The Baltic International Yearbook of
Cognition, Logic and Communication

October 2010 Volume 5: *Meaning, Understanding and Knowledge*
pages 1-25 DOI: 10.4148/biyclc.v5i0.285

Peter Ludlow
Northwestern University
peterjludlow@gmail.com

Having Linguistic Rules and Knowing Linguistic Facts

Abstract: 'Knowledge' doesn't correctly describe our relation to linguistic rules. It is too thick a notion (for example, we don't believe linguistic rules). On the other hand, 'cognize', without further elaboration, is too thin a notion, which is to say that it is too thin to play a role in a competence theory. One advantage of the term 'knowledge'—and presumably Chomsky's original motivation for using it—is that knowledge would play the right kind of role in a competence theory: Our competence would consist in a body of knowledge which we have and which we may or may not act upon—our performance need not conform to the linguistic rules that we know.

Is there a way out of the dilemma? I'm going to make the case that the best way to talk about grammatical rules is simply to say that we have them. That doesn't sound very deep, I know, but saying that we have individual rules leaves room for individual norm guidance in a way that 'cognize' does not. Saying we have a rule like subjacency is also thicker than merely saying we cognize it. Saying I have such a rule invites the interpretation that it is a rule for me—that I am normatively guided by it. The competence theory thus becomes a theory of the rules that we have. Whether we follow those rules is another matter entirely.

The first time I saw Noam Chomsky give a talk, it was on the topic of knowledge of grammar. At the time (around 1980) I could not understand why he was using the term 'knowledge' in describing our relation to linguistic rules. Like most philosophers, I assumed that, to a first approximation, knowledge was justified true belief. The linguistic rules that Chomsky was talking about were not the sorts of things that people *believed* even upon reflection (e.g. the rule *subjacency*, which says that a move with an element like the word 'what' can't jump over an S node and an NP node without an intervening landing site), nor was it clear that linguistic rules were the sorts of things that were true or correct, since there wasn't a question of having the wrong linguistic rules. You just have the rules that you do. Finally, talk of justification didn't make sense for grammatical rules; there is no issue of my getting the right rule in the wrong way (say from an unreliable teacher or peer). So linguistic rules weren't believed, they weren't true in any sense, and even if they were true and we did believe them there is no reason to think those beliefs would be justified. That is 0 for 3.

Things are different for a prescriptive grammarian, for in that case there **is** a question of knowing the prescriptive rules of your language. Maybe some sort of official academy of language or someone in a power relation with respect to you determines what those rules are. In that case I might come to believe a rule like 'never split an infinitive', and (according to the prescriptivist) there could be a question of whether it is a "correct" rule for English, and there is even a question about whether I am justified in believing the rule (did I get it from a reliable source?). But for a generative linguist this picture is deeply confused. Generative linguists are engaged in an enterprise that is both descriptive and explanatory. It is descriptive in that they are interested in the linguistic rules that individual people actually have, and it is explanatory in that they are interested in why those people have the rules that they do. The explanation typically involves an innate language acquisition device that admits of parametric variation. For example, one common metaphor is that the language faculty is a largely prewired box with a finite number of discrete switches (parametric settings). When you are exposed to a language as a child the switches are set and you end up in a particular parametric state of the language faculty. Whatever we chose to call our relation to the resulting body of

linguistic rules, 'knowledge' is not a happy term.

There are of course exceptional cases where the idea of knowledge of rules makes perfectly good sense. There are "surfacey" rules like lexical rules (e.g. that 'snow' refers to snow) which are accessible to us and which are the kinds of things that could be got right or wrong since I might be prepared to defer to my community on what the correct rule is, but these are not the kinds of rules that linguists typically concern themselves with.

Of course we can talk about trained linguists knowing rules like subjacency, but this isn't the kind of knowledge that typical language users have; it is a kind of theoretical knowledge that is only available to those acquainted with professional linguistics. It is scientific knowledge.

About the time Chomsky gave the talk I heard, he also (Chomsky (1980); 69-70) introduced the term 'cognize'—suggesting that it was a kind of technical precisification of the term 'knowledge' and that we might also say that we "cognize grammars". While 'cognize' may have started out as a sharpening of 'knowledge' its use has drifted in the linguistics literature to the point where it simply means that we mentally represent grammars. This weakening seems to retire the worries with using 'knowledge' (i.e. that rules of grammar aren't true and we don't believe them), but it raises questions of its own. Suitably watered down, 'cognize' suggests that linguistic rules are like data structures in the computational system that is the language faculty. (For example, we could think of the rules as being like lines of code that are accessed by a natural language processing system.)

This makes sense for an account of linguistics that takes linguistics to be a *performance* theory, but it seems inadequate if we take linguistics to be (as Chomsky still does) a *competence* theory. The problem with the term 'cognizing' once it is watered down to mean 'represents' is that it is too thin. A competence theory suggests that linguistic rules are more than just data structures involved in our computations.

We have the makings of a dilemma here. On the one hand, 'knowledge' just doesn't correctly describe our relation to linguistic rules. It is too thick a notion. On the other hand, 'cognize', without further elaboration, is too thin a notion, which is to say that it is too thin to play a role in a competence theory. One advantage of the term

'knowledge'—and presumably Chomsky's original motivation for using it—is that knowledge would play the right kind of role in a competence theory: Our competence would consist in a body of knowledge which we have and which we may or may not act upon—our performance need not conform to the linguistic rules that we know.

Is there a way out of the dilemma? I'm going to make the case that the best way to talk about grammatical rules is simply to say that we *have* them. That doesn't sound very deep, I know, but saying that we have individual rules leaves room for individual norm guidance in a way that 'cognize' does not. I'll say a bit more about the details of this (like what it means to have a linguistic rule), but for now I just want to be clear on how this avoids our dilemma. The problem with 'knows' was that it was too thick, and introduced features that are simply not appropriate for the rules that generative linguists are concerned with. We don't *believe* that we have rules like subjacency, nor is there some sense in which subjacency is the correct rule for us. But it is certainly appropriate to say that we *have* subjacency (or that my idiolect has the subjacency rule or some parametric variation of it).

Saying we have a rule like subjacency is also thicker than merely saying we cognize it (or at least it can be made thicker). Saying I have such a rule invites the interpretation that it is a rule *for me*—that I am normatively guided by it (we will explore this in detail in section 2 of the paper). The competence theory thus becomes a theory of the rules that we have. Whether we follow those rules is another matter entirely.

I've rejected talk of knowledge for linguistic rules, but I haven't rejected it for all aspects of linguistics; I think that there are kinds of linguistic knowledge that we have (even if tacit). I'll argue that there are certain kinds of linguistic facts or phenomena that we can have knowledge of, although the nature of this knowledge is going to be somewhat partial.

Accordingly, I have two jobs to complete in this paper. I need to makes sense of what it means to have a linguistic rule (and how it can be normatively guiding), and I need to makes sense of what it means to know linguistic facts and phenomena. I'll take up these topics in reverse order.

1. KNOWING LINGUISTIC FACTS

Let's say that Universal Grammar (UG) is the system that accounts for the different individual grammars that humans have. UG is thus not to be confused with the theory of grammar itself; rather UG is an object of study in the theory of grammar.

For example, if we think of UG as being the initial state of a parametric system of rules, then individual grammars are the result of parameters being set. I have the grammar that I do—call it G_{PL}—because of the way the parameters were set in response to the linguistic data I was exposed to. Let's also not confuse the grammar that I have with the resulting state of the parameter setting of UG in me. Let's call the resulting state of my parameters being set as UG_{PL}. We can say that I have the grammar G_{PL} because I am in parametric state UG_{PL}.

As a further preliminary, let's say that a *grammar* generates a language (or *language narrowly construed* in the sense of Hauser, Chomsky and Fitch (2002)). We can now make a distinction between the language narrowly construed that is generated by my grammar G_{PL}—we can call this language $L_{G_{PL}}$—and other phenomena that we might pretheoretically take to be linguistic, or part of my "language" understood loosely speaking. Let's call this pretheoretical collection of phenomena that involve my language L_{PL}. To illustrate the distinction, consider the contrast between the following two sentences involving center embedding.

(1) The cat the dog bit ran away
(2) The mouse the cat the dog bit chased ran away

We might hypothesize that although I judge (2) to be unacceptable, it is still well formed or legible in $L_{G_{PL}}$; perhaps I merely judge it to be unacceptable because of processing difficulties. So (2) is well formed according to $L_{G_{PL}}$ but not acceptable in L_{PL}.

Accordingly, $L_{G_{PL}}$ should not be expected to line up with all of the phenomena that we pre-theoretically take to be linguistic or part of my language. The range of phenomena in $L_{G_{PL}}$ are determined by theoretical investigation and they at best overlap with the range of phenomena in L_{PL}. Clearly there can be disagreement about the range of phenomena that fall under $L_{G_{PL}}$. Some linguists have pressed for a very broad understanding of what the theory of grammar might be expected to account for, and at times they have suggested that we expand the range of phenomena explained by G_{PL} to include most if not all of L_{PL}. I think it is fair to say that most generative linguists today take the range of phenomena explained by the theory of grammar in isolation to be limited. However, it also seems fair to say that the interaction of the theory of grammar with other considerations can contribute to the explanation of a broad range of phenomena - perhaps even most of the phenomena falling under L_{PL}. More generally, let's say then that L_{PL} is a function of $L_{G_{PL}}$ + processing considerations + pragmatic considerations + socio-cultural factors, etc.

I've already given a case (center embedding) where processing limitations interact with the grammar to explain a phenomenon that fell under L_{PL}. Other phenomena might be explained by the fact that I have been inculcated with prescriptive rules. For example, I might find a sentence like 'I ain't got no money' unacceptable. Is the explanation for this that it violates G_{PL} or is it because I was drilled by grammar school teachers not to use 'ain't' and "double negatives" (actually, this is not really a double negative, but the use of a negation as a negative polarity item). In advance of inquiry there is no way to know, because phenomena do not wear their explanations on their sleeves.

I have been using the terms 'phenomena' and 'facts' interchangeably. For the moment let's stick with the term 'facts' and make a distinction between two kinds of facts (or at least two ways of individuating facts). Let's say that there are surfacey facts (S-facts) and explanatory facts (X-facts) about L_{PL}. S-facts are facts like this: 'Who did you hear the story that Bill hit' is not acceptable. X-facts incorporate information about the explanations for these surfacey linguistic facts—for example this: 'Who did you hear the story that Bill hit' is unacceptable because it violates subjacency. I will also be making the case that the source of our knowledge about S-facts includes our judgments (what are sometimes called linguistic intuitions).

Of course, our knowledge attribution reports are often forgiving of what we don't know, so that sometimes we might say that an agent recognizes a subjacency violation in a sentence, when the agent would merely report that the sentence is not right and have no idea what subjacency is. I don't have an issue with this sort of knowledge report, but to be clear we should call this a *charitable attribution* of knowledge

of an X-fact.

To understand the role of linguistic judgments in this picture it will be useful if we can get clear on the difference between linguistic *theory*, linguistic *phenomena or facts*, and linguistic *data*. Following more general work in the philosophy of science by Bogen & Woodward (1988) we can illustrate the relation between theory, phenomena, and data as follows.

Theory $\Longrightarrow_{\text{explains/predicts}} \Longrightarrow$ **phenomena** $\Longleftarrow_{\text{is evidence for}} \Longleftarrow$ **data**
$\Longleftarrow_{\text{is evidence for}} \Longleftarrow$

Theory is the theory of grammar in this case. Following Bogen and Woodward, I will take *phenomena* to be stable and replicable effects or processes that are potential objects of explanation and prediction for scientific theories. In this case the phenomena will include the pre-theoretical domain of language-related facts. While pre-theoretically we can't say which facts *provide evidence* for the theory of grammar (the theory of UG), some facts will provide such evidence (more or less directly). We will also say that the theory *contributes to* the explanation and prediction of these facts. We can also say that I have knowledge of some of the linguistic phenomena.

I will take *data* to be observational evidence for claims about phenomena. The data come from token events of observation and experimentation. For example, an act of measuring the freezing temperature of a liquid might yield the datum that the fluid froze at n degrees (this is not to be confused with a written record of the measurement - we can call this a *record of the datum*). This datum is a piece of evidence for the more general phenomenon (fact) that the fluid freezes at n degrees. The data are token-based, and the phenomena are type-based.[1] We can, of course, aggregate data. So, for example, we might aggregate the results of several observations to show that the average freezing temperature in our experiments is n degrees. It still counts as data on my view because we are aggregating over token experiments/observations.

This distinction between data, records of data, and sources of data applies also to linguistic methods that appeal to corpora of written sentences. In this case an occurrence of a sentence in a written corpus is not the datum. Rather the datum is *that* the sentence was found in the

corpus on a particular search. This datum provides evidence for several kinds of phenomena. In the first place it provides evidence for the phenomenon *that* the sentence occurs in the corpus. We can aggregate this data (either by counting occurrences or statistically generalizing over data) to show, for example, that it occurs 1000 times or a certain percentage of the time. Again, this is still a datum in my usage because it is an aggregation over token observations. It may also, via inference, provide evidence for the phenomenon that the sentence is acceptable to a number of language users.

As with all data, linguistic data come from observation and experimentation (for example, they may be found in a corpora or they may be the result of acts of judging that tokenings of linguistic forms are unacceptable; as we will see there are many other potential sources of data). Such data provide *evidence for* phenomena (both surfacey and explanatory linguistic facts) that are in turn *explained by* the theory of grammar.

Let's make this a bit more concrete with some specific examples from linguistics.

Consider subjacency, and the case where an act of judgment by me is the source of a datum. As noted earlier, we do not have judgments about rules like subjacency, nor do we have judgments that a particular linguistic form violates subjacency. Rather, our judgments of acceptability provide evidence for the existence of these phenomena. We can illustrate the idea as follows:

Grammatical Rule for G_{PL}
Subjacency: Moved elements can't jump an NP and an S node without an intervening landing site
Explanatory fact about $L_{G_{PL}}$ (potential object of theoretical knowledge for PL)
'$[_S$ who$_i$ did you hear the story that Bill hit ei]' violates subjacency
Explanatory fact about L_{PL} (potential object of theoretical knowledge for PL)
'who did you hear the story that Bill hit' is unacceptable in L_{PL} because it violates subjacency
Surfacey fact about L_{PL} (potential object of knowledge for PL)
'who did you hear the story that Bill hit' is unacceptable for PL
Datum (content judged by PL)
That a particular tokening of 'who did you hear the story that Bill hit?' is

unacceptable to PL
Source of datum (act of judgment by PL)
PL's act of judging that 'who did you hear the story that Bill hit?' is unacceptable

The next illustration involves the case of reflexives. We again distinguish between the linguistic phenomenon, which involves complex notions from binding theory and the data (in this case linguistic judgment), which are much more "surfacey".

Grammatical Rule for G_{PL}
reflexives must be bound in their governing category
Explanatory fact about $L_{G_{PL}}$ (potential object of theoretical knowledge for PL)
'$[_{NP}$ himself]' can be bound by '$[_{NP}$ John]' in '$[_{S}$ Bill said that John likes himself]' because it is in the same governing category, and '$[_{NP}$ Bill]' cannot bind '$[_{NP}$ himself]' because it is not in the same governing category.
Explanatory fact about L_{PL} (potential object of theoretical knowledge for PL)
'himself' can be bound by 'John' in 'Bill said that John likes himself' because it is the same governing category, and 'Bill' cannot bind 'himself' because it is not in the same governing category.
Surfacey fact about L_{PL} (potential object of knowledge for PL)
'himself' can be associated with 'John' but not 'Bill' in 'Bill said that John likes himself'
Datum (content judged by PL)
Judgment that 'himself' can be associated with 'John' but not 'Bill' in a given tokening of 'Bill said that John likes himself'
Source of datum (act of judgment by PL)
PL's act of judging that 'himself' can be associated with 'John' but not 'Bill' in a given tokening of 'Bill said that John likes himself'

Summarizing thus far, grammatical rules generate linguistic facts—facts about our language. We have knowledge about some of those facts (mostly the surfacey facts), and for the most part our knowledge is underwritten by the judgments that we make about those surfacey facts. We can get things wrong, of course, but this doesn't undermine our knowledge. Indeed, you could make the case that the possibility of getting it wrong is a prerequisite for something being a candidate

object of knowledge.

But this leaves open the question of what role this knowledge plays for us and how it interacts with the individual grammatical rules that each of us has (e.g., in my case the rules specified by G_{PL}). I'm suggesting that it is our knowledge of these facts that can tell us if we are following the rules of our individual grammars. That is, we use our knowledge about linguistic facts as cues to let us know if we are in accord with our individual linguistic norms.

2. HAVING RULES

As noted earlier, linguists and philosophers of linguistics typically think of linguistic rules and principles as being descriptive; they don't often think of linguistic rules as being normative, but that doesn't mean the idea is a complete nonstarter. One problem, of course, is that the rules (like subjacency) are so abstract that few would be in a position to consciously entertain them, much less reflect on their normative pull.

This, in effect, is why thinking about normative linguistic rules is not very attractive to linguists. It certainly doesn't make sense to think of us following these rules based on any sort of reflective capacity. It is much easier to think of linguistic principles as being part of a project of *describing* linguistic competence, rather than *normatively guiding* linguistic competence.

But perhaps we can still make sense of the idea. In part 1 we looked at linguistic judgments and the role they play in our linguistic knowledge. Now the question is whether we can also think of linguistic judgments as playing a role in *directing* or *monitoring* our linguistic competence.

As also noted earlier, this makes perfect sense for prescriptive grammarians. Certain rules are ingrained in you (for example: "don't end a sentence with a preposition"), you come to have judgments that comport with those rules, and they guide your linguistic practice. Now from the perspective of generative linguistics this idea is so confused it would be a task to even sort out and enumerate all the mistakes. Clearly, we aren't interested in artificial prescriptive rules, and when we get to rules like subjacency, it seems fairly implausible to think that we have transparent judgments about such things, much less that judg-

ments of that form could provide normative guidance. So how can we be normatively guided by our grammar when it is construed as an abstract object that is the product of a parametric state of the language faculty?

Consider the following passage from John Lawlor's online course notes, where he discusses the role that Ross island constraints might play in our linguistic planning and performance:

> Violations of Ross Constraints are very ungrammatical. Most people never encounter them. We appear to formulate our discourse to avoid them. Occasionally, we get in a bind and see one looming at the end of the clause, and have to do something quick. What we do is often illuminating about the relative importance of syntactic rules.
>
> For instance, consider the following:
>
> **?That's the book$_i$ [that Bill married the woman$_j$ [who$_j$ illustrated it$_i$]].**
>
> ***That's the book$_i$ [that Bill married the woman$_j$ [who$_j$ illustrated ____$_i$]].**
>
> Neither sentence is terrifically grammatical, but the first seems more appropriate (and common as a type) than the second, though the last word in the first sentence still feels strange. The ordinary rule of relative clause formation operating on the last clause should result in its deletion at the end of the clause (and thus the sentence). However, it appears inside another relative, an island, and is thus safe from such "movement" by the Complex NP Constraint.
>
> Sentences like the first one are generated when, at the last minute, the speaker realizes what is going to result, and cancels the deletion, substituting an alternative relative-formation rule (called a ***Resumptive Pronoun*** in the trade), which merely pronominalizes the coreferential NP, instead of deleting it in the object position.
>
> This is not the way English forms its relative clauses (though other languages use it frequently, e.g, Hebrew), and the sentence is thus ungrammatical. But this turns out to be a venial syntactic sin by comparison with a violation of a Ross constraint, which typically produces extreme ungrammaticality.

Notice Lawlor's description of the case. He is not saying that we are consciously aware of Ross constraints (although to some extent

his description is charitable in the sense discussed above), nor even that we judge that there has been compliance with such rules. Rather he is saying that we see that something is wrong, and that we act so as to make it right. We make it as right as we can, and in so doing (and strictly speaking unbeknownst to all but theoretical linguists) we have acted so as to avoid a violation of the Complex NP constraint (in effect, a general rule that encompasses subjacency). What I am suggesting is that this is a case in which we are normatively guided by the Ross constraint even though we have no conscious knowledge of such guidance.

This gambit is tricky to play. If it is true that we have judgments that play a role in the normative guidance of our linguistic performance, we don't want to be in a position that only true believers can be so guided. These should be judgments that not only are *available* to all competent linguistic agents, but in fact are *used* by them as ways of checking or regulating their linguistic performance. *All* linguistic agents. Even agents that believe their judgments are not connected to their linguistic rules but are actually judgments about language construed as an external social object, even agents that believe the judgments are generated by high level central processing mechanisms, even agents that don't believe they have linguistic rules of any form.

Given all this, the idea of linguistic rules providing normative guidance must be pretty hopeless, no? I'm going to suggest that the problem is difficult but not necessarily hopeless. The puzzle is to figure out how we can have judgments that can tell us if certain linguistic forms are well-formed according to a linguistic rule system if those very judgments are not sufficient to tell us that they are *about* compliance with the linguistic rule system. Can we have a judgment about the output of rules of type *t*, use that judgment to regulate our behavior, and not even recognize that the judgment involves compliance with rules of type *t*? This might *seem* hopeless, but it is really a quite widespread phenomenon, or at least closely related to a widespread phenomenon.

An example comes from work in ethics by Arpaly (2003) and Railton (2006). They discuss cases where an agent is following an ethical principle, but does not recognize this, and indeed even interprets their behavior as ethically unprincipled and indeed morally wrong. Arpaly illustrates this idea with a literary example from Mark Twain's book

Huck Finn. Huck, who decides not to turn in the escaped slave Jim, even though he thinks the *moral* option is to do precisely that. Huck believes that Jim is someone else's property, after all. He judges that turning in Jim is not the thing to do, but he takes the judgment to have a non-moral etiology. Why does Huck refuse to turn in Jim? Well, Huck is really not able to articulate the reason. As Railton (2006) has described such situations, perhaps Huck just has a nagging feeling of discomfort at the idea of doing it. He doesn't feel that this discomfort or his decision of what to do is based on an ethical principle, but Arpaly and Railton argue that this is precisely what Huck is doing - he is following a moral principle, but he describes his action as being immoral. Indeed he thinks he is a bad boy for that very reason.

Now let's return to the case of judgments of acceptability for linguistic forms. We can imagine someone in the position of Huck, only with respect to linguistic rules rather than moral rules. Let's call this hypothetical agent 'Michael'. Michael has a grammar as part of his cognitive architecture, and he has judgments of linguistic acceptability that he uses to guide his linguistic performance. Yet Michael, like Huck, is deeply confused. Although his judgments guide him in such a way that he generally follows the rules/principles of grammar he does not recognize that he is so guided. He doesn't believe that there is a grammar construed as a chapter of his cognitive psychology - to the contrary he takes language to be a social object, and he thinks that he isn't following rules at all, but rather thinks he has a kind of knowledge-how that accounts for his linguistic competence. He also misdescribes his judgments, taking them as being the byproducts of high level processing over social linguistic facts.

Michael, like Huck who believes he is immoral, believes he is ignorant of language in the relevant sense. But on Arpaly and Railton's view we don't need to feel bad for either Huck or Michael. Huck really is acting on ethical principles; he is not really a bad boy after all. Similarly, Michael really does have robust knowledge of his language and he really is following grammatical rules—in spite of what even Michael himself insists, he is not ignorant of language after all.

As I said before, this gambit is subtle. One needs to develop a position on the normativity of language which can allow that linguistic rules are very abstract and currently outside the reach of our best linguistic theorizing yet have some normative pull on us.

It's interesting to note that Railton (2006) uses a linguistic case to illustrate the kind of rule governance he is attempting to get clear on.

> Interestingly, such pressures for consistency can be triggered and felt even when the norm of the agent in question is one of which she herself is unaware. One intriguing piece of evidence for this is the phenomenon of overregularization in children's speech. As their linguistic ability develops, some children who have previously mastered the past tenses of irregular verbs begin 'correcting themselves' by forming irregular past tenses using the <verb stem + -ed> rule for regular verbs, for example saying 'go-ed' instead of 'went'. This occurs despite the fact that these children have never heard 'go-ed' spoken by adult speakers, and have never been sanctioned for using 'went' as the past tense of 'go'. As adults, we feel similar pressures toward consistency in language use. We can sense that grammatical anomaly is creeping into a sentence we are uttering, and struggle to correct ourselves on the fly. We treat such anomalies as mistakes, even when they have no effect on—or even improve—sentence intelligibility, and even when we would be at a loss to identify the particular incompatibility with grammatical rules involved. (2006; p. 12).

But precisely what notion of rule governance works in cases like linguistics and the Huck Finn case we discussed earlier? Railton (2006) walks us through a series of accounts of rule governance until he gets one that he thinks fits the bill. He begins with the following formulation.

> **(RG1)** Conduct C is guided by norm N only if C is in accord with N

The problem with this formulation, as we all know too well, is that conduct can be rule guided but fall short of successfully being in accord with N. That is, I might be guided by a rule, try very hard to follow the rule, and yet fall well short of my goal. Rule guidance may fall well short of rule accord.

This leads us to Railton's second formulation.

> **(RG2)** Conduct C is guided by Norm N only if C is the manifestation of a reliable disposition to act in a way conducive to compliance with N.

This is better, but Railton suggests a problem. Consider the case of Harry the receptionist. He is disposed to dress in accord with company dress code, but is guided by his own sense of style rather than by the company dress code, of which he is only vaguely aware. In this case we would say that Harry is disposed to act in a way that is conducive to compliance with N, but it does not seem that he is *guided* by the company dress code. We need to make it clear that one's disposition to be in compliance with N is somehow connected with one's being guided by N itself.

This leads us to the third formulation.

> **(RG3)** Conduct C is guided by norm N only if C is the manifestation of a disposition to act in a way conducive to compliance with N, such that the fact that C conduces to compliance with N plays an appropriate role *in the explanation* of the agent's C-ing.

Obviously everything here turns on what would count as being explanatory in the agent's C-ing. Railton provides the following formulation: "A has a mental representation of N, judges that C-ing would conduce to compliance with N, takes this to be a reason for C-ing, and this judgment (partially) causes A's C-ing [in] virtue of its content."

The problem with this formulation is that most cases of norm guidance don't have this explicit character—in either the ethical or the linguistic realm. As Railton puts it, "in many cases of norm-guided behavior, individuals do not even form the belief that their conduct conduces toward normative compliance."

To see this, consider the case of Fred, who is disposed to validate his ticket whenever he gets on the bus. When he forgets to do so, he corrects his behavior.

> "Fred does feel discomfort upon discovering that he is riding without validating. . . Fred tends to treat departures from his usual practice as calling for correction."

Notice however, that Fred may not be aware that he is motivated by some sort of rule. He may only be aware of the vague sense of discomfort that he feels. (As we have seen, linguistic cases and the Huck Finn case work like this too.) What we need is a notion of regulative behavior that allows that we can be guided by a rule without explicitly being aware of the content of the rule (under relevant descriptions). We could achieve this, if the rule functioned as a kind of regulator in the sense deployed by engineers – something that regulates the system for error and corrects when necessary. More explicitly, Railton offers the following explication of regulative explanation:

> **regulative explanation:** For an engineer, a regulator is a device with a distinctive functional character. One component continuously monitors the state of the system—the regulated system—relative to an externally set value, e.g. temperature, water pressure, or engine velocity. If the system departs from the set-point value, the monitor sends an 'error signal' to a second component, which modulates the inputs into the system. . . until the set-point value is restored.

This helps us get closer to the notion of regulative behavior, and suggests the following formulation for rule governance:

> **(RG4)** Agent A's conduct C is guided by the norm N only if C is a manifestation of A's disposition to act in a way conducive to compliance with N, such that N plays a regulative role in A's C-ing, where this involves some disposition on A's part to notice failures to comply with N, and to feel discomfort when this occurs, and to exert an effort to establish conformity with N

The problem with this formulation is that regulative behavior isn't quite enough. Not all regulated behavior counts as rule governed behavior. To illustrate this point, Railton offers another disposition of Fred's.

Fred usually has a snack around 10 AM every day. Typically, when he sees the clock strike 10, he goes and buys a snack. One day, however, buried under a pile of work, Fred works straight through to lunch. As Railton describes the situation, "Fred does not regard this failure as

something that calls for correction. Instead, he thinks only, 'Funny, I didn't even notice'."

Railton calls this a non-consequential and unsanctioned failure to fit his standing behavioral expectations. The failure is non-consequential because, as it happened, Fred suffered no ill effects from the omission. The failure is unsanctioned (it might be better to say it is "not sanctioned") because no authority would take any interest in his missed snack or impose penalties.

In Railton's terminology, the morning snack is a *default plan*, and it plays a role in regulating Fred's behavior, but it isn't normative in the way that his ticket validating was. The clue to its not being normative is that Fred shrugs off the missed snack, whereas if, for example, he realized that he failed to validate his bus ticket earlier in the day his inner dialogue might invoke excuse-making and bargaining ("I've lost tickets before that I never used", or "I'll validate twice text time") even though his failure to comply with the validating rule has no real consequences when he realizes his omission.

Are linguistic cases like this too? We don't always exert effort to comply with our judgments of acceptability, but often enough we do, even when compliance is not required, and even when the editing gets in the way of smoothly communicating.

We need more than (RG4). Genuine rule governance is not just about following a default plan, it involves a felt need to fix things even when we are not under external pressure to do so. Thus Railton ultimately suggests we opt for (RG5).

> **(RG5)** Agent A's conduct C is guided by the norm N only if C is a manifestation of A's disposition to act in a way conducive to compliance with N, such that N plays a regulative role in A's C-ing, where this involves some disposition on A's part to notice failures to comply with N, to feel discomfort when this occurs, and to exert effort to establish conformity with N *even when* the departure from N is unsanctioned and non-consequential.

Railton's formulation here is a bit misleading because it seems to suggest we might be in a position to consciously recognize we are complying with N, but this can't be what Railton intends, since he expressly

states that we may not have access to N. So let's try the following more austere formulation.

> **(RG5')** Agent A's conduct C is guided by the norm N only if C is a manifestation of A's disposition to act in a way conducive to compliance with N, such that N plays a regulative role in A's C-ing, where this involves some disposition on A's part to feel discomfort at failing to C, and to exert effort to C even when failing to C is unsanctioned and non-consequential.

Even on this austere formulation, there are a couple of assumptions built into Railton's analysis that we may not want to buy into. For example, Railton is focused on the discomfort one might feel after failing to comply with a rule. Must it always involve such discomfort? For example, one might feel no discomfort at all from a failure to comply with rules but plenty of satisfaction from compliance. To illustrate, consider Church Lady (borrowed from the old Saturday Night Live routine), who is puffed up with moral pride. Let's suppose she never feels guilt or discomfort at failure to follow a rule (indeed she cannot recognize that she fails to be moral). She never forgets to validate her bus pass. On some days, however, the bus is very crowded and it becomes inconvenient to validate. She realizes that on such days she will not be sanctioned for failing to validate. Nevertheless, she validates her ticket and feels a sense of satisfaction with herself and her conduct.

Railton is showing his Kantian stripes here in thinking that Church Lady is not ethically rule guided. Kant would say that Church Lady is not acting out of respect for the law; she is merely acting out of the sense of pride she feels in her action. I'm not an expert on moral psychology, but I am suspicious of theories that take actions motivated out of moral pride to not be cases of behavior that is governed by moral rules. Church Lady was puffed up with moral pride, but I have trouble thinking that she was not for all that often guided by moral rules.

Now to be sure if we allow that cases like these are instances of rule governance we need to rethink the weight put on the notion of a regulator in Railton's proposal.[2] A regulator, after all, detects that something is wrong and attempts to return the system to equilibrium. That works in the case of someone who feels discomfort at not validating their ticket (validating returns the system to equilibrium by

eliminating the sense of discomfort) but it doesn't seem to work for Church Lady, who is not trying to return to a state of equilibrium but is attempting to accumulate as much satisfaction as she can.

Regulators aren't the right metaphor for this, but there are other metaphors that can be employed here. Obviously, there are systems that strive to maximize for certain properties (for example consider a system that attempts to absorb as much sunlight as possible). Let's call these systems *optimizing systems*.

Here is how we might have to structure the theory for Church Lady.

> **(RG5-CL)** Agent A's conduct C is guided by the norm N only if C is a manifestation of A's disposition (i) to act in a way conducive to compliance with N, such that N plays a role in A's C-ing, where this involves some disposition on A's part to notice that C-ing will optimize for her sense of pride, and (ii) to exert effort to C *even when* failure to C is unnoticed and non-consequential.

If we wanted, then we could construct a disjunctive analysis that would allow either route to normative guidance - one where the primary moral sentiment is discomfit at failure to comply and one where the primary moral sentiment is pride at compliance. But does this analysis have to be hooked to a particular moral sentiment at all? That is, does normative rule guidance involve a distinctive phenomenology?

Consider Zombie Girl, who is phenomenologically impaired as regards her rule compliance. If she fails to validate her ticket she feels no discomfort. If she goes out of her way to validate her ticket, even when it is inconvenient to do so, she feels no satisfaction. She validates, alright, but there is no moral sentiment involved. She simply judges that it is the thing to do, and that is why she does it.

Humeans like Prinz (2007) might argue that these sorts of cases aren't possible—that normative judgment necessarily involves some emotional content. But again I wonder if this is right.

For example, there is a version of Kantianism described by Korsgaard (1996) according to which not only is Church Lady not acting morally (she is acting out of pride rather than respect for the law) but the same might be said of the person who is acting merely so as to ease discomfort. After all, how is acting to ease discomfort the same as acting out of respect for the law? On this view the phenomenology attending our acts of rule governance might provide *evidence* that we are acting in a rule governed way, but the sentiment is not *constitutive* of our so acting.[3]

I am pressing against the role of moral sentiment here because if we want to extend Railton's analysis of rule governance to the linguistic case there is a very serious question as to what the relevant sentiments might be. In the case of traditional prescriptive grammar there is an easy answer: I might feel shame or embarrassment at not speaking properly. That's fine for learned linguistic rules like using 'whom' not 'who' in dative case, but it's less clear how to make sense of the kinds of linguist rules generative linguists are concerned with. It also isn't clear that the sense of embarrassment one feels at failure to comply with prescriptive linguistic rules is about failing to comply with a linguistic norm—it might simply be due to telegraphing one's social or educational status. The prescriptive rules are just like secret handshakes in this case. If you don't know the prescriptive rules then you are marked as an outsider or a social inferior and thus you feel bad. Obviously these sorts of prescriptive norms are established by power relations, and obviously they have nothing to do with the individual norms that would be operative in the context of generative linguistics.

Take the pending violation of the Ross Constraint the agent sees in the example discussed by Lawlor. Does the agent correct and add the resumptive pronoun because she sees that she would have felt shame or embarrassment at failure to do so? Just what is the attending phenomenology? And is there any reason to believe that her phenomenology is shared by other agents? That is, even if there is some sentiment that she has when she sees an impending violation of a Ross Constraint, why should she suppose that others have a similar sentiment in similar circumstances? Maybe violations of Ross Constraints are like fingers on a blackboard to her but like a dull headache to me. Or maybe like Zombie Girl there is no phenomenology at all attending my insertion of a resumptive pronoun.

If we want to extend the analysis to at least allow such possibilities then it seems we want to revise Railton's analysis to something like (RG6):

> **(RG6)** Agent A's conduct C is guided by the norm N only

> if C is a manifestation of A's disposition to act in a way conducive to compliance with N, such that N plays a role in A's C-ing, where A judges that her efforts to C, whether successful or not, are the thing to do—particularly when so acting is difficult and unsanctioned, and even when inconsequential. Some of A's failures to try and C may give rise to A's feeling discomfort and A's efforts to C, whether successful or not, may have lead A to feel a sense of satisfaction—of having "done (tried to do) the right thing". These sentiments may signal to A that she is acting appropriately but they are not constitutive of her being rule governed.

While this unlinks rule governance from moral sentiment there is a sense in which it falls short of some of the goals of Railton's proposal. Notice that to escape talk of moral sentiments we are now leaning on the idea that A is judging that C is the thing to do, and thus it looks like our analysis of rule governance has normative judgment packed within the analysis. Is this ok?

If we were looking for a completely reductive analysis of rule governance then it isn't ok as it stands. On Railton's proposal we had a reduction in which the analysis of rule governance bottomed out with a particular sentiment—a particular feeling of discomfort. Now we bottom out with a normative judgment.

Still, we might think that there is a difference between bottoming out in a normative judgment and bottoming out in rule governance. That is to say, maybe it was never part of the game to expunge normative elements from the analysis—maybe the goal was to have the analysis of normative rule governance bottom out with the right kinds of normative elements. In this case it bottoms out with our judgment that something is either right or not right.

For the record, it's not clear that Railton has a clean reduction to the non-normative either. After all, we might ask how moral sentiments are to be individuated. Not every physical affect is a candidate (for example the need to sneeze); only special affects are—the emotions that are distinctively *moral*. So we might think that the relevant sentiments can't be basic.

Meanwhile, if we are engaged in a reductive project, perhaps the detour through moral *judgments* is necessary to get the reduction. For

example, these normative judgments could be given an analysis in terms of expressivism. When I judge that C is the thing to do, perhaps that just comes to me having a Pro-attitude about C-ing in this instance (I would express this attitude as "yay C-ing!"). Or to use Mark Schroeder's (2008) formulation, perhaps I am "*for* C-ing." In the linguistic case, this invites the formulation that I have a grammar that establishes rules *for me*, and when I am normatively guided by those rules I am *for* acting in a way conducive to compliance with those rules.[4]

Alternatively, there is the Gibbard (2003) formulation in which I *plan* to do C or have a kind of contingency plan according to which I would do C. Personally, I like the Schroeder formulation better, because it does a better job of handling the case of Fred and his morning snack. Recall that we did not want to say Fred is normatively guided, but surely he did *plan* on a morning snack. It is not at all clear, however, that he is *for* having a morning snack. In fact, if we asked him we can imagine him saying "of course I *planned* on having a morning snack, but it's not like I was *for* it—it was just something I always do."

Whichever formulation we eventually end up with, the point is this: If we are interested in giving a reductive account of rule governance we might find the move to normative judgments a more productive first step than attempting to make the reduction directly to moral sentiments. (Notice that we are not necessarily talking about a reduction of normative facts here, just a reduction of the normative judgments.)

Whatever we might say about standard moral cases, I have to say that I like this approach better for accounts of individual normative governance in the case of linguistic rules. We encounter a subjacency violation and we reject it, not because we have a feeling of discomfort about it, but rather because we simply are *against* saying that sort of thing. And that is the end of it. Similarly, we are *for* there being a binding relationship between 'himself' and 'John' in 'John saw himself', but we are against such a relationship in 'John's mother saw himself'.

So far I've been supposing that the kinds of moral judgments made in cases like Huck Finn are parallel to those cases in which we judge a particular example to be unacceptable or we judge a particular interpretation to be possible for a structure. But the parallel is inexact.

In part 1, I made a distinction between the grammar, linguistic facts/phenomena, and our judgments about those linguistic phenom-

ena. In the Railton proposal there doesn't seem to be a place for moral facts/phenomena. Suppose we tried to introduce that element before we developed our general theory of individual normative guidance. The parallel moral case would be something like this:

Moral rules/principles → moral facts/phenomena ← ethical judgments
(establish) (provide evidence for)

So in the case of Huck Finn, let's suppose that the moral principles are something like "always treat persons as members of the Kingdom of Ends" or some other version of the categorical imperative. This general principle establishes and explains the moral fact that Huck ought not to turn in Jim. Huck's judgment is that turning in Jim is not the thing to do; we can formulate this in various ways—for example that Huck is *for* not turning in Jim (Schroeder's formulation), or he *plans* to not turn in Jim (Gibbard's formulation). As in the linguistic case Huck has no access to the deep underlying principle (it took a genius like Kant to discover that, after all). Huck does not even have direct access to the explanatory description of the moral fact—this too is a deep and important discovery. Huck merely judges that turning in Jim isn't the thing to do. We as theorists can take this as evidence for the fact that turning in Jim is morally wrong. Obviously Huck doesn't take it as evidence for this fact. But it is not necessary that he do so in order to be morally guided by the categorical imperative.

We are still short of having a satisfactory notion of rule governance, however, since we may want to incorporate the Lawlor case in which some norms trump others (and in which we correct ourselves on the fly). In linguistics, rule ordering is standard fare. Kant and Peter Geach famously argued that moral conflicts just don't happen, but this seems optimistic to me. I see no reason why the notion shouldn't be employed in ethics. We might also allow that circumstances can make conduct conforming with N imprudent (e.g. validating may require muscling aside a street gang).

> **(RG7)** Agent A's conduct C is guided by the norm N (where N is an element in the set of norms N*) only if C is a manifestation of A's disposition to act in a way conducive to compliance with N, such that N plays a role in A's C-ing, where this involves some disposition on A's part to judge that C is the thing to do unless there are extenuating circumstances and/or conflicting actions C' that are conducive to compliance with higher ranked norms in N*, and to exert effort to C even when any failure to C would be unsanctioned and non-consequential.

(Norm Ranking) Given conduct C that is conducive to compliance with N (where N plays a role in A's C-ing), and conduct C' that is conducive to compliance with N' (where N' plays a role in A's C'-ing), N is higher ranked than N' only if A has a disposition wherein having to chose between C-ing and C'-ing A judges that C-ing is the thing to do, all other things being equal.

This formulation not only works in the ethical case but also seems apt for the sorts of linguistic rule—following cases we are interested in. Agents may not be aware of the linguistic rules—indeed, how could they be?-but their behavior is nonetheless guided by the rules. Like Lawlor's agent, we judge that something is not right and we repair it, or do the best we can to repair it, even though our failure to do so is unsanctioned and inconsequential.

3. CONCLUSION

I've spent rather a lot of time on what it means to have (and be guided by) a linguistic rule and much less time on the question of what it means to know a fact about one's language, but the time allocation is not driven by importance, just by what needs to be clarified most. If I am right, then there is a tight connection between our knowledge of linguistic facts and our having linguistic rules. The linguistic rules provide normative guidance for us, and our linguistic knowledge (of S-facts) helps us to stay in compliance with those rules. Having rules without also having knowledge of linguistic facts would leave us blind. At this point you are probably expecting me to say that having linguistic knowledge without having rules of grammar would be empty, but it is worse than that, for without the rules there are no facts, and hence no linguistic knowledge to be had.

Notes

[1] Thanks to Herman Capellan for discussion here.

[2] Thanks to my Spring 2009 grad seminar participants for pointing this out.

[3] Once again, thanks are due to my Spring 2009 seminar participants for discussion of this point.

[4] Thanks to Barry C. Smith for this very clever formulation.

References

Arpaly, N. 2003. *Unprincipled Virtue: An Inquiry into Moral Agency*. Cambridge: Cambridge University Press.

Bogen, J. & Woodward, J. F. 1988. 'Saving the Phenomena'. *The Philosophical Review* 97, no. 3: 303–352.

Chomsky, N. 1980. *Rules and Representations*. New York: Columbia University Press.

Gibbard, A. 2003. *Thinking How to Live*. Cambridge: Harvard University Press.

Korsgaard, C. 1996. *The Sources of Normativity*. Cambridge: Cambridge University Press.

Prinz, J. 2007. *The Emotional Construction of Morals*. Oxford: Oxford University Press.

Railton, P. 2006. 'Normative Guidance'. In R. Schaefer-Landau (ed.) 'Oxford Studies in Metaethics', 3–34. vol. 1, Oxford: Oxford University Press.

Schroeder, M. 2008. *Being For: Evaluating the Semantic Program of Externalism*. Oxford: Oxford University Press.