Bericht Nr. 78, August 1996

Arbeitspapiere des
Sonderforschungsbereichs 340

ISSN 0947-6954/96

Sprachtheoretische Grundlagen für die Computerlinguistik

Arthur Merin:

Neg-Raising, Elementary Social Acts, and the Austinian Theory of Meaning

Universität Stuttgart Universität Tübingen
IBM Deutschland Informationssysteme GmbH

Arbeitspapiere des SFB 340
Bericht Nr. 78, August 1996

NEG-RAISING, ELEMENTARY SOCIAL ACTS, AND THE AUSTINIAN THEORY OF MEANING

ARTHUR MERIN

ABSTRACT. A quirk of natural languages, noted by St. Anselm, Quine and Hintikka, is 'neg-raising' (NR): the apparent self-duality, with respect to negation, of epistemic or deontic complement-taking predicates including *believe, seem, expect, should, intend, want.*(E.g. 'I don't think that A' being near equivalent to 'I think that not-A', etc.) Larry Horn discovered that the items in question are 'mid-scalar' on scales of 'strength'. This notion is shown to be still in want of further explication. Flaws in Horn's standard 1978/1989 account of Neg-raising are demonstrated. A directly act-denoting, kinematic semantics is proposed for the expression paradigms involved. It hinges on a paradigm of Elementary Social Acts—claim, concession, denial, and retraction—defined by decision-theoretic parameters of contexts, their transformations, groups under composition, and quotient groups. 'External' and ' internal' occurences of *not* map to context- and act-type-changing operations defined in terms of dominance (a.k.a. warrant) and preference. Mid-scalars are explicated as instances of a quotient act-type 'suggestion', also representable as a mixture, probabilistic or otherwise, of claim and concession. The analysis predicts the effects of external negation on 'weak' and 'strong' members of scales, as well as NR for mid-scalars. Rather than rely on metaphoric appeal to operator paradigms in modal logic it offers a formal rationale for the obvious analogy. It exemplifies thereby how a semantics for speech acts, directly in terms of operations on context (and thus arguably in the performative spirit of J.L. Austin), will condition much-debated linguistic data, bypassing the traditional route via truth-conditions. The modal *actually* and its usage in Leibniz are briefly addressed along the way.

ISSN 0947-6954/96

1. Neg-Raising: A Systematic Quirk of Natural Languages

One of the least tractable problems in syntax / semantics / pragmatics of natural languages, say Gazdar and Klein (1980:96), is 'neg-raising' (NR): the apparent equivalence—*mutatis mutandis* in many languages—of (a) and (b), in sentence pairs such as

(1) a. I don't {think/believe/suppose} that Kim is drunk
b. I {think/believe/suppose} that Kim isn't drunk

(2) a. I don't {want/expect} Kim to be drunk
b. I {want/expect} Kim not to be drunk

(3) a. It isn't likely that Kim is drunk
b. It is likely that Kim isn't drunk

whenever the verbs or adjectives in question are lightly stressed.

The name NR, as usual in linguistics, derives from a 'transformational' account. Negation, so the story went, was 'raised' or 'transported' out of the 'deep-structure' equivalent of the embedded clause (a position reflected in the b. versions) past the predicate in question, into the matrix clause (a. versions). By classical definition of transformation, indeed by its very raison d'être, this meant: under preservation of meaning. (NR pre-dates Partee's work on quantifiers, which killed 'deep structure' at least as the input to semantic interpretation.)

In fact, scholars had quite early on noted subtle differences of meaning, the (a) versions generally being held to be 'weaker' or 'more polite'. But *pace* Horn, this is not always evident. Example from Horn (1978:132): *I want him not to come* may imply that it is for his sake that I want him to refrain, rather than as in *I don't want him to come* for reasons of my own. If so it is no longer clear that the first is felt to be, as Horn presupposes, 'stronger or ruder' than the second. Quite the contrary.

Hence—and however the explanatory burden for NR is apportioned among syntax, semantics and pragmatics—the origins of NR usage need not lie in politeness phenomena which have become sedimented over time to acquire a well-nigh syntactic phenomenology. Rather more forceful as a functional rationale is Horn's concluding hypothesis that there is a premium on signalling 'negative force' as early as possible in the negativesentence (1978:215). This is a rationale also motivated by the overlapping 'embedded negation constraint' [ENC] (Horn 1978a) which militates against lower clause negation.[1]

Horn's (1978) monograph (summarized in Horn 1989:308-330) remains to this day what the reviewers then judged to be 'without question the definitive paper' on the subject and, incidentally, 'the most thorough exploration of a single rule of grammar since Postal's epic 1974 book on raising'. Most of the linguistic and historical facts I shall refer to are thus reviewed or established in that paper. And so is the major descriptive thesis about NR. That being so, read a tacit 'as Horn shows/relates/...' wherever the context does not imply otherwise.

NR has been treated as a freak phenomenon of little significance by philosophers, including Quine (1960) and Hintikka (1962), attending to expressions of propositional attitude. St. Anselm, by contrast, seems to have taken it more seriously in his discussions of *non debere peccare* and *debere non peccare*.[2] But the systematicity of the quirk across languages, indeed its near-universality, suggests something else: that natural languages work in ways quite different from those envisaged, say, in Montague's 'English as a Formal Language'.

What is so interesting about NR? 1. NR predicates are complement-taking verbs, adjectivals or participles with epistemic or deontic-volitive meanings (they include *seem, should, ought to, suppose(d to), be a good idea*, in some dialects also *suggest*); not, for instance, literally interpreted perceptuals (*see*) or unambiguous causatives (*persuade*). 2. Neighbouring predicates within the respective semantic fields fail to neg-raise: the equivalences do not hold e.g. for *know*, *suspect*, *desire*, *demand*, *certain*, *assert*, *admit* or *possible*. 3. NR predicates (in English, German, ...) do not take prominent stress, and tend to acquire readings that fail to neg-raise when they are stressed. 4. NR verbs raise most securely in first person present tense usage. 5. NR verbs appear to denote 'states' rather

[1] In fact, the proposal I shall make below involves an assumption which needs relatively few auxiliary assumptions to predict an ENC.

[2] The 'seems' bears on 'it': if *debere* were to translate 'must', —not a mid-scalar on any account—then this might be a phenomenon of 'negative-attraction' akin to that which lends *mustn't*, unlike *needn't* or *can't*, the interpretation of an internal negation, xsupposing that *must* designates deontic necessity throughout. Cp. also the history of non-mid-scalar modal German *d"urfen*, which changed its meaning from 'must' to 'may'.

than 'acts' or 'activities' on those readings which neg-raise.[3] 6. NR predicates appear to indicate 'uncertainty' on their agent's part about the evidential or similar status of the complement.

Properties 1, 3 and 4 suggest that neg-raisers (NRs) function somewhat in the manner of 'parentheticals'.[4] I.e. that they signify, intimate, 'show' or—let us revive this unloaded term—*ostend* the speaker's epistemic (evidential) or ethico-aesthetic attitude to the proposition denoted by the embedded clause. Rather than describe the speaker's state of mind, activity, or as in exx. (3) an ostensibly objective property of the world. Property 5 is prima facie at odds with the performative element suggested by 4, and I shall say more about it below, in line with the intuition engaged by 6.

Property 2 poses a major problem of description, solved by Horn (1978): neg-raisers are situated at (or rather just above, as Horn specifies) the intuitive middle of a positive 'scale' reflecting an intuitable ordering in terms of 'strength' of broadly lexical or phrasal semantico-syntactic paradigms. Intuitively, says Horn, NRs instantiate the notion 'opinion' (about what is, or ought to be, the case).

The scales are operationalized by linguistic diagnostics standard since Horn (1972):

(4) It is {possible if not indeed likely/likely if not indeed certain} that Kim will come

(5) I am not only allowed to talk, I (indeed) want to talk

(6) I not only want to talk, I (indeed) must talk

(7) X not only R, R' X

where R $<$ R' and $<$ is a strict partial order relation.

However, ordering by 'strength' does not always lead to unambiguous results; consider alongside (6) the converse diagnosis in

(8) I not only want to talk, I am (indeed) allowed to talk

This apparent multi-dimensionality of scales, the most obvious explanation for the ambiguity, suggests that there might be problems about the

[3] As a necessary condition, Georgia Green's characterization 'non-factive verbs of [i.e. designating a] mental state', e.g. the 'opinion' sense of *believe* rather than the 'accept' sense, appears to be apt. However, Horn showed that it was far from being a sufficient condition.

[4] After Urmson, 1952, broadly in line with the later Wittgenstein and with Austin 1946 and Austin's subsequent publications collected in the posthumous 1961, 1962 volumes.

definition of scales: it remains unclear what our diagnostics are diagnostics of. In fact, as Horn noted, scales come dually in pairs, one 'positive' and a corresponding one 'negative'; e.g.

(9) 0 possible likely certain

(10) impossible unlikely not certain 0

The 'direction of implication' or 'ordering of strength' reverses at the point marked 0 if one decides to stand the scales end to end.

Motivation for speaking in terms of 'implication' is this: one cannot coherently deny that something is possible or likely if one maintains that it is certain; similarly, one cannot deny that it is not certain or unlikely if one holds it impossible. However, such a rationale is no longer so compelling when some of the verbs in question are involved: one can very well deny that one is conceding a point when one is, in fact claiming it. (As distinct from denying the point.)

There is another mystery about the scales. What Horn observes (1978:195) about dual pairs is the following:

- a. The (contradictory) negation of a weak item on the positive scale [Pos] (e.g. *possible, allow*) is strong on the corresponding negative scale [Neg] (*impossible, forbid*).
- b. The negation of a strong item in Pos (e.g. *certain, have to*) is weak in Neg (*not certain, do(es)n't have to*).
- c. The (contradictory) negation of a mid-scalar in Pos (e.g. *likely, advisable*) will be intermediate in Neg (*not likely, not advisable*).

(Horn excludes 'contrary' negatives such as *unlikely, inadvisable* from direct consideration and suggests derivation by NR from structures embedding a lower negative.)

A predicate F taking a propositional complement Φ, says Horn (1978:195), is above the mid-point of Pos iff $F\Phi$ & $F\neg\Phi$ is a contradiction. This seems to work as a condition on NRs (if we exclude mid-sentence shifts of deontic point of view for *ought to* etc.). But the 'if' part is too strong.

1.1. **Problems with the Definition of Scales.**

Intuitively, *admit*, local synonym of *concede*, seems to be one end of a scale of which *assert* would be the other and for which *suggest*, *suppose* or *advise* might be mid-scalars. And yet, admitting Φ and also $\neg\Phi$ looks dangerously close to contradicting oneself.

Preserving consistency with this phenomenon, Horn classes *admit* as a 'strong assertive' (1978:196); which sounds itself close to being a contradiction in terms. I take it that *I do not assert that* Φ comes out intuitively 'weak', whereas *I do not admit that* Φ comes out as a denial, i.e. rather stronger than the former or *I do not suggest that* Φ. Which, by Horns's generalization, says that it was weak(er) on Pos to start with. If *admit* is not a synonym of *assert* then the interpretation given to scales leaves unclear what differentiates this instance of apparent contradiction from those arising with more obviously higher items on the scale. In fact what is the interpretation?

Already for *likely*—seemingly the best candidate for a semantics that motivates the scale—matters are far from neatly arranged. As the natural denotation range Horn appears to have had in mind probabilities of complement sentences. For his scale is stronger than purely ordinal, admitting, as we shall see, of additive operations, and he notes that *likely* does not, in fact, denote a propositional attitude *per se* (1989:327). Hence probabilities would, at least here, be a reasonable space of denotata to impute to the author. That is, our scale would have to be, or be a representation of, the real interval $[0, 1]$.

Apparently renormalizing, then, Horn maps Neg onto the half-open interval [-1,0) and Pos onto (0,1]. Direction of pragmatic entailment reverses at 0. Let v stand for a real-valued function which determines the infimum of an interval which, according to Horn, represents the range of 'strength' of its argument.[5] For the "weak positive scalar", *possible*, Horn has $v(possible) \in (0, 1]$; $v(impossible) = -1$, and, presumably then, $v(certain) = 1$. This makes it hard to interpret the scale as a probability scale if we are to accomodate the intuition that anything not impossible is possible, i.e. if we do not assume a scale v' such that $-1 < v'(possible) \leq 1$. And it seems that this interpretation is not in fact intended.

But even within its own purely formal terms the specification is problematic. Horn's general rule is essentially (and we do not worry about distinguishing *min* and *inf* and let Pos and Neg now designate sets of numbers representing sets of lexical items):

[HR] If $Y \in \{X : min[v(X)] = k \in \text{Pos}\}$ then $v(\neg Y) = k - 1 \in \text{Neg}$.

[5]'The position which a logical operator occupies on a strength scale is ... actually an interval bounded on the lower end by the point assigned to it ... and on the upper end by the end-point on the relevant scale' (1978:195).

For symmetry's sake we might then add: $\neg Y \in \{X : max[v(X)] = k-1 \in \text{Neg}\}$. Given the scale infimum at -1 this yields a point value.

Now *likely* is situated "just above .5 on the epistemic scale, and its contradictory at $-.5$ on the negative counterpart. It is this position ... which harbors all neg-rasing predicates." (1978:195f.) Thus, presumably

(11) $v(\textit{likely}) = 0.5 + \delta$; $v(\textit{not likely}) = -0.5 + \delta$ $(0 < \delta$, where δ is small.

But $v(\textit{possible}) = 0 + \delta$, and since semantically $|\textit{impossible}| = \neg|\textit{possible}|$, [HR] predicts $v(\textit{impossible}) = (-1 + \delta) > -1$. If impossibility is the end of the scale, i.e. at -1, we are faced with a contradiction.

To get things right one should have to amend [HR] to [HR'] yielding $v(\neg X) = k - (1 + \delta)$. This is not, perhaps, the most elegant of operations. Note also that, by [HR], $v(\textit{not certain}) = 0 < 0 + \delta = v(\textit{possible})$, whereas by [HR'] we should have $v(\textit{not certain}) < 0$. If the composite-scale interval [-1,1] did normalize to a probability-scale interval [0,1], then presumably *likely* would be at $p \approx 0.75$ and *not likely* at $p \approx 0.25$. Is "just above" merely the reflex of a confusion between this scale (where *likely* should correspond to $p > 0.5$) and its renormalization? Not so, it appears. The same positioning is claimed to hold also for *believe* and *want* on their respective scales (Horn 1989:325f.).

These problems of consistency are perhaps repairable.[6] But they bespeak uncertainty about conceivable denotational semantics for neg-raisers. And there is no independent motivation for the $k-1$ rule let alone its δ-emendation; nor an obvious non-arbitrary way to extend it to other scales.

Of course, one might now object that a straw person of an argument is being attacked here. For *if* 'scales' are given by entailment relations among expressions α, β derived from entailments among minimal pairs of sentences $\sigma = X\alpha Y$; $\sigma' = X\beta Y$ (as proposed by Horn 1972 and formalized in Gazdar 1979), then the effect of 'external' negation is, in essence, no more than a concomitant of the validity of contraposition, i.e. of $p \rightarrow q$ iff $\neg q \rightarrow \neg p$. Or, in the symbols appropriate to the term 'scale', $p \geq q$ iff $\neg q \geq \neg p$. Negation is an inversely monotone operator with respect to the order relation of entailment.

There are, however, problems with this view. Fauconnier, an early analyst of scale reversal and, concurring, Gazdar (1979) maintain that the ordering is 'pragmatic'—not strictly one of entailment giving rise to an ordering in terms of semantic informativeness. This leaves wide open

[6]One might consider 'fuzzy' specifications if inclined to deviate minimally from Horn's explication.

what 'pragmatic' means;[7] and Gazdar, not an author given to gratuitous diffidence, concludes that the scales are 'simply given to us'.

Problems with the entailment view of 'scales'[8] have been well-known in the past. Thus, Grice's (or was it not originally Austin's?) early example of what became known as 'scalar implicature', induced by the pair *It is red* and *It looks red to me*, does not rest on a relation of entailment between the two sentences, and hence cannot establish a derivative one for *is* and *looks red to me*. Well, we could say: let's sanitize with a 'normally'. But where do we stop then?[9]

Even supposing that we have an ordering 'given to us' and have identified (external) negation as an inversely monotone operator with respect to all such 'given' orderings: there is no obvious link between these very general structural properties and the specific intuitions attaching to NR phenomena, including the related ENC.

Another point to note—one which I hope to defend in detail elsewhere—is that the characterization of NR verbs in terms of mental *state* might usefully be amended to one in terms of *disposition*. Positive evidence for this is the paraphrastic equivalence of most NR occurrences, at least in the first person present, to a construction involving the NR expression *would say*. More positive evidence is the metapragmatic intuition of agent's 'uncertainty', noted by scholars. It suggests, if anything, a stochastic aspect to the definition of 'state'. On either count we have a motivation for explicating 'non-action' of sorts as a potential for action.

The problems with explication of 'scales' and the suggestion about disposition should raise the attractiveness of attempting a treatment broadly in line with Urmson's Austinian approach, by way of characteristically ostended speech act. This option, whose origins lie in Austin's and Wittgenstein's denials of the 'descriptive fallacy' underlying contemporary analyses of first-person uses of *know* or *believe* is briefly considered by Horn under the labels 'performative' and 'semiperformative' (1978:191). In Austin's terms one might simply say: 'primary (as opposed to 'explicit') performative'.

While acknowledging the option, Horn demands not unreasonably that it also account for peculiarities of uses other than first person present (FPP), those which Urmson described as cases of '*oratio obliqua*'. Horn

[7] If it means 'entailment relative to background assumptions' it is like a license to print Deutschmarks: who can stop you from maintaining that $p \to q$ is part of the background?

[8] A usage of the term introduced, it seems, in Urmson (1952)

[9] Horn, too, notes that the NR quasi-modal *ought to* and its presumable upstairs neighbour *must* do not stand in relations of entailment (1978:202).

notes a number of constraints on such embedded uses which a comprehensive account ought to explain. But even a more modest account, as he rightly says, will not characterize the class of NRs simply by appeal to performativity. Any number of non-NRs fit Austin's and Urmson's description.

One should add a question about possible mechanisms. Horn (1978:177), in a discussion of 'queclaratives' (negatively pointed questions) mentions the idea of 'illocutionary force' being read off surface structure directly. But how is it to be read off? And what is the relation between reading off and the hygienic demands of compositionality? In other words: do we simply choose the soft option of a deeply noncompositional theory of meaning, after the manner associated with the later Wittgenstein? Or do we—contra Austin—compute standard truth-conditional meanings on which holistic, non-compositional analyses of force operate?

Suppose we take the Austinian route seriously. Then we want something more revealing than a bland statement that, somehow, the predicates under investigation designate 'illocutionary forces' that happen to confer upon the sentences their apparent structural properties. A look at the (neo)-classical literature (Searle and Vanderveken 1985; Vanderveken 1991) cannot raise hopes that insights are to be gained by augmenting 'illocutionary points' with rankings according to 'strength', and weakening or strengthening operations to match ("isomorphic to the abelian group of integers"). Nor is there any obvious merit in proposing an 'illocutionary negation' (Searle and Vanderveken 1985) defined on all and sundry speech act types which somehow—mysteriously—happens to deliver results reminiscent of the modal paradigm on a few select of them.

However, before proposing the rudiments of a solution, we must attend to a structural property of NR which, obvious as it is, offers a useful heuristic on the way.

2. Structural Properties of NR: Self-Duality

On a broadly truth-conditional theory of meaning (pick your favourite surrogate for truth), neg-raisers (NRs) are most readily represented as propositional operators $\mathbf{R}$ more or less *self-dual* with respect to negation:

[SELF] $\quad \neg\mathbf{R}\Phi \simeq \mathbf{R}\neg\Phi.$

'More or less' means: give or take a bit of idealization, i.e. disregarding the subtle differences of meaning noted or suspected in the literature.

Of course, self-duality suggests kinship with the dual propositional operators familar from modal logic (and their object-level quantificational analogues) for which the following relations hold:

(12) $\neg\mathbf{L}\Phi \leftrightarrow \mathbf{M}\neg\Phi$

(13) $\neg\mathbf{M}\Phi \leftrightarrow \mathbf{L}\neg\Phi$

English near-correlates of these two operator types, i.e. alethetic, epistemic or deontic versions of necessity (L) and possibility (M), do indeed crop up as endpoints of some of Horn's 'positive' scales. However, in standard one-dimensional (S$n.m$) modal logic there is no operator which, intuitively stands between the two.

Well, one might say, perhaps we have not looked hard enough. Why not introduce, redundantly, an identity-operator **I** by way of

(14) $\mathbf{I}\Phi \leftrightarrow \Phi$

for which it is harmless to demand

(15) $\neg\mathbf{I}\Phi \leftrightarrow \mathbf{I}\neg\Phi$.

An English alias, in a classical framework, might then be the word *true*. Considering its ordinary use conditions in discourse, compare

(16) a. It is not true that Φ
b. It is true that not Φ —

which might suggest a duality of *acts* reminiscent of one also found in NR predicates: external negation goes with a denial of another's claim that Φ, internal negation with an assertion that $\neg\Phi$. In a bivalent framework supporting a redundancy account of *true*, the sole difference in meaning between the two utterance-types would be in respect of an ostension about who took the *initiative* in the ongoing conversational exchange. In the first case we could say: ostensibly not the speaker.

The widespread intuition—not in any obvious way related to truth-conditions—that external negation is somehow less impositive would be immediately accounted for. To deny an assertion is either to refuse to assume a new epistemic commitment or else to respond with a commitment to the contrary (either by way of a bivalence convention brought into force by the raising of Φ as an issue, or because a prior and unshaken commitment to $\neg\Phi$ is at the root of one's denial).

One problem with this account is that it misprepresents the linguistic facts. The internally negated version is used to assert (rather than admit) only in the exceptional, 'marked' case where the speaker *insists*. In general, and as Oxford philosophers were well aware of in the 1950s, *It is* *true that* is the prefix of a symbol string used to *assent to*, indeed *concede* or *admit* its complement proposition. (Insistence requires emphatic stress on *true*.)

If so, (16 b) will also assent to another's assertion: this time to that of the negative proposition. And this should, if anything, be an even less impositive way of bringing across one's commitment to $\neg\Phi$. In a two-person universe of dialogue partners it would be akin to a qualifier: 'You said it'. A suitably purified operator *True* might do the trick; but we would no longer be making fullest use of the data.

There is also a larger problem: how to translate a basically Austinian account of *meaning intuitions* into a piece of formal semantics that will predict a few facts into the bargain. But before addressing that issue, there is another modal approach to mention.

The most obvious alethetic modal operator that satisfies schema [SELF] is that expressed by *actually.* For, intuitively, we have

(17) Kim didn't actually come $\simeq$ Kim actually didn't come

For Leibniz, the most anthropological of metaphysicians, actuality stands as a *contingency* midway between possibility and necessity: indeed, as an optimal compromise between the demands of simplicity and requisite variety facing the Creator. As such, it instantiates the kind of 'moral necessity' which Leibniz sees characteristically warranted by those 'motives which incline without necessitating' (Leibniz 1716:V,§9). This is remarkably close to a being gloss for the argumentative force of NRs, but it does not offer a semantics for them.

In contemporary logic, *actually* is explicated by two-dimensional modal operators whose purpose, though not invariably the finer structure of definition (cp. Crossley and Humberstone 1977), is analogous to that proposed by Kamp (1971) for *now.* On all such accounts it is either by explicit stipulation (Crossley and Humberstone) or trivially implicitly self-dual. Its non-redundant use is exemplified by

(18) Kim thinks she is taller than she (actually) is

The version without *actually*, in essence due to Russell, gives rise to ambiguity among readings one of which imputes contradictory belief to Kim. With *actually* we have a clear indication that Kim's belief is, according to the speaker, false; but not that it is contradictory. Like *now* does for the moment of utterance in

(19) A boy was born that would (now) be King

it shifts the point of reference to that of the speaker or more generally, the index deemed actual.

However, it is not clear that such truth-conditional explications of *actually* offer solutions to our problem. For a start, they are non-redundant only in the scope of intensional operators or relations (e.g. belief, possibility, tense), whereas *believe* in NR occurrences does not in general so occur on currently defensible analyses.

Nor is it clear that the general framework of two-dimensional modality which turns on the notion 'denotation (e.g. truth-value) of an expression (e.g. sentence) at an index (e.g. world) i with respect to an index (e.g. context) j' offers scope for an interesting distinction between negation internal and external to *actually*. The best we could look out for is some distinction between, say, 'not true at i from the speaker's (marked as the actual) point of view' and 'true at i not from the speaker's (here: the actual) point of view'.

This might address a distinction that engages the intuition about the use of *actually* which Crossley and Humberstone call 'rhetorical'. It serves for 'dispelling or forestalling a misapprehension on the part of the hearer, as when one says "Actually it was March, not April, when we bought the house ... "'. The empirical observation to make immediately is that *it is true that* introduces *assent*, whereas *actually* ostends, if anything, a *denial*. Thus *not actually* might be glossed as 'not, as you (or presently third parties) hold'

Again, it is not obvious how this should capture the required distinction. Versions with negation apparently inside the scope of *actually* are a bit clumsy, which goes for many of the other NR examples too and is, presumably, an instance of the ENC. But they do not seem to differ in meaning from the externally negated version.

On the other hand, the two-dimensional analysis is suggestive; and some of its suggestiveness does carry over to the analysis I propose. For suppose the j ranges over a set $\{j_1, j_2\}$ whose elements are respectively interpreted as the points of view of speaker and ostensible addressee. And consider now also unnegated occurrence. The intuition to be pursued is of a tacit presumption that, for relevant propositions, truth-values will as a rule be assigned differently with respect to the two indices.

This is a wholly extra-logical assumption, of course; and best seen as a piece of social psychology.[10] If we make it, there would be two ways of denying a proposition even while retaining bivalence: (i) directly by

[10] But then, unlike pure logic, natural languages are, if anything, the province of (mathematical) social psychology.

declaring it false with respect to one's *own* index (asserting the negative, obtained by propositional negation to be true w.r.t. it); or by noting it to be true w.r.t. the *other* index or indices (i.e. operating on the set of indices, let us say, to make the point vividly, by complementation applying to the singleton whose member is one's own index.)

3. A Discourse Institution and Elementary Social Acts

Suppose we follow the Austinian route of bypassing the computation of truth-conditions or taking them to be pre-imputable proto-meanings: raw inputs to those computations of meaning that warrant the curious properties of NRs. Then an evaluation map $|\cdot| := |\cdot|_U$ should map expressions to use-conditions of some kind. What kind? Well, the kind that will (i) account for the NR data, and (ii) will do so in a rationalizable way. How to specify them? Most obviously as conditions on the ostensible context of use. This leads to felicity conditions of which Searle (1969) and its descendants have given familiar examples, albeit without so far explaining things that call for explanation with some urgency.

An alternative, it seems, is to propose a dynamic or kinematic interpretation. This is in line with common intuitions about acts: after all, an act changes the course of the world in some respect. It is also in line with current epistemic re-construals of truth-conditional meaning in terms of changes of context brought about by an utterance. Typically, the meaning of an indicative sentence Φ is seen as a partial function on epistemic states. (Undefined e.g. for any context i at which a presupposition of Φ is unsatisfied.)

In its most basic versions this view abstracts away from difference of speech act type such as between *assertion* and *admission*, although proposals such as van der Sandt's (1991) introduce a distinction between denial (erasure leading to non-commitment) and assertion of a negative. It is a moot point to what extent it would be useful to pack what is needed to remedy the situation into an extended presuppositional component.

The more important question concerns the goods to be packaged: what predictively useful tacit conventions might there be, governing tacit or explicit actions by participants *prior* to adoption of a proposition into the context? But who does the adopting?

The convention I propose, let us call it a Discourse Institution, is modelled on the situation explicated by two-person bargaining games (Nash 1953). Two dialogue participants, Ego (E) and Alter (A) have to decide whether a certain proposition Φ, proposed by one of them, is to become a joint commitment (JC) constraining their beliefs ('epistemic') or conduct ('deontic'). I.e. we have, technically, a decision to be made about adoption of a 'jointly randomized strategy', where randomization for present purposes only admits of the 'pure' cases.

Suppose E proposes, i.e. *claims* Φ ($\mathrm{CLA_E}\Phi$). Then A has two options: (i) admit, i.e. *concede* Φ ($\mathrm{CON_A}\Phi$), whereupon Φ becomes a JC; (ii) *deny* Φ ($\mathrm{DEN_A}\Phi$). If (ii), then E has two options: (iii) *retract* the claim ($\mathrm{RET_E}\Phi$); or (iv) *insist* ($\mathrm{INS_E}\Phi$) i.e., to simplify matters, repeat the *claim* ($\mathrm{CLA_E}\Phi$); and so on.

In the case of retraction two options for consequences for JC are readily conceivable: either that no JC regarding Φ is undertaken, or else that $\neg\Phi$ becomes JC. The latter option might be motivated as a constraint against frivolous claims, supposing that personal interests attach to adoption.

If we disregard changes in the object of the claim or in its backing, both the Joint Commitment States (JCS) and the Negotiation States can be represented unsurprisingly by a pair $(\mathcal{J}, N)$ of feedback-linked finite state automata.

In what follows I shall skip most of the representation steps. For reassurance, however, here is a summary of the reconstruction as given in (Merin 1994).

I there consider a paradigm of *elementary social acts* (ESAs): claim, concession, denial, and retraction—performed with respect to social commitments in the boulomaic- deontic ('demand') and the epistemic-alethic ('assertion') domains. Completed transactions are represented as strings ('sentences') of a regular formal *transaction description language* over a vocabulary of 'words' such as $CLA_E\neg\Phi$. Examples are (where $\cdot$ is concatenation: $CLA_A\Phi \cdot DEN_E\Phi \cdot INS_A\Phi \cdot DEN_E\Phi \cdot INS_A\Phi \cdot DEN_E\Phi \cdot RET_A\Phi$; $CLA_E\neg\Phi \cdot CON_A\neg\Phi$. Non-examples: $CLA_E\Phi$; $CLA_A\neg\Phi \cdot RET_A\neg\Phi$; $DEN_E\Phi \cdot RET_E\Phi$. The words are, in turn, generated by a vocabulary of constituent 'morphemes' such as CLA, $_E, \neg$, Φ.

Sentences of a regular language are 'accepted' by a finite state automaton, $\mathcal{N}$. Our example above involved four states, plus a starting state. Multiply those four by 2 (propositional polarities) and again by 2 (person indices) and there are 16 states, in addition to the starting state.

$\mathcal{N}$ has output which drives transitions of an automaton $\mathcal{J}$ whose 4 states represent joint commitment to one of $\Phi, \neg\Phi$, or the lack of it, or conflict.[11] A transition to conflict state (however occasioned) emits a symbol which triggers an appropriate claim in $\mathcal{N}$. States of $\mathcal{N}$ reached

[11]The number 4 is familiar from Belnap's (1977) considerations on types of information states.

by concession or retraction afford transitions back to the neutral starting state, on emission of a symbol which moves $\mathcal{J}$ into an appropriate commitment state. This division of labour among automata is a banal bookkeping device which corresponds more or less to a division between Austin's 'illocutionary' and, as Davis (1979) showed, sometimes heavily conventional 'perlocutionary' aspects of communicative action.

A procedural semantics interprets fully specified ESAs as partial functions on the state set of $\mathcal{N}$ such that there is a bijection between them and their target states. Thus, acts denoted by expressions of our transaction description language, such as $CLA_E\neg\Phi$, are *constant partial functions* (*cpf*s) here symbolized by upright characters: $|CLA_E\neg\Phi|_U = \mathrm{CLA_E}\neg\Phi$.

A semantics for states, which interprets them as 'contexts' representable by vectors of binary-valued decision-theoretic parameters, allows us to define a set of coordinate-wise context transformations which form a group under function composition.

Act-*types* are defined as equivalence classes of acts, and thus identified with sets of states; a particular act-typology thus corresponds to a partition of the state set. A compositional semantics for recursion on the 'morphology' by a semigroup homomorphism is specified (example: $|CLA \cdot_E \cdot\Phi| = |CLA| \cap |_E| \cap |\Phi|$ where $\cdot$ is again concatenation). Transformations of states can then be lifted to transformations of act-types for suitable typologies, again forming a group under composition.

Denotata for natural language expressions with direct context-changing potential could now, in principle, be found in three domains: (i) the original *cpf*s (corresponding to explicit performatives); (ii) more abstract context transformations or partial restrictions thereof; (iii) transformations of act types operating, say, on a default act type or else anaphorically on an act-type imputed to the interlocutor for a prior utterance. It is option (iii) that I shall attend to almost exclusively in what follows.[12]

[12]See Merin (1994) for more on (i) and (ii). In Merin (1994a) I also characterize relations between ESA types and modal operators by an iso-functor between (I) a category whose objects are act-types (interpreting elements of the 'morphology' mentioned above) and whose morphisms are salient transformations thereof, and (II) a category with objects deontic modal operators and morphisms internal and external negation. Some of this will be summarized below. In the 1994a paper Anderson's (1956) construal of deontic necessity in terms of consequences interpretable as 'sanctions' provides a substantive link between the two categories. One moderately unexpected result discussed in 1994a is that the isomorphism does not extend to one which maps order relations on acts defined in terms of preferences to preference relations among objects compatible with interpretation in terms of entailment familiar from the 'Aristotelian' rectangle. This has a number of empirical consequences for linguistic analysis which I hope to explore elsewhere.

Time now for zooming down to some detail.

3.1. **A Paradigm of Elementary Social Acts.** States of $\mathcal{J}$ represent contexts of epistemic commitments, i.e. equivalence classes w.r.t. commitment or otherwise to Φ and its complement. (And we disregard here all issues of consistency with other propositions.) State transitions of $\mathcal{J}$ thus represent changes of joint epistemic commitments.

But what is a semantics for states and transitions of $\mathcal{N}$? The proposal is: Negotiation steps as outlined above are state-transitions; states represent negotiation contexts with respect to a single issue given by Φ whose essential features are characterized by vectors (5-tuples or shorter) of binary parameters.

For the paradigmatic 2-by-2 issue (2 persons, 2 complementary propositions) each context c is characterized by a 5-tuple $\langle P_S(c), P_O(c), P_P(c), P_D(c), P_I(c)\rangle$ of binary *parameter* values. Their intuitive interpretations are:

- $P_S(c) \in \{E, A\}$: '*acting subject*', (self)-imputed with the act;
- $P_O(c) \in \{\Phi, \neg\Phi\}$: '*issue orientation*', the propositional object of the initial claim;
- $P_P(c) \in \{E, A\}$: '*preference*', who ostensibly and *ceteris paribus* prefers adoption of Φ or, more generally, that proposition $\Theta \in \{\Phi, \neg\Phi\}$ whose natural language equivalent is unnegated. Preferences are strict, and E's are converse to A's.
- $P_D(c) \in \{E, A\}$: '*dominance*', who (ostensibly) dominates, i.e. here whose preference ostensibly decides the issue. Dominance means warrant, most basically: overwhelming incentives in a take-it-or-suffer-the-consequences *ultimatum game*
- $P_I(c) \in \{E, A\}$: '*initiative*', who ostensibly initiatiated the transaction in progress by a claim.

Strictness and duality of preferences factors out indifference and unanimity—i.e. those things which make for a non-issue. With obvious symbols,

$$(20) \qquad (\Phi \succ_E \neg\Phi \wedge \neg\Phi \succ_A \Phi) \vee (\Phi \succ_A \neg\Phi \wedge \neg\Phi \succ_E \Phi).$$

The question: 'Why should I ({do/believe} that)?' is always implicit, 'pointed', and openly so by un-deletion of your favourite deep-down expletive.

Our convention on preferences—for the 'positively' expressed item of an issue pair rather than for the actual propositional object—looks a bit roundabout. (If you look at Figure 1 it would be far easier to make sense of things if preference was always assigned with respect to the embedded propositional literal, i.e. even if negated.) But it is not unnatural for linguists surrounded by markedness phenomena. And it has the great advantage, to be used later, of turning $P_O(c)$ into redundant information. Moreover, there is a phenomenon consistent with it: the embedded negation constraint (ENC) which militates against lower clause negation.

Given an extra starting state, the four main act types will give us acts such as $\text{CLA}_E\Phi$ defined as *constant partial functions* (*cpf*s) on the state set of $\mathcal{N}$. Partial because they are defined only for states that represent proper points of departure in the transaction sequence. Constant because as things are defined they have unique target states, whatever the point of departure. Thus we have a bijection between acts and their target states. I propose the following semantics for acts: They bring about contexts whose relevant properties are given by Figure 1:

	P_S	P_O	P_P	P_D	P_I		P_S	P_O	P_P	P_D	P_I
$\text{CLA}_E\Phi$	E	Φ	E	E	E	$\text{CLA}_A\Phi$	A	Φ	A	A	A
$\text{CON}_E\Phi$	E	Φ	A	A	A	$\text{CON}_A\Phi$	A	Φ	E	E	E
$\text{DEN}_E\Phi$	E	Φ	A	E	A	$\text{DEN}_A\Phi$	A	Φ	E	A	E
$\text{RET}_E\Phi$	E	Φ	E	A	E	$\text{RET}_A\Phi$	A	Φ	A	E	A
$\text{CLA}_E\neg\Phi$	E	$\neg\Phi$	A	E	E	$\text{CLA}_A\neg\Phi$	A	$\neg\Phi$	E	A	A
$\text{CON}_E\neg\Phi$	E	$\neg\Phi$	E	A	A	$\text{CON}_A\neg\Phi$	A	$\neg\Phi$	A	E	E
$\text{DEN}_E\neg\Phi$	E	$\neg\Phi$	E	E	A	$\text{DEN}_A\neg\Phi$	A	$\neg\Phi$	A	A	E
$\text{RET}_E\neg\Phi$	E	$\neg\Phi$	A	A	E	$\text{RET}_A\neg\Phi$	A	$\neg\Phi$	E	E	A

Fig. 1: Elementary Social Acts and associated Contexts

Example: if Ego retracts Φ ($\text{RET}_\text{E}\Phi$) the ostension (from Ego's point of view) is among other things that Ego *ceteris paribus* prefers Φ, that Ego took the initiative along those lines, and that Alter dominates (hence that Alter's preference for $\neg\Phi$ will turn into a JC.)

Five binary parameters distinguish 32 contexts. Halving the set is a link between preference and initiative: agents do not launch counterpreferential initiatives.

4. Act Types and Transformations

An ESA such as $\text{CLA}_E\Phi$ is an act: a finest act type within the resources of our specification. Thus a generic claim CLA should be called simply an *act type*. The bijection from ESAs to the set $[c^\diamond]$ of contexts associates possible act-types with subsets τ of $[c^\diamond]$. A partition π of $[c^\diamond]$ is then a *typology.* A π of some interest is ACT $= \pi_1 = \{$CLA,CON,DEN,RET$\}$ where

(21) a. $\mathrm{CLA} = \{c \in [c^\diamond] : P_S(c) = P_D(c) = P_I(c)\}$

b. $\mathrm{CON} = \{c \in [c^\diamond] : P_S(c) \neq P_D(c) = P_I(c)\}$

c. $\mathrm{DEN} = \{c \in [c^\diamond] : P_S(c) = P_D(c) \neq P_I(c)\}$

d. $\mathrm{RET} = \{c \in [c^\diamond] : P_S(c) \neq P_D(c) \neq P_I(c)\}$

Let now $[c]$ be a set of 32 contexts extending $[c^\diamond]$ by those which do not satisfy the constraint linking propositional polarity, preference and initiative. Let the set $Trans := \{f_X : X \in \{S, O, P, D, I\}\}$ of bijections $f_X : [c] \to [c]$ be such that

(22) $P_X(f_X(c)) \neq P_X(c) \wedge [P_Y \neq P_X \Rightarrow P_Y(f_X(c)) = P_Y(c))]$.

Intuitively, f_X changes c into a context $c' = f_X(c)$ at which P_X assumes a different value, but which is in all other pertinent respects like c. Since all P_X are binary-valued, $(f_X(f_X(c)) = c$.

Transformations permit formal specification of the intuitive constraint linking preference and initiative:

(23) $P_P(c) = P_I(c) \Leftrightarrow P_P(f_O(c)) \neq P_I(f_O(c))$.

The closure T^* of *Trans* under composition is the abelian *group of context transformations* on $[c]$ generated by *Trans*. T^*. Let now

(24) $T^{\diamond *} := (f_{X'}|[c^\diamond]) = (f_{X'} : X' = S, OP, D, OI)$

be the group generated by $Trans^\diamond := \{f_{X'}|[c^\diamond] : f_X \in Trans\}$, the set of restrictions to $[c^\diamond]$ of the $f_{X'}$. (The set $[c^\diamond]$ is not closed under application of f_O, f_P or f_I: the constraint linking the respective parameters, which defines the subset $[c^\diamond]$ of the 2^5 contexts discriminable by 5 binary parameters is reflected in the intuitively composite nature of the generating elements of the group $T^{\diamond *}$ of transformations defined on it.) $T^{\diamond *}$ extends to a proper subgroup of $T^* := (f_X : X = O, S, P, D, I)$.

The next step to consider is the one from acts to act types. A useful act-typology π_i of $[c^\diamond]$ will have the *substitution property* (S.P.) , i.e. that for all $c, c' \in [c^\diamond]$, and $f = f_X \in T^{\diamond *}$

(25) $c \equiv c' \pmod{\pi_i} \quad \Rightarrow \quad f(c) \equiv f(c') \pmod{\pi_i}$.

Suppose $\tau \subseteq [c^\diamond]$ and $\tau \in \pi_i$. Any $c, c' \in \tau$ are then equal modulo the equivalence relation inducing π_i. If π_i has S.P., any $f \in T^{\diamond *}$ will map c and c' to elements of a single $\tau' \in \pi_i$ (where $\tau \neq \tau'$ or $\tau = \tau'$, depending on f). Thus, f maps 'blocks' of π_i to 'blocks' of π_i. Now for any S.P.

partition π_i of $[c^\diamond]$ the *difference between two act-types* $\tau, \tau' \in \pi_i$ will be definable as a unique element $f_{\tau-\tau'} = f_{\tau'-\tau} \in T^{\diamond *}$. We can now define a group $TT^{\diamond *}$ of *transformations* $\mathbf{f}_X$ on *act-types* (of whatever granularity) by

(26) $\mathbf{f}_X(\tau) := \{f_X(c) : c \in \tau\}$.

Examples: Let $\Theta \in \{\Phi, \neg\Phi\}$ and $\neg\neg\Theta = \Theta$; and let $\mathrm{ACT} \in \{\mathrm{CLA}, \mathrm{CON}, \mathrm{DEN}, \mathrm{RET}\}$; $H, H', K, K' \in \{E, A\}$ where $H \neq K$, $H' \neq K'$. Writing $f_{XY}(c)$ for $f_X(f_Y(c))$ and $\mathbf{f}_X : Type1 \rightleftharpoons Type2$ for $\mathbf{f}_X(Type1) = Type2 \wedge \mathbf{f}_X^{-1}(Type2) \equiv \mathbf{f}_X(Type2) = Type1$:

(27) $\mathbf{f}_{OP}: \quad \mathrm{ACT}\Theta \rightleftharpoons \mathrm{ACT}\neg\Theta;$

(28) $\mathbf{f}_D: \quad \mathrm{CLA} \rightleftharpoons \mathrm{RET};\ \mathrm{CON} \rightleftharpoons \mathrm{DEN};$

(29) $\mathbf{f}_{PI}: \quad \mathrm{CLA} \rightleftharpoons \mathrm{DEN};\ \mathrm{CON} \rightleftharpoons \mathrm{RET};$

(30) $\mathbf{f}_{PDI}: \quad \mathrm{CLA} \rightleftharpoons \mathrm{CON};\ \mathrm{DEN} \rightleftharpoons \mathrm{RET};$

(31) $\mathbf{f}_S: \quad \mathrm{CLA}_H\Theta \rightleftharpoons \mathrm{CON}_K\Theta;\ \mathrm{DEN}_{H'}\Theta \rightleftharpoons \mathrm{RET}_{K'}\Theta;$

(32) $\mathbf{f}_{SD}: \quad \mathrm{CLA}_H\Theta \rightleftharpoons \mathrm{DEN}_K\Theta;\ \mathrm{CON}_{H'}\Theta \rightleftharpoons \mathrm{RET}_{K'}\Theta.$

(33) $\mathbf{f}_{SPDI}: \quad \mathrm{ACT}_H\Theta \rightleftharpoons \mathrm{ACT}_K\Theta.$

Application of $\mathbf{f}_{SPDI}$ corresponds to changing the acting subject while keeping act-type and object constant: conditions for imputation must change too, then. $\mathbf{f}_S$ intuitively corresponds to a *change in point of view* on the same ostensibly objective conditions. By contrast any two of $\mathbf{f}_D$, $\mathbf{f}_{PI}$, $\mathbf{f}_{PDI}$ generate $\mathrm{ACT}_H\Theta$ from any one of its elements, for constant H, Θ.

We can now eliminate the 'subject' parameter, writing '+' for $P_X(c) = P_S(c)$ and '–' for $P_X(c) \neq P_S(c)$, where $X \in \{P, D, I\}$). That is: for '+' read 'self / speaker / agent is the one who prefers Φ (or is dominant, or initiated with a claim)'; for '$-$' read 'the other does (or is, or did)'.

Note that our representational convention already makes $P_O(c)$ redundant information: it correlates *negation* of the propositional object with a change of *preference* polarity. And it is precisely with those changes in mind, i.e. under a 'kinematic' point of view, that all clumsiness of specification that made Figure 1 a less than easy read disappears.

Act types of grain $\mathrm{ACT}\Theta$ can now be specified by (in)equations homogeneous in person-valued parameters (Figure 2). In fact: given the constraints of the discourse institution we have *represented* negation by an operation which amounts to conversion of preference. Thus, setting $\Theta := \Phi$,

	P	D	I		P	D	I
CLAΦ	$+$	$+$	$+$	CLA$\neg\Phi$	$-$	$+$	$+$
CONΦ	$-$	$-$	$-$	CON$\neg\Phi$	$+$	$-$	$-$
DENΦ	$-$	$+$	$-$	DEN$\neg\Phi$	$+$	$+$	$-$
RETΦ	$+$	$-$	$+$	RET$\neg\Phi$	$-$	$-$	$+$

Fig. 2: Act Types

Example: RET$\neg\Phi$, a retraction of a claim of $\neg\Phi$, has initiative assigned to the actor, but preference for Φ and dominance to the other.

Intution suggests that one obtain CLA$\neg\Phi$ by transformation $\mathbf{f}_O$ ('negation of content') of CLAΦ. But in view of the parameter definitions reflected in Figures 1 and 2, we exploit the 'markedness' convention (call it M), which assigns preference with respect to the unnegated proposition of an issue pair, to obtain it by $\mathbf{f}_P$. I.e., CLA$\neg\Phi =_M \mathbf{f}_P$(CLA$\Phi$); DEN$\Phi =_M \mathbf{f}_{PI}$(CLA$\Phi$), hence DEN$\Phi =_M \mathbf{f}_I$(CLA$\neg\Phi$).

Suppose next that we factor out initiative by projection along the P_I axis of our parameter space, setting ACT$\Phi \simeq \mathbf{f}_I$(ACTΦ). This yields the Fregean (1918) identification,

(34) CLA$\neg\Phi \simeq$ DENΦ

along with RET$\neg\Phi \simeq$ CONΦ, DEN$\neg\Phi \simeq$ CLAΦ, and CON$\neg\Phi \simeq$ RETΦ.

We can thus eliminate negation and hence the propositional object, which is understood from the global context (the issue). What is left is the foursome of pure act types, strongly reminiscent of a modal 'Aristotelian' rectangle.

The analogy can be made precise. A functor $G = (G_{Obj}, G_{Mor})$ relates the two categories of acts and modals isomorphically. G takes as its argument a small, finite category (in the mathematician's sense) (call it *Civil*) with objects {CLA,CON,DEN,RET} and an abelian group of morphisms generated by $\{\mathbf{f}_P, \mathbf{f}_D\}$. Objects, morphisms, and their action on objects are represented in the left-hand panel of Figure 3, where a '–' in the column $X \in D, P$ represents application of operator $\mathbf{f}_X$ to CLA, considered as the distinguished generating element of the set of operands. The value, $G(Civil)$, is another category, call it *Modals*, whose objects are the modal operators with internal and external negations and whose group of morphisms is generated by external (η; $\neg_{ex}$) and internal (ι; $\neg_{in}$)

negation. (In Figure 3 the previous order of columns P and D has been switched around to conform visually with the standard syntactic icons for these operations.)

G_{obj} maps CLA to NECessary (NEC), CON to POSsible (¬NEC¬), DEN to IMPossible (NEC¬), and RET to the fourth, UNNecessary (¬NEC). The respective deontic instances of these, which make the mapping more intuitive yet, are OBLigatory, PERmitted, PROhibited, and (let's call it with John Lyons) EXEmpted.

Now, whatever interpretive instances of modality we choose (deontic, epistemic, alethetic), given our choice of G_{Obj}, functoriality of G demands a unique correlation: 1. *preference reversal*, $\mathbf{f}_P$, with *internal negation*; 2. *dominance reversal*, $\mathbf{f}_D$, with *external negation*; 3. their *composition*, $\mathbf{f}_{PD}$, with *dualization*, δ. I.e., to be a functor (a structure-preserving map) G must have a morphism map

(35) $\quad G_{\mathrm{Mor}}: \quad \mathbf{f}_P \mapsto \neg_{in};\ \mathbf{f}_D \mapsto \neg_{ex};\ \mathbf{f}_{PD} \mapsto \delta.$

	D	P		η	ι
CLA	+	+	NEC	+	+
CON	−	−	POS	−	−
DEN	+	−	IMP	+	−
RET	−	+	UNN	−	+

Fig. 3: Parameter Transformations in $Civil$ and $Modals$

Of course, the association of $\mathbf{f}_P$ with $\neg_{in}$ is made by definition. But note that $\mathbf{f}_D$ is not simply 'negation', i.e. Boolean complementation, of acts. ¿From the fact that retractions, or rather pre-emptive 'disclaimers', on the one hand, and denials on the other, are often expressed by way of potentially performative act-descriptions as 'I don't {claim/concede} that Φ' we cannot simply derive a transformation of 'force' induced by negation.[13]

[13] At least no intelligible account of such a transformation has been offered prior to the present proposal. The 'illocutionary logic' of which Searle and Vanderveken (1985) purport to give the foundations with a set of 'illocutionary connectives' including conjunction, negation, etc. appears to be a logic at best in a Pickwickian sense; and the unexplained absence of any formal reference in Vanderveken (1991) to what, if anything, ought to make it a logic, i.e. its connective structure, suggests that this judgment no longer needs much arguing for.

5. 'Scales' and Transformations Explicated

Given the mapping in (35) and Figure 3, the equality CON $= \mathbf{f}_{PD}$(CLA) warrants calling claim and concession *dual act types*. My proposal is that the 'positive' scales elicited by Horn are bounded (as it were) by items which, under the use-conditional interpretation, denote instances of these two ESAs. This act-interpretation is not wholly surprising in view of the relative ease of NR in first-person-present (FPP) occurrence. (If a predicate neg-raises at all, it will do so in FPP usage, and sometimes only so. Example for some: *suggest.*) And I believe that in third or second person contexts a case can be made for imputation of the appropriate instances of acts to third parties or to the addressee.

Suppose this could be worked out in detail. Does it have any direct empirical consequences? Yes. For a start, it gives an explanation for the effect of 'external' application of the morpheme *not* (Horn's 'contradictory negation') to non-mid-scalar items. Recall the intuitions concerning 'strong' and 'weak' elements of Pos and Neg. Given the obvious instantiations, we have CLA strong in Pos (s^+), CON weak in Pos (w^+), DEN strong in Neg (s^-), and RET weak in Neg (w^-).

If 'external' occurrence of the negation morpheme denotes a transformation $\mathbf{f}_D$ of act type then we get just the required mappings, since

(36) $$\mathbf{f}_D(\text{CLA}) = \text{RET}; \quad \mathbf{f}_D(\text{CON}) = \text{DEN}$$

and, if you will, vice versa (cp. *not impossible*). I.e. we have, loosely stated,

(37) $$\mathbf{f}_D : s^+ \mapsto w^-;\ w^+ \mapsto s^-.$$

What have we gained? For a start, an interpretation for the notion 'transformation of an act type'. For clearly, just noting that scales are bounded by cousins of modal operators does not in itself supply one. Secondly, this interpretation does not explicate the intuitions about 'weak' and 'strong' elements through appeal to unilateral entailment, which we have seen to be less than reliable.

In fact, it explains the intuitions in terms which make the metapragmatic glosses supplied by Horn quite unmetaphoric. 'Strong' means: being ostensibly in a position of strength with respect to the issue at hand. 'Weak' means being ostensibly in a position of weakness. Thirdly, there is no need for doubtful arithmetic in specifying the transformations. But, one may ask: will mid-scalars fall out from this scheme?

6. Quotient Types: First Approximations to Suggestion and Doubt

Consider a coarser partition π_5 of $[c^\diamond]$ that conflates claim with concession, and denial with retraction. Let nonce labels SUG and DUB designate the act-types represented by its elements:

(38) $$\mathrm{SUG} := \mathrm{CLA} \cup \mathrm{CON} = \{c \in [c^\diamond] : P_D(c) = P_I(c)\}$$

(39) $$\mathrm{DUB} := \mathrm{DEN} \cup \mathrm{RET} = \{c \in [c^\diamond] : P_D(c) \neq P_I(c)\}.$$

Once we have factored out initiative, the 'Fregean' identification (34) makes this tantamount to the more intuitive (and primary) constraint

(40) $$\mathrm{SUG} := \mathrm{CLA} \cup \mathrm{CON} = \{c \in [c^\odot] : P_D(c) = P_P(c)\}$$

(41) $$\mathrm{DUB} := \mathrm{DEN} \cup \mathrm{RET} = \{c \in [c^\odot] : P_D(c) \neq P_P(c)\}.$$

where $[c^\odot]$ is the appropriate set of more coarsely individuated quotient contexts.

Extensionally, $\pi_5 = \{\mathrm{SUG}, \mathrm{DUB}\}$ is then a *quotient typology* on ACT which yields, in a sense open to immediate inspection and assessment for adequacy, 'neutral', 'attenuated' or 'uncertain' acts. The set-valued denotation leaves undefined whose preference (*cum* initiative) and dominance characterize the context ostended by imputation of $\mathrm{SUG_E}\Phi$. Ego is still the acting subject, but the important decision-theoretic parameters are left vague. Likewise for other acts.[14]

The nonce labels are, of course, intended to suggest *suggestion* and *doubt*. I hedge: quotient-type acts defined in Discourse Institutions that permit P_O, P_P and P_I to vary independently might in certain respects do rather better as explicata of the familiar natural act types. But as a first approximation, SUG and DUB will get us closer than do CLA and DEN to the discursive neutrality preferred by philosophers and polite conversationists.

A look at the left-hand side of Figure 3 will immediately get us the action of $\mathbf{f}_D$ and $\mathbf{f}_P$ on SUG and DUB. Conceptualizing these as (plain,

[14] And since elements of S.P. partitions are, in a straightforward way, the pre-images of elements making up the range of a corresponding homomorphism $h^\diamond$, we could of course write, e.g.

$$\mathrm{SUG_E}\Phi = h^\diamond(\{[\mathrm{E}\Phi\mathrm{EEE}], [\mathrm{E}\Phi\mathrm{AAA}]\})$$

and construct a distinct state-homomorphic automaton with half the number of target-context-denoting states.

unordered) sets of ESAs on which $\mathbf{f}_D$ and $\mathbf{f}_P$ operate pointwise (recalling their definition makes this quite unabusive) yields the following identity:

(42) $$\begin{aligned} &\mathbf{f}_D(\text{SUG}) = \mathbf{f}_D(\{\text{CLA}, \text{CON}\}) \\ &= \{\mathbf{f}_D(\text{CLA}), \mathbf{f}_D(\text{CON})\} \\ &= \{\text{RET}, \text{DEN}\} = DUB = \{\text{DEN}, \text{RET}\} \\ &= \{\mathbf{f}_P(\text{CLA}), \mathbf{f}_P(\text{CON})\} \\ &= \mathbf{f}_P(\text{CLA}, \text{CON}) = \mathbf{f}_P(\text{SUG}). \end{aligned}$$

Note that $\mathbf{f}_D$, designated by 'external' *not*, will act in accordance with Horn's observation (b). For just as SUG is mid-scalar on a Pos scale (situated 'between' CLA and CON), so DUB is mid-scalar on the corresponding Neg scale (situated between DEN and RET).

The scare-quotes round 'between' can in fact go. For one way of interpreting these derived acts is as convex *linear superpositions* of functions, represented by superposition of states represented by parameter vectors. And the semantics of *between* in geometry (and, I should say, in English: cp. Merin 1992) is explicated by convex linear combination (a.k.a. weighted average.) Example: let there be a real number $k \in]0, 1[$ and let

(43) $$\text{SUG}_\text{E}\Phi = k[\text{E}\Phi\text{EEE}] + (1-k)[\text{E}\Phi\text{AAA}]$$

For $k = 1/2$ our intuitions should be indistinguishable from those attaching to the plain set-representation. The 'smear' or 'mixture' of parameter values is nicely brought out by a nonce vector symbol: $\text{SUG}_\text{E}\Phi$ =[EΦÆÆÆ]. What does 'mixture' mean?. Well, k might be thought of as a probability. This has the added advantage of letting SUG be explicated as a *disposition* which, given a certain state of moral affairs, might either lead to an act of claiming or to one of conceding on the agent's part.[15] Ignorance about which state, if any, is likely to obtain will then reflect in the probability distribution. Note that this affords a decision-theoretic explication to the 'mixed' act: namely as a as a strategy in an ostensible game of *incomplete information* where players may be ignorant of their own and others' preferences and resources. But on might also come up with intuitions supporting non-probabilistic mixture.

If we deviate from the default $k = 1/2$ we can weight the mixture to take account of, say, an assertive bias. If anything, neg-raisers tend to swerve closer to the claim than to the concession type of act. And of

[15]Urmson (1952) objects to dispositional accounts of *I believe*, suggesting, it seems, that analysis as 'parenthetical' obviates the rationale for such construals. But to the extent that the structure, rather than the evidentiality- or attitude-indicating function of parenthetical usage needs explaining I remain unconvinced. We do not, incidentally, have a disposition to extra-social conduct here, as is usually included in dispositional accounts of belief.

course we get predictions consistent with Horn's observation about external negation. If $1/2 < k << 1$, say 0.6, and our transformations operate linearly as demanded by pointwise application, the analysis predicts that

(44) $$\begin{aligned} &\mathbf{f}_D(\text{SUG}) = \mathbf{f}_D(k[\text{CLA}]) + \mathbf{f}_D((1-k)[\text{CON}]) \\ &= k\mathbf{f}_D([\text{CLA}]) + (1-k)\mathbf{f}_D([\text{CON}]) \\ &= k[\text{RET}]) + (1-k)([\text{DEN}]) \end{aligned}$$

will give us a mixture of 0.6 RET and 0.4 DEN, i.e. an act type DUB which is closer to retraction than to denial, tending to weak rather than to strong 'negative'. By contrast, application of $\mathbf{f}_P$ would now yield an act type tending more to strong than to weak 'negative'. Affixal negation by way of *un*$-$ (as in *unlikely* would be an instance, 'internal' by Horn's 1978 account.

7. Conclusion

The objective of the argumentation was to outline the rudiments of an essentially Austinian semantics for certain epistemic and deontic complement-constructions in English. Not simply as an exercise in filling in a certain philosophical position which, in its most famous application—the phenomenon of explicit performatives—is anything but compelling in the strong form proposed by Austin. (After all, in the light of Lemmon (1962) it became difficult to continue insisting that explicit performatives did not perform partly in virtue of their primary performative aspect of stating what they were intended to perform, as indeed suggested by their indicative mood.)

The present analysis of NR is motivated largely by a desire to account for the composition of meaning in natural languages.[16] I have not given an explicit 'fragment' here: but it would in all essential aspects be no more complex or subtle than a straightforward translation from the formal transaction description language. I.e.

(45) I don't believe that the earth is flat

would come out as an expression such as $dn'tSUG_i\Phi$ where Φ = 'the earth is flat', $|dn't| = \mathbf{f}_D$, and SUG_i has its interpretation given above. We get equivalence to $DUB_i\Phi$. From there to

[16] See Lewis (1979), Kamp (1979) and Merin (1991, 1992) for similar motivation, centering around disjunctive permissions. These approaches also broach the question of directly act-denoting expressions and their composition.

(46) I doubt that the earth is flat

it is still some way through the lexicology of antonyms. But given the emotive concomitants of such distinctions (and of 'negative' affixes) we are closer to an account of those than if we had to rely exclusively on truth-conditional negation and its presumed action sites.

There are two *prima facie* reasons for considering act-denotations: (i) a large proportion of the complement-taking expressions in question appear to deviate systematically from their original descriptive meanings when licensing NR; (ii) the structure of the paradigms involved which bears a 'static' modal interpretation at best by analogy and which would otherwise remain a brutish regularity.

The proposed semantics relies heavily on a Discourse Institution (DI) with its paradigm of Elementary Social Acts. If we speak of a semantics, we must ask about the status of the DI. The best I can do at present is to say that it is a socio-psychological representation, part of people's language or language-related faculties, of a kind similar to or extending the Discourse Representations introduced by Kamp (1981) with particular attention to trans-sentential anaphora.

The assumption then is that intuitions of paraphrase and acceptability are conditioned by such representations, and that the NR expressions in question denote transformations of instances of such representations. In our case the relevant parameters concern the ostensible evidential and preferential relations between speaker, addressee and propositional content. A simplistic view of things, no doubt. But simple enough to inform computations so routine that their empirical reflexes are neither obvious to nor call for much introspection.

Perhaps I should end with an empirical point barely broached so far: the ENC. Consider the markedness convention M on the representation of negation, which was instrumental in explaining NR. If we assume that (a) M informs spontaneous native speaker intuitions on acceptability and paraphrase and (b) difficulties of the analyst in making sense of preference assignments reflect, in this instance, a naturalness constraint[17] then precisely those natural language instances of sentences representing the acts in Figure 1 should be 'marked' in which there is an (internally) negated complement. Meaning that there should be a premium on avoiding them.

[17]Namely: that preferences should be read off directly from the embedded clause; i.e. that there should not be a need for an extra inferential or computation step.

I hesitate to call this an explanation of the ENC, subtle as it is. But it looks like being a useful part of one, rather than merely an ad hoc device to translate truth-conditional negation into an operation of conversion on preferences.[18]

The rationale for the proposed account are linguistic data. Its implications will, I believe, retain some philosophical interest.

[18]In Merin (1994) I sketch how this could retain the elegance of Ladusaw's (1979) largely semantic treatment of Negative Polarity Items *ever, at all, ...* in the face of examples that have prompted rather less elegant 'implicatural' proposals (Linebarger 1987) with arbitrarily large explanatory power.

References

Austin, J.L. (1946) Other minds. *Proc. Arist. Soc. Suppl. Vol XX.* 148-187. Repr. in Austin (1961).

—— (1961) *Philosophical papers*. Oxford: Clarendon Press.

—— (1962) *How to do things with words*. Oxford: Clarendon Press.

Anderson, A.R. (1956) The formal analysis of normative systems. Repr. in N. Rescher (ed.) *The logic of decision and action*. Pittsburgh: Pittsburgh University Press, 1967.

Belnap, N.D. (1977) A useful four-valued logic. In G. Epstein and M. Dunn (eds.) *Modern uses of multiple-valued logic* Dordrecht: Reidel.

Crossley, J.N. and Humberstone, L. (1977) The logic of 'actually'. *Reports on Mathematical Logic* 8, 11-29.

Davis, S. (1979) Perlocutions. *Linguistics and Philosophy* 3, 245-266.

Frege, G. (1918) Logische Untersuchungen: Die Verneinung, *Beiträge zur Philosophie des deutschen Idealismus* I, 143– 157.

Gazdar, G. (1979) *Pragmatics*. New York: Academic Press.

Gazdar, G. and Klein, E. (1980) [Review of Horn 1978 etc] *Journal of Pragmatics* 4, 95-103.

Hintikka, J. (1962) *Knowledge and belief*. Ithaca NY: Cornell University Press.

Horn, L.R. (1972) On the semantic properties of logical operators in English. Ph.D. Diss. UCLA. Repr: Bloomington: Indiana University Linguistics Club 1976.

—— (1978) Remarks on Neg-Raising. In P. Cole (ed.) *Syntax and Semantics 9: Pragmatics.* New York: Academic Press.

—— (1978a) Some aspects of negation. In J. Greenberg, C. Ferguson, and E. Moravcsik (eds.) *Universals of human language* Vol 4,'. Stanford CA: Stanford University Press.

—— (1989) *A natural history of negation.* Chicago: University of Chicago press.

Kamp, H. (1971) Formal properties of 'now'. *Theoria* 37, 227-273.

—— (1979) Semantics versus pragmatics, in F. Guenthner and S.J. Schmidt (eds.), *Formal semantics and pragmatics for natural languages*. Dordrecht: Reidel.

—— (1981) A theory of truth and semantic representation. In J. Groenendijk, T. Janssen, and M. Stokhof (eds) *Formal methods in the study of language.* Amsterdam: Mathematical Centre Tracts; repr. Dordrecht: Foris 1984.

Ladusaw, W. (1979) Polarity sensitivity as inherent scope relations. Ph.D. diss. U. of Texas, Austin; repr. New York: Garland.

Leibniz, G.W. (1716) Correspondence with Clarke. In L.E. Loemker (ed.) *Leibniz: Philosophical papers and letters.* Dordrecht: Reidel, 1970.

Lemmon, E.J. (1962) Sentences verifiable by their use. *Analysis* 22, 86-89.

Lewis, D. (1979) A problem about permission. In E. Saarinen, R. Hilpinen, I. Niiniluoto and M. Provence-Hintikka (eds.), *Essays in honour of Jaakko Hintikka*. Dordrecht: Reidel.

Linebarger, M. (1987) Negative polarity and grammatical representation.*Linguistics and Philosophy* 10, 325-387.

Merin, A. (1991) Imperatives: linguistics vs. philosophy, *Linguistics* 29, 669–702.

—— (1992) Permission sentences stand in the way of boolean and other lattice-theoretic semantices, *Journal of Semantics* 9, 95–162.

—— (1994) Algebra of elementary social acts. In S.L. Tsohatzidis (ed.) *Foundations of speech act theory*, London: Routledge.

—— (1994a) From civil procedure to deontic modality. In A.J.I. Jones and M. Sergot (eds.) *Proc. 2nd Int. Workshop on Deontic Logic in Computer Science, Oslo January 5-8 1994*. Oslo: TANO.

Nash, J.F. (1953) Two-person cooperative games, *Econometrica* 21, 140–152.

Quine, W.V.O. (1960) *Word and object*. Cambridge MA: MIT Press.

Searle, J.R. (1969) *Speech acts.* Cambridge: Cambridge University Press.

Searle, J.R. and Vanderveken, D. (1985) *Foundations of Illocutionary Logic*. Cambridge: Cambridge University Press.

Urmson, J.O. (1952) Parenthetical verbs. *Mind* 61, 480-496.

van der Sandt, R. (1991) Denial. In *Parasession on Negation, 27th Regional Meeting, Chicago Linguistic Society.*

Vanderveken D. (1991) *Meaning and speech acts* Vol. 2. Cambridge: Cambridge University Press.

[This expands a paper delivered at the Workshop on Negation, 2nd International Conference of the (German) Society for Analytical Philosophy, Leipzig, September 1994.]

Institut für Maschinelle Sprachverarbeitung, Formale Logik und Sprachphilosophie, University of Stuttgart, Azenbergstr. 12, 70174 Stuttgart, Germany

E-mail address: arthur@ims.uni-stuttgart.de