4

Implying and inferring

Laurence R. Horn

> To draw inferences has been said to be the great business of life.
>
> John Stuart Mill, *Logic* (1843), Intro. §5

4.1 A "vulgar conflation"

In Bach's short manifesto unveiling his list of the top ten misconceptions about implicature (2006c: 23), #2 is the thesis "Implicatures are inferences." For Bach, such a claim – whether explicit (as in the subtitle of Levinson 2000, "Generalized conversational implicatures as default inferences") or implicit – is a "misdenomer" amounting to a "slight variation on the vulgar conflation of implying with inferring." The distinction in each case is seen as a straightforward one: implying (or, more specifically, implicating) is something the producer or sender of a message (speaker or writer) does, while inferring pertains to the cognitive effort of the receiver. Bach submits the entry from the *American Heritage Book of English Usage*:[1]

> When we say that a speaker or sentence implies something, we mean that information is conveyed or suggested without being stated outright . . . Inference, on the other hand, is the activity performed by a reader or interpreter in drawing conclusions that are not explicit in what is said.

The distinction is vital for pragmatic theory because an interpreter may "recover" an implication (presupposition, implicature) that was not intended by the utterer, and a speaker may imply (presuppose, implicate) something that the interpreter does not grasp. In I. A. Richards's words (MWDEU 1994: 541), "An utterer may imply things which his hearers cannot reasonably

I am grateful to Barbara Abbott, Keith Allan, Kent Bach, Betty Birner, Bart Geurts, Kasia Jaszczolt, Nicole Palffy-Muhoray, and Gregory Ward for their helpful comments and pointers; the usual disclaimers apply.

infer from what he says"; in other cases, the expectation of inference may be reasonable but nevertheless unfulfilled, whether through inattention or a mismatch of shared beliefs. But what of the straightforward distinction between implying and inferring itself?

Most usage manuals endorse the distinction, but a closer look shows that it's not that simple. *WDS* (1942), for example, begins with this prescription, complete with pointy finger:

> ☞Do not confuse *infer* with *imply*.

But then the rockier landscape of actual usage is surveyed (*WDS* 1942: 449–50):

> The use of *infer* in the sense of to hint at or to intimate (as, by his remarks he *infers* [correctly *implies*]) is still regarded as erroneous. However, in the past *infer* sometimes meant, and still to some extent means, to give grounds for believing (something that is stated) or to permit (something) to be inferred. In such use, a personal subject is to be avoided, for in precise English only that which gives the grounds for or permits an inference, or which leads to a given conclusion, can rightly be the subject; as "This doth *infer* the zeal I had to see him" (*Shak.*); "Consider first that great Or bright *infers* not excellence" (*Milton*); "Matters were by no means so far advanced between the young people as Henchard's jealous grief *inferred*" (*Hardy*).

So, in effect, "Don't use *infer* to mean *imply* – but if you do, make sure your subject is inanimate." The secure footing appears to have become a bit slippery, presumably because a misstep with impersonal subjects yields no ambiguity.

In *Merriam Webster's Dictionary of English Usage*, the linguistically best informed usage compendium, the extensive entry for **infer, imply** (MWDEU 1994: 541–4) disentangles three uses of *infer*, eloquently illustrating that "Real life is not as simple as commentators would like it to be" (541). Besides the universally accepted 'deduce, conclude' sense for *infer*, *MWDEU* differentiates the impersonal-subject *infer*-for-*imply* illustrated in the Shakespeare, Milton, and Hardy examples above[2] from the "personal *infer*" attested in an Ellen Terry letter from 1896: "I should think you *did* miss my letters. I know it! but . . . you missed them in another way than you infer, you little minx!" The *MWDEU* suggests that this latter use, the specific target of twentieth-century prescriptive opprobrium, has always been largely restricted to informal spoken language.

Similarly, the *OED*'s sense 4 for *infer* reads as follows:

> To lead to (something) as a conclusion; to involve as a consequence; to imply. (Said of a fact or statement; sometimes, of the person who makes the statement.)
>
> This use is widely considered to be incorrect, esp. with a person as the subject.

But even if incorrect, not inexistent; cites range from the sixteenth century to the nineteenth ("Socrates argued that a statue inferred the existence of a sculptor") and the twentieth ("I can't stand fellers who infer things about good clean-living Australian sheilahs").

It should be noted that the direct target of objections like that of Bach (or Horn 2004) is the nominal form (*inference*) rather than the verbal *(infer)*, as in Levinson's implicatures as default inferences. But the speaker-oriented sense here boasts its own distinguished pedigree: the relevant *OED* sense 2 for *inference* is glossed as

> That which is inferred, a conclusion drawn from data or premises. Also an implication; the conclusion that one is intended to draw. Cf. INFER *v.* 4.

– with no disparagement of the latter usage and with attestations back to a 1612 essay by Francis Bacon on judicial practice warning that "Judges must beware of hard constructions and strained inferences."

What does not seem to be noted in any of the dictionaries and manuals is the asymmetry exhibited in these purportedly erroneous uses. While *infer* has been used for almost five centuries in the sense of 'imply, convey' (whether with impersonal or personal subjects), *imply* is always sender-, not receiver-oriented; it is never used for 'infer', to refer to the cognitive processes of the hearer or reader. Although this asymmetry may be seen as betokening the primary status of the speaker or producer of the message, we can find other cases in which a predicate exhibits an analogous ambiguity but where the direction of the meaning shift is less clear.[3] Thus *entendre* exhibits a range of meanings in French from the speaker-oriented 'mean, intend' to the hearer-oriented 'understand, hear'. This is especially significant for pragmatic theory because, as we shall see, conversational implicature first appeared in Mill's invocation of the SOUS-ENTENDU, that which is literally under-meant or under-understood.

4.2 Implication and implicature

Since classical rhetoricians first described figures in which we say less and mean more (*minus dicimus et plus significamus*),[4] semanticists and pragmaticists have explored the boundaries between what is said and what is meant-but-not-said. The latter is the realm of the implied. Recognition of the distinction between the said and the (merely) implied is not, however, limited to philosophers, linguists, and rhetoricians. Consider this exchange between Elinor and Marianne Dashwood, Austen's eponymous Sense and Sensibility, respectively (1811: Chapter 29), concerning the reprehensible but not actionable misbehavior of the latter's erstwhile beau Willoughby:

> 'But he told you that he loved you?'
>
> 'Yes – no – never absolutely. It was every day implied, but never professedly declared. Sometimes I thought it had been, but it never was.'

This is a distinction vital to lawyers as well as cads, as seen in the myriad devices for exploiting the difference between what is said and what is implied under oath. An important precedent in this domain is *Bronston* v. *United States*. Samuel Bronston, president of a film production company, responded as follows to a cross-examining prosecutor in his 1973 trial (409 U.S. 352–354, cited in Solan and Tiersma 2005: 213):

Q: Do you have any bank accounts in Swiss banks?
A: No, sir.
Q: Have you ever?
A: The company had an account there for about six months, in Zurich.

In fact, besides the company account, Bronston had actively maintained a large personal account in a Swiss bank. Thus, while his first response was truthful (depending, as President Clinton might have said, on the meaning of the word *do*), his second answer was at the very least misleading or "non-responsive." But was it false? Bronston was convicted of perjury, on the grounds that his last response, while literally true, "falsely implied that he had never had a personal Swiss bank account," but the judgment was reversed by a unanimous US Supreme Court. The particulars of this and related cases are illuminated by Solan and Tiersma (2005: 212–35), who point out that Bronston's violation concerned what is implicated (via the quantity and relation maxims) rather than what is literally said and endorse the *Bronston* "literal truth" defense against perjury charges, whether for sleazy movie producers or jesuitical presidents.

The difference between lying (based on the falsity of what is said) and misleading (based on the falsity of what is implied), as instantiated above and in a variety of other fictional and all too real settings over the last two millennia from the Oval Office to everyday conversation, can be taken to support an orthodox Gricean conception of what is said that hugs the syntactic ground of the spoken or written sentence as opposed to an "inflationary" view that incorporates pragmatically derived aspects of the intended communication; see Horn 2009b for elaboration.

But not just any (non-logical) implication is an implicature. In particular, conversational implicature in the Gricean model typically arises from what the speaker didn't say but (given rationality and cooperation) would have been expected to say if she had been in a different epistemic position. This point, rightly associated with Grice's William James lectures, was actually made exactly a century earlier.

In the locus classicus, a speaker uttering *Some F are G* implies that (for all she knows) not all F are G because she would have been expected by the hearer to have expressed the stronger proposition if she had been in a position to do so. The key insight is provided in this passage in which John Stuart Mill rejects Sir William Hamilton's (1860) treatment of *some* as logically expressing 'some only, some but not all':

> No shadow of justification is shown . . . for adopting into logic a mere *sous-entendu* of common conversation in its most unprecise form. If I say to any one, "I saw some of your children today", he might be justified in inferring that I did not see them all, not because the words mean it, but because, if I had seen them all, it is most likely that I should have said so: though even this cannot be presumed unless it is presupposed that I must have known whether the children I saw were all or not. (Mill 1867: 501)

Mill invokes here the two-stage process allowing the hearer's move from the weaker recovered implication ('for all the speaker knows, not all . . . ') to the stronger ('the speaker knows that not all . . . ') when epistemically licensed. These are the primary and secondary implicatures of Sauerland (2004), built into the rationality-driven Gricean model (cf. Horn 1989, 2009b: §2; Geurts 2009) but not captured in current alternative grammatical theories of "blind mandatory scalar implicature" (Chierchia *et al*. forthcoming; Magri 2009; see Geurts 2010 for discussion).

Mill's allusion to a tacit principle requiring the speaker's choice of the stronger *all* over the weaker *some* when possible, and inviting the hearer to draw the corresponding inference when the stronger term is eschewed, is echoed by others in his own time –

> Whenever we think of the class as a whole, we should employ the term All; and therefore when we employ the term Some, **it is implied** that we are not thinking of the whole, but of a part as distinguished from the whole – that is, of a part only. (Monck 1881: 156, emphasis added)

– and in Grice's (e.g. Nowell-Smith 1954, Fogelin 1967: 20–22; see Horn 1990, Chapman 2005 for discussion).[5]

The principle tacitly invoked by Mill and Monck for generating such implications (or sous-entendus) is formulated by Strawson (1952: 178–9) as a "general rule of linguistic conduct" he attributes to "Mr H P Grice": "One should not make the (logically) lesser, when one could truthfully (and with greater or equal clarity) make the greater claim." The implicational relation between the subcontraries *some* and *some not* is captured independently in this overlooked passage that stresses the role of cancelability in distinguishing what is said from "what can be understood without being said" while also touching on the roles of relevance, economy, and epistemic insecurity:

> What can be understood without being said is usually, in the interest of economy, not said . . . A person making a statement in the form, "Some S is P", generally wishes to suggest that some S also is not P. For, in the majority of cases, if he knew that all S is P, he would say so . . . If a person says, "Some grocers are honest", or "Some books are interesting", meaning to suggest that some grocers are not honest or that some textbooks are

> not interesting, he is really giving voice to a conjunctive proposition in an elliptical way.
>
> Though this is the usual manner of speech, there are circumstances, nevertheless, in which the particular proposition should be understood to mean just what it says and not something else over and above what it says. One such circumstance is that in which the speaker does not know whether the subcontrary proposition is also true; another is that in which the truth of the subcontrary is not of any moment. (Doyle 1951: 382)

Grice's contribution, beyond securing the naming rights to the relation in question, was to ground the operation of Mill's "sous-entendu of common conversation"[6] within an overall account of speaker meaning and the exploitation of conversational principles based on assumptions of the interlocutors' rationality and mutual goals. In fact, like presupposition, implicature was (re)introduced into the philosophical literature and thence into the consciousness of linguists not with a specialized label but as a species of implication distinct from logical implication or entailment:

> To say, "The king of France is wise" is, in some sense of "imply" to *imply* that there is a king of France. But this is a very special and odd sense of "imply". "Implies" in this sense is certainly not equivalent to "entails" (or "logically implies"). (Strawson 1950: III)
>
> If someone says "My wife is either in the kitchen or in the bedroom" it would normally be implied that he did not know in *which* of the two rooms she was. (Grice 1961: 130)

Just as Strawson (1952) carved out a dedicated relation of presupposition two years after his first broadside at the non-existent French king,[7] so too Grice ([1967]1989) advances specialized labels – conventional and non-conventional (specifically conversational) implicature – for what he had earlier (1961: §3) described as varieties of (non-logical) implication delineated by the diagnostics of cancelability and detachability.

Conversational implicature arises from the shared presumption that S and H interact to reach a shared goal. A speaker S saying *p* and implicating *q* counts on her interlocutor's ability to compute what was meant from what was said, based on the assumption that both S and H are rational agents. On Grice's view, speakers implicate, hearers infer; such inferences may or may not succeed in recovering the speaker's intended implicature(s), if any. Nevertheless, it is S's assumption that H will draw the appropriate inference that makes implicature a rational possibility.[8]

The governing dictum is the Cooperative Principle: "Make your conversational contribution such as is required, at the stage at which it occurs, by the accepted purpose or direction of the talk exchange" (Grice 1989: 26). This principle is instantiated by a set of general maxims of conversation whose exploitation potentially yields implicatures:

(1) QUALITY: Try to make your contribution one that is true.
    1. Do not say what you believe to be false.
    2. Do not say that for which you lack evidence.

QUANTITY:
    1. Make your contribution as informative as is required (for the current purposes of the exchange).
    2. Do not make your contribution more informative than is required.

RELATION: Be relevant.

MANNER: Be perspicuous.
    1. Avoid obscurity of expression.
    2. Avoid ambiguity.
    3. Be brief. (Avoid unnecessary prolixity.)
    4. Be orderly.

A year after introducing the notion (or, more precisely, the label) of implicature in the William James lectures, Paul Grice published Lecture 6 as Grice 1968, situating his take on linguistic semantics within his broader project for speaker meaning (Grice 1968: 225; cf. also Grice 1989: 118):

> The wider programme . . . arises out of a distinction I wish to make within the total signification of a remark, a distinction between what the speaker has *said* (in a certain favored and maybe in some degree artificial, sense of "said"), and what he has "implicated" (e.g., implied, indicated, suggested, etc.), taking into account the fact that what he has implicated may be either *conventionally* implicated (implicated by virtue of the meaning of some word or phrase which he has used) or *non-conventionally* implicated (in which case the specification of implicature falls outside the specification of the conventional meaning of the words used).

The characterization of implicature as an aspect of speaker meaning – what is meant without being said – was set forth at a time when similar notions were in the air, especially that of Oxford and its ordinary-language sphere of influence. A case in point is "contextual implication" as invoked by Nowell-Smith (1954: 80–82) and revisited by Hungerland (1960) in her eponymous paper: "A statement *p* contextually implies *q* if anyone who knew the normal conventions of the language would be entitled to infer *q* from *p* in the context in which they occur." The *locus classicus* for contextual implication is the context familiar from Moore: "When a speaker uses a sentence to make a statement, it is contextually implied that he believes it to be true" (Nowell-Smith 1954: 81). But the relation between (my saying) *He has gone out* and my believing that he has gone out cannot be assimilated to conversational implicature, for reasons Grice himself would later provide:[9]

> On my account, it will not be true that when I say that *p*, I conversationally implicate that I believe that *p*; for to suppose that I believe that *p* (or rather

> think of myself as believing that *p*) is just to suppose that I am observing the first maxim of Quality on this occasion. I think this consequence is intuitively acceptable; it is not a natural use of language to describe one who has said that *p* as having, for example, "implied", "indicated", or "suggested" that he believes that *p*. The natural thing to say is that he has expressed (or at least purported to express) the belief that *p*. (Grice 1989: 42)

The difficulty of canceling such a (putative) implicature without epistemic or doxastic anomaly also argues against such an analysis. This applies as well to other cases of sincerity conditions Hungerland cites, e.g. the relation between *I promise to p* and *I intend to p*. Another of Nowell-Smith's examples of contextual implication does qualify as prefiguring Grice on the maxims, specifically Relation, although not on implicature as such: "What a speaker says may be assumed to be relevant to the interests of the audience." This maxim may indeed be overridden, as Nowell-Smith and Hungerland both observe. But Hungerland (1960: 212) is properly skeptical of the heterogeneity of a construct that extends from this relevance injunction to the sincerity condition on assertions and promises. Nowell-Smith himself concedes that the violation of the latter leads to "logical oddity" – "It's raining, but I don't believe it is" – while the non-observance of the relevance rule, according to him (and Hungerland), runs the mere risk of boredom. Even when relevance violations produce confusion or a recognition of the different conversational goals of speaker and hearer, such consequences do not rise to the level of the "logical oddity" of Moore's paradox.

4.3 Scalar implicature and the maxim map

Conversational implicature differs from contextual implication, and non-demonstrative implication more generally, in being defined as a relation between a speaker (not a sentence!) and a proposition that typically arises from the exploitation of the maxims (Grice 1989: 26ff.; cf. Horn 2004); in the case of scalar implicature in particular, what is implicated depends on what isn't (but could have been) said. The crucial principle of informative strength or quantity – whether formulated à la Strawson 1952 channeling Grice ("One should not make the (logically) lesser, when one could truthfully (and with greater or equal clarity) make the greater claim"), à la Grice 1961 ("One should not make a weaker statement rather than a stronger one unless there is a good reason for so doing") or à la Grice [1967]1989 ("Make your contribution as informative as is required (for the current purposes of the exchange)") – balances what the speaker can and does say with what she doesn't and hence presumably can't (or shouldn't) say. A speaker may opt for a weaker utterance from a belief that to utter its stronger counterpart might violate considerations of relevance, brevity, clarity, or politeness (note

the codicils and parentheticals in each of the formulations above[10]), but especially – as Mill and Doyle foresaw – from a lack of certainty that the stronger counterpart holds.

This reasoning, exploiting Grice's first quantity maxim, is systematically exploited to yield upper-bounding SCALAR IMPLICATURES associated with relatively weak scalar operators, those configurable on a scale defined by unilateral entailment as in <*all*, *most*, *many*, *some*>. What is ***said*** in the use of a weak scalar value like those boldfaced in (2) is the lower bound (. . . *at least* . . .); what is ***implicated***, in the absence of contextual or linguistic cancellation, is the upper bound (. . . *at most* . . .). What is ***communicated*** is the "two-sided reading" that combines what is said with what is implicated. Thus in (2c), to quote Mill (1867: 512), "If we assert that a man who has acted in a particular way must be either a knave or a fool, we by no means assert . . . that he cannot be both" – but we do typically communicate this "exclusive" understanding of the disjunction.

(2)		1-SIDED READING	2-SIDED READING
a.	You ate **some** of the cake.	'some if not all'	'some but not all'
b.	It's **possible** she'll win.	'at least possible'	'possible but not certain'
c.	He's a knave **or** a fool.	'. . . and perhaps both'	'. . . but not both'
d.	It's **warm**.	'at least warm'	'warm but not hot'

The alternative view on which each scalar predication in (2) is lexically ambiguous between one-sided and two-sided readings contravenes Grice's (1989: 47) Modified Occam's Razor: "Senses are not to be multiplied beyond necessity." Scalar implicature was introduced and formalized in work by Horn (1972, 1989), Gazdar (1979), Hirschberg (1991), and Levinson (2000); cf. also Matsumoto (1995), Katzir (2007), and Geurts (2010) for insightful discussions of certain problems arising in the implementation of the central notions involved and Bontly (2005) for a defense of Modified Occam's Razor based on its role as a heuristic in acquisition.

The implicature-based approach to scalar predications has been vigorously challenged by relevance theorists (see Carston 1988, 2002, 2004a and work reviewed therein), who take such sentences to involve propositional ambiguity, with the pragmatically enriched two-sided meanings constituting not implicatures but EXPLICATURES, pragmatically derived components of propositional content.[11]

Two major challenges to the Gricean picture of implicatures involve the number and status of the maxims and the relationship between implicature and propositional content. To begin with the former issue, Grice himself later acknowledged (1989: 371ff.) that the four macroprinciples (inspired by Kant) and nine actual maxims in his inventory are somewhat overlapping and non-coordinate. The number of maxims has been revised both upward (Leech 1983) and downward. The dualistic program of Horn (1984b, 1989,

2007a) follows Grice (1989: 371) in ascribing a privileged status to Quality, on the grounds that without the observation of Quality, or Lewis's (1969) convention of truthfulness, any question of the observance of the other maxims fails to arise (though relevance theorists, beginning with Sperber and Wilson 1986, offer a dissenting view). The remaining maxims are subsumed under two countervailing functional principles governing the economy of communication. On Horn's Manichaean model, implicatures may be generated by either the Q Principle (essentially 'Say enough', generalizing Grice's first sub-maxim of Quantity and collecting the first two 'clarity' sub-maxims of Manner) or the R Principle ('Don't say too much', subsuming Relation, the second Quantity sub-maxim, and brevity).

The hearer-oriented Q Principle is a lower-bounding guarantee of the sufficiency of informative content, exploited to generate upper-bounding (typically scalar) implicata. The speaker-oriented R Principle reflects Zipf's principle of least effort dictating minimization of form, exploited to induce strengthening implicata; it is responsible for euphemism, indirect speech acts, neg-raising, and meaning change (Horn 2007a). Opposition and equilibria between speaker's and hearer's communicative economies have been posited since Paul (1889: 351ff.) and Zipf (1949: 20ff.). According to the division of pragmatic labor (Horn 1984b), a relatively unmarked form – briefer and/or more lexicalized – will tend to become R-associated with a particular unmarked, stereotypical meaning, use, or situation, while its periphrastic or less lexicalized counterpart, typically more complex or prolix, will tend to be Q-restricted by implicature to those situations outside the stereotype, for which the unmarked expression could not have been used appropriately (as in *kill* vs *cause to die*, or *mother* vs *father's wife*). Formalizations of the division of pragmatic labor have been undertaken within bidirectional optimality theory and game-theoretic pragmatics; cf. e.g. Blutner 2004, van Rooij 2009, and references cited therein.

Levinson's (2000) framework posits an interaction of three heuristics: Q, I (for Informativeness, ≈ Horn's R), and M (Manner). Levinson's reconstruction of the division of pragmatic labor involves not Q but the M heuristic, given that *some* differs from *all* in informative content whereas *kill* differs from *cause to die* in complexity of production or processing. As Levinson acknowledges, however, the Q and M patterns are closely related, since each is negatively defined and linguistically motivated: S tacitly knows that H will infer from S's failure to use a more informative and/or briefer form that S was not in a position to have used that form. Unlike Q implicature, R/I-based implicature is not negative in character and is socially rather than linguistically motivated, typically yielding a culturally salient stereotype (cf. Huang 2007 for a useful overview).

Relevance theorists (e.g. Sperber and Wilson 1986, Carston 2002) posit one pragmatic principle, that of Relevance, defined in non-Gricean terms. It may be argued, however, that the RT program is itself covertly Manichaean, given that Relevance itself is calibrated as a minimax of effort and effect. In the

words of Carston (1990: 231), "Human cognitive activity is driven by the goal of maximizing relevance: that is . . . to derive as great a range of contextual effects as possible for the least expenditure of effort."

4.4 What is meant and what is said: the -plicature family

We now return to the perennial dispute over the shape of the landscape of implied meaning. In recent years a partial consensus has formed as to semantic underspecification and pragmatic enrichment, one that transgresses the view inherited from Grice that the pragmatics can be simply "read off" the semantics. When we turn from the relatively straightforward cases of reference fixing and ambiguity resolution acknowledged by Grice himself to the more problematic phenomena of completion, saturation, and free enrichment (cf. Bach 2001; Recanati 2001, 2002; Carston 2002; and references therein, as well as the relevant chapters in Horn and Ward 2004), it is clear we must grant what Bach (2005) terms the "contextualist platitude":

> Linguistic meaning generally underdetermines speaker meaning. That is, generally what a speaker means in uttering a sentence, even if the sentence is devoid of ambiguity, vagueness or indexicality, goes beyond what the sentence means.

Thus, the speaker uttering the non-bracketed material in each example in (3) may well communicate the full sentences indicated, enriched by the bracketed addenda. As seen from the cancelability evidence in (4), however, this process, resulting in truth-conditionally relevant propositions not directly expressed, is pragmatic in character.

(3) a. I haven't had breakfast {today}.
b. John and Mary are married {to each other}.
c. They had a baby and they got married {in that order}.
d. Dana is ready {for the exam}.

(4) a. John and Mary are married, but not to each other.
b. They had a baby and got married, but not necessarily in that order.

Those enrichments constituting necessary conditions for the expression of truth-evaluable propositions involve what Recanati has called saturation and Bach completion. Recanati (2002) distinguishes the bottom-up processes linguistically triggered by indexicals (*I*, *today*) and other expressions requiring saturation from those top-down modulation and free enrichment processes motivated on purely pragmatic grounds. At issue here are, for example, the underspecification of genitives (*John's car* – the one he owns? is driving? is following? is painting? is repairing?), unspecified comparison sets (*Chris is tall* – for an adult? for an adult American of the relevant sex?), and various expressions with apparent free variable slots: *You are late* (for what?), *Robin*

is too short (for what?). For Recanati – *contra* Stanley 2000 and King and Stanley 2005 – "unarticulated constituents" are real and cannot be reduced to independently motivated elements in abstract syntax or logical form.

Since Grice, the pragmatic landscape has exploded with aspects of meaning variously identified as conversational implicatures, conventional implicatures, presuppositions, implicitures, and explicatures. These are not simply diverse labels for given subclasses of implication but different ways of mapping the territory between the said and the meant. Situating "what is said" along this spectrum is itself controversial; what is said for Recanati (2001), Ariel (2008b), and the relevance theorists is enriched by pragmatically derived material (hence constituting an explicature). Levinson (2000), on the other hand, responds to the apparent need to allow "pragmatic intrusion" into what is said by allowing conversational implicatures to have truth-conditional consequences for the propositions in question, *contra* Grice; in cases like (3c) or Deirdre Wilson's aperçu *It's better to meet the love of your life and get married than to get married and meet the love of your life*, an implicature ("P precedes Q") can feed into (rather than just being read off) what is said. (See Carston 2002 and Russell, Benjamin 2006 for illuminating discussions of the complexity of conjunction buttressing.)

For orthodox Griceans, the pragmatically enriched proposition in such cases – what is communicated – is distinct from what is said. As we saw in §4.2, an "austere" conception of what is said (to borrow Jenny Saul's phrase; cf. Borg 2004, Horn 2009b), corresponding closely to the syntax of the sentence uttered and excluding pragmatically derived material, may have more to recommend it than first appears. Further, as Bach (2001) observes, once we abandon "OSOP" (the One Sentence, One Proposition assumption) we can recognize that a sentence may express not only more than one proposition but fewer than one. What is said in *Dana is ready* constitutes not a truth-evaluable proposition but a propositional radical. Completing such a radical within a given context to yield e.g. *Dana is ready to write a dissertation* yields not what is said (which is tightly constrained by the actual syntax) or an explicature (since there is nothing explicit about it), but rather an IMPLICITURE, a proposition implicit in what is said in a given context as opposed to a true implicature, a proposition read off what is said (or the way it is said). What Grice failed to recognize, argues Bach, is the non-exhaustive nature of the opposition between what is said and what is implicated.

"Scalar 'implicatures' are implicatures" is #9 in Bach's hit parade of misconceptions (2006c: 28–9): since a speaker uttering "Some of the boys went to the party" means not two separate things but just one, i.e. that some but not all of them went, this enriched proposition is an implic-i-ture (built up from what is said), not an implicature. But on the Gricean account (Horn 1972, 1989; Gazdar 1979; Hirschberg 1991), the strong scalar implicature here is "Not all of the boys went to the party"; this combines with what is said ("Some . . . ") to yield what is communicated ("Some but not all . . . "). Thus the impliciture includes the scalar implicature rather than supplanting it.[12]

While Levinson (2000) defines generalized conversational implicatures as default inferences, others argue that they are neither inferences – an implicature is an aspect of speaker's meaning, not hearer's interpretation[13] – nor true defaults. This last point is especially worth stressing in the light of much recent work in experimental pragmatics (see e.g. Noveck and Posada 2003; Bott and Noveck 2004; Breheny *et al*. 2006; Katsos 2008) suggesting that children and adults do not first automatically construct implicature-based enriched meanings for scalar predications and then, when the "default" interpretation is seen to be inconsistent with the local context, undo such meanings and revert to the minimal implicature-free meaning. To the extent that this empirical work on the processing of implicature recovery can be substantiated and extended, this is a very interesting result, but not (contrary to some claims) one that threatens the actual Gricean tradition, which predicts no automatic enrichment or default interpretation. This is clear from the passage distinguishing generalized and particularized implicature (Grice 1989: 37, emphases added):

> I have so far considered only cases of what I might call 'particularized conversational implicature' . . . in which an implicature is carried by saying that p **on a particular occasion in virtue of special features of the context**, cases in which there is no room for the idea that an implicature of this sort is normally carried by saying that p. But there are cases of generalized conversational implicature. Sometimes one can say that the use of a certain form of words in an utterance would **normally (in the absence of special circumstances)** carry such-and-such an implicature or type of implicature.

The classic contrast here dates back to Grice 1961: §3 – the particularized implicature with the "Gricean letter of recommendation" for a philosophy job candidate (*Jones has beautiful handwriting and his English is grammatical*) vs the generalized implicature with logical disjunction (*My wife is in Oxford or in London*, implicating I don't know which). Crucially, an implicature may arise in an unmarked or default context without thereby constituting a default or automatic inference. (See Bezuidenhout 2002a; Jaszczolt 2005; and Geurts 2009 for different views on defaults and their relation to implicature.)

Despite their substantial differences (from each other and from Grice) as to the role of implicature and the relation between what is implicated and what is said, the proponents of the approaches touched on above share Grice's commitment to situating implicature within a rationality-based pragmatics. On a competing view that has recently been elaborated by Chierchia (2004) and his colleagues, scalar implicatures in particular are generated locally as part of the grammar and/or the conventional lexical semantics of weak scalar operators. Support for this variety of defaultism involves an appeal to cases in which the Gricean model appears to yield the wrong results, thus arguing for local computation of "embedded implicatures." Others (e.g. Sauerland 2004; Russell, Benjamin 2006; Horn 2006) have challenged these conclusions and defended a global account of implicature along Gricean lines. In particular,

Geurts (2009, 2010) provides a broad survey of the landscape. Drawing a distinction between marked L[evinson]-type cases and unmarked C[herchia]-type cases of putative locality effects, Geurts (2009) argues that unlike the latter type, the Levinsonian contrast-induced narrowings represent true problems for a classical Gricean (or neo-Gricean) theory of implicature but shows that these can be handled by allowing upper-bounding to enter into the reinterpretation of what scalar operators express, a reinterpretation that is itself pragmatic in nature. In his treatise on Q-implicatures, Geurts (2010) argues that the conventionalist alternative to a Gricean approach is not only stipulative but also empirically flawed in predicting the full range of implicature-related results.

4.5 Conventional implicature from Frege to Grice (and beyond)

Alongside the successful conversational implicature model, Grice's category of CONVENTIONAL IMPLICATURE – a non-cancelable but truth-conditionally transparent component of encoded content – plays the role of ugly stepsister. The coherence of this category has evoked much skepticism: Bach (1999a) consigns it to the dustbin of mythology, Carston (2002: 134) remarks that "there simply is no such thing as 'conventional' implicature in relevance theory (or, we would argue, in reality)," while Potts (2005, 2007b) rehabilitates it in a different guise. But Grice's account of conventional content that does not affect the truth conditions of the asserted proposition has a rich lineage. Frege (1892, 1897, 1918) delineates a class of meanings that, while of linguistic interest, do not "affect the thought":

> With the sentence "Alfred has still not come" one really says "Alfred has not come" and, at the same time **hints** [*andeutet*] that his arrival is expected, but it is only hinted. It cannot be said that, since Alfred's arrival is not expected, the sense of the sentence is therefore false. The word 'but' differs from 'and' in that with it one **intimates** [*andeutet*] that what follows it is in contrast with what would be expected from what preceded it. Such suggestions in speech make no difference to the thought. A sentence can be transformed by changing the verb from active to passive and making the object the subject at the same time . . . Naturally such transformations are not indifferent in every respect but they do not touch the thought, they do not touch what is true or false. (Frege 1918: 295–6)

While recent scholarship largely follows Dummett (1973) in dismissing Frege's positive proposals in this area as representing a confused and subjective notion of "tone," this mischaracterizes Frege's actual account of the relevant phenomena. The two verbs in Geach's rendering highlighted above – *hint* and *intimate* – both translate Frege's *andeuten*, i.e. 'conventionally implicate'; no subjectivity or confusion is involved.

For a range of constructions including discourse particles (*but*, *even*, Ger. *ja*, *doch*), subject-oriented adverbs, epithets, and other "loaded" words, a version of the approach proposed by Frege and Grice remains eminently plausible (Barker 2003; Horn 2007b, 2008; Gutzmann 2008; Williamson 2009). Such an approach extends naturally to a range of other linguistic phenomena, including the familiar vs formal second person singular ("T/V") pronouns of many modern European languages, evidential markers, and arguably the uniqueness/maximality condition on definite descriptions. In addition, certain syntactic constructions can be profitably analyzed along these lines such as the southern US English "personal dative," a non-argument pronominal appearing in transitive clauses that obligatorily coindexes the subject as exemplified in *I love me some squid*, truth-conditionally equivalent to, but not fully synonymous with, 'I love squid' (Horn 2008). In each case, we find aspects of conventional content that are not entailed and do not fall inside the scope of logical operators.

The category of conventional implicatures poses a complication for the distinction between what is said and what is meant. Such expressions present a recalcitrant residue for Grice (who was concerned with delineating what is said and what is conversationally, and hence calculably, implicated) as they did for Frege (who was concerned with the thought, i.e. with sense and potential reference); for both, detecting a conventional implicature facilitates the real work by clearing away the brush. But Grice did undertake to situate this relation within what we refer to (though he did not) as the semantics/pragmatics divide. His contributions in this area, if not always accepted, are widely recognized, as in this passage from Davidson (1986: 161–2):

> It does not seem plausible that there is a strict rule fixing the occasions on which we should attach significance to the order in which conjoined sentences appear in a conjunction: the difference between "They got married and had a child" and "They had a child and got married." Interpreters certainly can make these distinctions. But part of the burden of this paper is that much that they can do should not count as part of their *linguistic* competence. The contrast in what is meant or implied by the use of "but" instead of "and" seems to me another matter, since no amount of common sense unaccompanied by linguistic lore would enable an interpreter to figure it out. Paul Grice has done more than anyone else to bring these problems to our attention and help to sort them out.

But how, exactly, does this sorting work? If descriptive content, reflecting what is said, is clearly semantic and if what is conversationally implicated (e.g. the "for all I know, not both p and q" upper-bounding implicatum associated with the utterance of the disjunction *p or q* or the negative effect of the Gricean letter of recommendation) is pragmatic (*pace* Chierchia 2004, among others), where is conventional implicature located? One standard view is that by falling outside what is said, the conventionally implicated

must be pragmatic (see e.g. Gutzmann 2008: 59). One argument on this side is terminological; in Kaplan's words (1999: 20–21):

> According to Grice's quite plausible analysis of such logical particles as "but", "nevertheless", "although", and "in spite of the fact", they all have the same descriptive content as "and" and differ only in expressive content . . . The arguments I will present are meant to show that even accepting Grice's analysis, the logic is affected by the choice of particle . . . If this is correct, then generations of logic teachers, including myself, have been misleading the youth. Grice sides with the logic teachers, and though he regards the expressive content as *conventional* and hence (I would say) semantic (as opposed to being a consequence of his conversational maxims), he categorizes it with the maxim-generated *implicatures*.

To be sure, conventional implicatures are implicatures. But then again, they are conventional; we are indeed dealing here, unlike in the maxim-based cases, with aspects of content.

Two decades after the William James lectures, Grice revisited these categories in his Retrospective Epilogue (1989: 359–65), where he sought to establish central and non-central modes of meaning through the criteria of FORMALITY ("whether or not the relevant signification is part of the conventional meaning of the signifying expression") and DICTIVENESS ("whether or not the relevant signification is part of what the signifying expression *says*"). Thus, when a speaker says "*p; on the other hand, q*" in the absence of any intended contrast of any kind between *p* and *q*, "one would be inclined to say that a condition conventionally signified by the presence of the phrase 'on the other hand' was in fact not realized and so that the speaker had done violence to the conventional meaning of, indeed had misused, the phrase 'on the other hand'." Crucially, however, "the nonrealization of this condition would also be regarded as insufficient to falsify the speaker's statement" (Grice 1989: 361). Thus, formality without dictiveness yields conventional implicature. (As for dictiveness without formality, a plausible candidate is the pragmatically enriched content of relevance theorists, the TRUTH-CONDITIONAL PRAGMATICS of Recanati 2001.)

In uttering a given sentence in a given context, the speaker may intentionally communicate more than one truth-evaluable proposition, but these communicated propositions do not necessarily have equal status; in *They are poor but happy*, the conjunctive content is truth-conditionally relevant while the contrastive content is not. Yet *but* and *and* are not synonyms. Conventional implicatures constitute part of encoded content but not part of truth-conditional content per se; their falsity does not project as falsity of the expression to which they contribute (cf. Barker 2003). What they contribute is use-conditional meaning (Kaplan 1999, Gutzmann 2008).

Besides detachability and non-cancelability (Grice 1961, 1989), additional diagnostics for conventional implicatures include their tendency to project out of embedded contexts, their immunity to certain kinds of objection,

and their contextual variability or DESCRIPTIVE INEFFABILITY (Potts 2007b, Horn 2008). The last property is the difficulty of pinning down the precise contribution of *but* (contrast? unexpectedness?) or *even* (relative or absolute? unlikelihood or noteworthiness?) and in the intention prompting the use of a second person "familiar" (**T**[u]) or "formal" (**V**[ous]) pronoun (**T** can be affectionate, presumptuous, comradely, or condescending; **V** can be polite, aloof, diplomatic, or hostile). Ineffability has a plausible source: modulo the well-known problems associated with vagueness, it is plausible that the edges of truth-conditional meaning should for the most part remain discrete, while inconsistency in the mental representation of non-truth-conditionally relevant content is less pernicious. If you know that my use of *vous* rather than *tu* signals some aspect of formal respect, distancing, or lack of intimacy, my precise motives can be left underdetermined, but if you don't know whether I'm using a second person or third person pronoun, the indeterminacy is more costly.

For Frege and Grice, identifying the class of conventional implicature-licensing constructions – scalar particles, speaker-oriented sentence adverbs, epithets and slurs, prosodic features, evidential markers, "affected" pronominals, word order effects – serves to characterize them in terms of what they are not: they do not affect the thought or the truth-conditionally relevant meaning of a given expression, and at the same time they are not derivable from general principles of rationality. While they share the former property with conversational implicatures, they differ crucially from them in the latter respect.

4.6 Implication and speaker meaning

Whether conversationally or conventionally triggered, implicatures are generally understood to constitute a proper subset of speaker-meant implications. As we have seen, the assimilation of implicature to inference has been deplored as an instance of the *imply/infer* confusion. But just as the latter turns out to be more complex than meets the eye, so too the subsumption of conversational implicature within the category of speaker meaning is not entirely straightforward. Saul (2002a) has argued that the full range of Grice's remarks on the topic suggests that we need to allow for a category of AUDIENCE-IMPLICATURE along with the traditional UTTERER-IMPLICATURE. Others (e.g. Bach 2001, 2006c), while acknowledging the role of the speaker's expectations about the inferences that the hearer can reasonably be expected to draw, defend the classical view that implicatures constitute part of speaker meaning.

In this respect, it is worth touching on evidence that *imply* in ordinary language use is not always definable in terms of speaker intention. A Google search on the string *didn't mean to imply* returns 929 valid hits (retrieved 16 June 2010) – "The President didn't mean to imply that AARP supports

current health care legislation", "Bolivian President Evo Morales Didn't Mean to Imply Straight Guys Go Gay Over Chicken", etc. – with the sense of 'x didn't intend in saying p that y should infer q'. So we can unintentionally imply propositions, whether or not we can unintentionally implicate them. This is reinforced by the 785 googla instantiating *unintentionally imply/implied* – "I'm sorry if I unintentionally implied here anything that touched your feelings", "I realized that I might have unintentionally implied that no Christian would feel right about attending a pagan ceremony" – and the 705 hits for *accidentally imply/implied*, including "When you accidentally imply she's fat – send chocolates instead" and "Did you know you're supposed to say 'best wishes' to the bride instead of 'congratulations', lest you accidentally imply that the bride won her groom through trickery or deceit? Like she most likely did. But still."[14]

Ideally, *imply* and *infer* would map respectively into what the speaker intends and what the hearer grasps. Most likely, that would make for a simpler pragmatics. But still.