom one can express gratitude. Moreover, the rules of normative value dictate that one had better be nice to the gods as well, because if anyone is in a position to reward or retaliate, it's the gods. Thus the reasoning in this section leads to one of the important groundings for religion, one that is not to my knowledge explored in recent work such as Boyer 2001 (but appears in Freud and perhaps Nietzsche).

Now it's not as though the steps leading to (87)-(90) follow from any sort of formal reasoning. But intuitively they're entirely seductive. I leave this as a last example of the slipperiness of intuitions regarding value.

8. What does this all mean?

I can imagine the reader who has suffered through all this discussion (or picked up a novel instead) asking what the point is. Why translate lots of moral truisms into a relatively arcane formal system? Having gone through the exercise, I can offer three reasons.

The first reason is that it reveals the complexity of our value judgments and our reasoning about values. By attempting to express these judgments in systematic terms, with a relatively limited vocabulary, we can see the rich interrelationships among different notions of value. An important virtue of the present approach is that it acknowledges different kinds of value, each participating in the system in a different way. My impression is that previous approaches have been limited because they insist on a unitary notion of value, and because in many cases (especially in moral philosophy), they discount subjective value altogether (see discussion in chapter 4).

Within the present approach, value is revealed as a complex conceptual system, rich in hierarchical abstract structure. In the context of how other cognitive systems are now understood – especially language – this should not be too surprising. This outcome is not undermined by the fact that value judgments are often quick and intuitive. In language, judgments of grammaticality and meaningfulness are quick and intuitive: the computational reasons for these judgments are deeply unconscious. In vision, judgments of spatial configuration and motion are intuitive and

---

[21]This also explains why many Jews lost faith during the Holocaust: nothing they or their ancestors had done could be bad enough to justify *this*.

present themselves to awareness as "what is the case in the world." Thus value judgments are of a piece with the rest of cognition. However, unlike linguistic and visual judgements, aspects of them are available to awareness as well – they lie on the borderline between intuitive and conscious reasoning.

A second reason that these explorations are valuable has to do with the connection between the value system and its linguistic expression. Part of the job of linguistic semantics is to explicate the meanings of words and phrases. In the domain of spatial language, linguistic semantics has benefitted from the attempt to develop formal analyses of the conceptualization of space, motion, force, and agency in terms of a limited conceptual vocabulary and combinatorial system, and large portions of language have been subsumed under such analyses. The present exploration has begun a similar undertaking with the parts of the vocabulary whose meanings incorporate notions of value. The appendix lists words that have come up in the course of discussion here, as a demonstration of how wide-ranging the investigation is.

A third reason for taking this approach seriously is that it offers the possibility of making many longstanding questions more precise. At the scale of the individual lifespan: What is the course of development of value systems in humans (a la Piaget 1932, Kohlberg 1981/84, Turiel 1983, Macnamara 1990, Premack and Premack 1994, Bloom 2004)? Over historical time: To what extent are value systems a functional outcome of what it takes to make a society work well (and interact well with other societies) (Fiske 1991, Jacobs 1994)? Over evolutionary time (Hauser 2000, deWaal 1996): How much of the basis of human value systems is innate? Of that, how much is part of our primate heritage, and how much is unique to humans? By taking a formal overview of the entire system, it's possible to pose these questions in a more comprehensive context.

I don't want to pretend that the analysis here is an ultimate solution. Its complexity has a certain degree of arbitrariness that reminds me of early transformational grammar (Chomsky 1957). Why should these particular principles be ones that we find persuasive, and why are other conceivable principles that we can state within this formal system not felt to be valid? Taking my cue from the history of generative grammar, I might answer that we can approach such questions of explanation only when we have a formal system for accurately describing intuitions about value. For the moment, I am pleased with how much descriptive breadth and depth has been possible here in a relatively short chapter, and how many potential connections to other disciplines, while at the same time recognizing that this is only a first step.

9. Appendix: Words discussed

Section 2.1
good/bad for
beneficial/harmful

Section 2.2
good to have
valuable
worth

Section 2.3

terrible (quality)
excellent (quality)
miserable (quality)
good for
makes a good X

Section 2.4
good of
(Section 2.4)
bad of
should (prudential and normative)
[conscience]
[guilt]
illicit

Section 2.5
good at
excellent at
prowess

Section 2.6
esteem
respect
reputation
prestigious

Section 3
nice
notorious

Section 4
(in return)for
reciprocation
retaliation
retribution
restitution
(to make up) for
revenge
turn the other cheek
honor
prize
thank
humiliate
apologize

Section 5.1
exchange

Section 5.2
trade
buy
sell

Section 5.3
pay
hire
hire oneself out
bribe
price
expensive
cheap
owe
for (beneficiary)
on (adversative)

Section 6
fair
equality under the law
selfish
altruistic
vote

Section 7.1
wages

Section 7.2
toadying/brown-nosing

Section 7.4
deserve

References

Akerlof, George A., and Janet L. Yellen. 1993. The Fair Wage-Effort Hypothesis in Unemployment. In Hechter et al. 1993, 107-134.

Aronoff, Mark. 1980. Contextuals. *Language* 56: 744-758.

Barth, Fredrik. 1993. Are Values Real? The Enigma of Naturalism in the Anthropological Imputation of Values. In Hechter et al. 1993, 31- 46.

Berger, Peter L. and Thomas Luckmann. 1966. *The Social Construction of Reality*. Garden City, NY: Doubleday.

Bloom, Paul. 2004. *Descartes' Baby: How the Science of Child Development Explains What Makes Us Human*. New York: Basic Books.

Boyer, Pascal. 2001. *Religion Explained.* New York: Basic Books.

Chomsky, Noam. 1957. *Syntactic Structures*. The Hague: Mouton.

Cosmides, Leda. 1989. The logic of social exchange: Has natural selection shaped how humans reason? Studies with the Wason selection task. *Cognition* 31, 187-276.

Dehaene, Stanislas. 1997. *The Number Sense: How the Mind Creates Mathematics*. Oxford: Oxford University Press.

de Waal, Frans. 1996. *Good Natured: The Origins of Right and Wrong in Humans and Other Animals*. Cambridge, MA: Harvard University Press.

Fehr, Ernst, and Urs Fischbacher. 2004. Social norms and human cooperation. *Trends in Cognitive Sciences* 8.4.

Fillmore, Charles, and Beryl Atkins. 1992. Toward a Frame-Based Lexicon: The Semantics of RISK and its Neighbors. In Adrienne Lehrer and Eva Kittay (eds.), *Frames, Fields, and Contrasts*, 75-102. Hillsdale, NJ: Erlbaum.

Fiske, Alan. 1991. *Structures of Social Life*. New York: Free Press.

Greene, Joshua. 2003. From neural 'is' to moral 'ought': What are the moral implications of neuroscientific moral psychology? *Nature Reviews/Neuroscience* 4, 847-850.

Gruber, Jeffrey. 1965. *Studies in Lexical Relations*. Ph.D. dissertation, MIT. Published as part of *Lexical Structures in Syntax and Semantics*. Amsterdam: North-Holland, 1976.

Harman, Gilbert. 2000. *Explaining Value*. Oxford: Oxford University Press.

Hauser, Marc. 2000. *Wild Minds: What Animals Really Think.* New York: Henry Holt.

Hechter, Michael, Lynn Nadel, and Richard E. Michod (eds.). 1993. *The Origin of Values.* New York: Aldine de Gruyter.

Herrnstein, Richard J. 1993. Behavior, Reinforcement, and Utility. In Hechter et al. 1993, 137-152.

Jackendoff, Ray. *Semantic Structures.* Cambridge, MA: MIT Press.

Jacobs, Jane. 1994. *Strategies of Survival.* New York: Vintage Books.

Katz, Jerrold. 1966. *The Philosophy of Language.* New York: Harper & Row.

Kohlberg, Lawrence. 1981/84. *The Philosophy of Moral Development* (vols. I and II). New York: Harper & Row.

Macnamara, John. 1990. The development of moral reasoning and the foundations of geometry. *Journal for the Theory of Social Behaviour* 21, 125-150.

Mandler, George. 1993. Approaches to a Psychology of Value. In Hechter et al. 1993, 229-258.

Millikan, Ruth. 1984. *Language and Other Abstract Objects.* Cambridge, MA: MIT Press.

Piaget, Jean. 1932. *The Moral Judgment of the Child* (trans. M. Gabain). New York: Free Press.

Premack, David, and Ann James Premack. 1994. Moral belief: Form versus content. In Lawrence A. Hirschfeld and Susan A. Gelman (eds.), *Mapping the Mind: Domain-Specificity in Cognition and Culture*, 149-168. New York: Cambridge University Press.

Pustejovsky, James. 1995. *The Generative Lexicon.* Cambridge, MA: MIT Press.

Schwartz, Barry. 1993. On the Creation and Destruction of Value. In Hechter et al. 1993, 153-186.

Scitovsky, Tibor. 1993. The Meaning, Nature, and Sources of Value in Economics. In Hechter et al. 1993, 93-105.

Searle, John. 1995. *The Construction of Social Reality.* New York: Free Press.

Stevens, Jeffrey R. and Marc D. Hauser. 2004. Why be nice? Psychological constraints on the evolution of cooperation. *Trends in Cognitive Sciences* 8.2.

Stich, Stephen P. 1993. Moral Philosophy and Mental Representation. In Hechter et al. 1993, 215-228.

Turiel, Elliott. 1983. *The Development of Social Knowledge: Morality and Convention.* Cambridge: Cambridge University Press.