12 Vocal Strategies in Verbal Irony

Gregory A. Bryant

Conversation is the fundamental act of human communication, comprising multimodal social interaction rooted in language. The psychology that allows us to socialize with one another through language incorporates many behavioral tools, including, quite notably, the voice. Language is ordinarily spoken, and the voice constitutes a rich communicative source including not only speech but also other expressive elements such as indexical cues (e.g., age and sex) and myriad emotional signals that help convey people's communicative intent. Understanding intentions in conversation is not simply a matter of decoding surface features of utterances, of course. Consider the following real exchange (Bryant, 2011: 298) captured between two college roommates:

> KRISTEN: you know, cuz you can't necessarily go, like, away you know, like, when I get annoyed, like, with you, or just plain annoyed in general
> SHAYNA: it happens? *laughing*
> KRISTEN: *laughing* no, it never happens

Kristen is explaining the problems associated with being confined to a small space with other people for long periods of time, giving the example just prior to this exchange of living with her sister during her high school years. Shayna playfully asks Kristen if she becomes annoyed in their living arrangement, referring to past times when it happened (that were discussed earlier), making the question ironic. She signals her playful intent using laughter, which is immediately met with colaughter by Kristen, along with a playful response. Bryant (2011) described this as an ironic adjacency pair, referring to a common pattern in conversation where ironic utterances are met with ironic responses, forming a brief exchange.

This example illustrates a typical case of how verbal irony manifests in ordinary conversation, including the various devices people use to achieve their communicative effects. Speaker meaning is often indirect and must be inferred by receivers based on multiple sources of evidence, including the specific, rich contexts in which most communicative acts occur. Vocal signals facilitate this process and are used by speakers to make implied messages relevant for listeners. In this chapter, I explore the use of vocal patterns as strategic signals in communicating verbal irony, a specific category of indirect language use. I argue that connections between ironic meaning and vocal strategies are explainable with reference to general principles of ostensive communication and form–function relationships between vocal acoustics and speaker intent.

Vocal signals and indirect language

Many scholars have wondered why people use language indirectly. Part of the answer is that we cannot help it – pragmatic reasoning emerges from the integration of social cognition and linguistic communication. Language use evolved to exploit our social-inferencing psychology (Scott-Phillips, 2014). Accordingly, we can conceptualize language use as strategic. That is, interlocutors have social communicative intent motivated by higher-order goals (for a review see Lieberman, 2013), and engagement with language becomes a means to an end. This is not to say that people are conscious of their strategies – in many cases we have little or no awareness of the tactics we employ. But our evolved social cognition is shaped to produce and understand communicative acts that researchers can describe analytically using probabilistic models of interactants' reasoning with game-theoretic logic (e.g., Goodman & Frank, 2016; Pinker et al., 2008). This is how I am using the term "strategy." As I describe later, the assessment of communicative goals (or functions) in the context of an interactive "game" affords an analysis of what kinds of structural features (i.e., forms) we should expect in signaling to implement a given strategy. This form–function approach provides a theoretical framework from which vocal signaling and language use can be understood (Owren & Rendall, 2001; Pisanski & Bryant, 2019), and provides a solid basis to analyze specific phenomena such as how people use their voices when employing devices such as verbal irony.

Before we delve into irony, consider a simple scenario where vocal signals drive indirect communicative intent.

In a grocery store, John asked Mary, a store employee, a question requesting information about the price of peaches. This question contained the typical prosodic features of English interrogatives: characteristic terminal pitch rise, and local pitch movement on the initial word (how), among other subtler features that distinguish questions from statements (Eady & Cooper, 1986; Pell, 2001). The prosody in John's query was not likely necessary to effectively ask the question, but it was produced as automatically as the linguistic utterance itself – an interesting aspect of language production where distinct dimensions are subject to varying levels of volitional control. Mary's answer, and the possible prosodic variants (Figure 12.1), reveal the complexity of vocal signaling during everyday indirect speech. In her response, imagine that Mary provided the requested information, but produced it with a question intonation signaling an added implied message (Figure 12.1: versions *c* or *d*). Intonation phonologists (e.g., Ladd, 2008) would describe her pitch pattern as comprising at least two dimensions: relative prominence and tune. In English, both of the falling tunes constitute typical responses to questions, albeit with prominence signaling possible clarifications on one aspect or another (e.g., in version *a*, perhaps John and Mary's mutual recognition of the unclear writing on the sign caused Mary to clarify that it was "pound" and not some other abbreviation). But rising tunes in responses to questions indicate indirect meaning. In version c,

John: How much for peaches?

Mary: Dollar a pound?

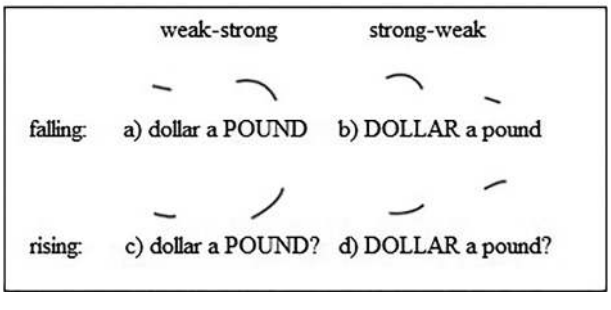

Figure 12.1 *Four possible intonation contours in English incorporating relative prominence and tune in response to a question. Adapted from Ladd (2008).*

we can label her prosody as a weak–strong, rising tune. When this structure accompanies an answer to a question, it indicates that Mary was, for instance, implicitly seeking approval for the price, and asking John if it was reasonable. It is worth noting that these particular prosodic patterns are characteristic of English, while other languages mark similar information in alternative ways (Féry, 2013; Jun, 2019).

As this example illustrates, prosody is often limited in its specificity, but powerful when interacting with context. The prosodic form in version *c* or *d* could equally well signal Mary's uncertainty (e.g., perhaps she needed to look it up to know exactly and she was guessing). Assume for a moment that Mary was not an employee at the store, and John asked her because he was unable to determine the price on his own. Mary could have answered as a guess and communicated that she agreed the mutually recognized information was not clear (e.g., a poorly made sign near the peaches). The prosodic structure could be highly similar in all of these examples, but listeners can use nuanced prosodic distinctions in context to derive proper inferences most of the time. Note, however, that while prosody is only part of the story in this example, without it (or some other signal such as a facial expression), the implied message could not be communicated effectively. Prosody becomes a crucial source of information in the speech stream, containing relevant communicative properties in an integrative social comprehension system. Conversely, this specific context would allow John to ask his question with no utterance at all – he could quite easily just hold up a peach in an intentional manner, and with a facial gesture and directed gaze, communicate his request for information and elicit a relevant response.

By this view, specific pragmatic phenomena need to be assessed in the theoretical context of ostensive communication (Sperber & Wilson, 1995). We have the tools of language and a variety of other communicative devices at our

disposal, including facial, vocal, and gestural signals, all manifesting in a rich social context, and used strategically to achieve social goals. But researchers have often preferred to explore specific tropes and attempt to identify reliable signals associated with them. A popular example is the study of how verbal irony is communicated.

Verbal Irony and the Voice

Verbal irony is best described as a type of indirect language where a speaker produces an explicit evaluative utterance that implicates an unstated, opposing evaluation (Burgers et al., 2011; Bryant, 2012). Scholars have defined it variably, often around the general idea of saying the opposite of what you mean. Prototypical examples of verbal irony can be illustrative of some important aspects, but do not represent well how it typically occurs in real talk. For example, a speaker might say "Nice weather we're having" while engaged in shared attention with a fellow conversant on a heavy rainstorm. Or a person might compliment a major life achievement by stating "I knew you wouldn't amount to much." While these examples are easily understood by students of verbal irony, and naïve listeners in experimental tests, they do not reflect the highly context-sensitive and esoteric ways that verbal irony regularly appears in actual discourse (Bryant, 2011; Gibbs, 2000). Ironies are messy in real language use and are easily misunderstood, sometimes by design (e.g., in the presence of particular overhearers). Given the importance of speech in linguistic communication, the voice naturally contains an abundance of information about emotion and intent that interacts with language. Not surprisingly, vocal features often play a significant role in how people produce and understand verbal irony.

Early theorizing about irony and the voice was done largely without acoustic or perceptual data, and instead relied mostly on descriptive analyses. For example, Cutler (1974) examined different forms of verbal irony, and pointed out correctly that vocal elements were not always necessary for successful communication. But when people did use their voice to signal irony, they typically used what Cutler called an "ironic intonation," characterized by some combination of nasalization, slowed speech, and stress or lengthening of key words. Sperber and Wilson (1981) questioned the validity of the concept of an "ironical tone," arguing that given the variety of attitudes speakers can express regarding a mentioned proposition in an ironic utterance, we should not expect a single intonation pattern but rather variation depending on the content of the ironic utterance (see also Muecke, 1978). I will argue a very similar point here. Nevertheless, the ironic tone of voice idea has stuck in many ways. For instance, Clark and Gerrig (1984) claimed that the pretense theory of irony provided "a natural account of the ironic tone of voice" (p. 122), a point Sperber (1984) disputed, arguing instead that Clark and Gerrig were really describing a parodic tone. These ideas are compatible – speakers can adopt a pretended role of another speaker and use a special voice to convey

that role, and at the same time mention attributed propositions that also drive particular vocal patterns. In any case, the debate regarding how the voice helps convey ironic meaning goes back decades.

More recently, researchers began examining vocal correlates of verbal irony with acoustic analyses and perceptual experiments. A good deal of this work, using actors, has found that speakers tend to produce prosodic patterns somewhat consistent with earlier notions of an ironic tone, such as lowered pitch, slowed duration, and a sneering tone (for reviews, see Bryant & Fox Tree, 2002, 2005; Cheang & Pell, 2008). But numerous studies have revealed variations, especially when different irony subtypes were explored or different languages were examined. For example, "blame by praise" and "praise by blame" ironies in Italian revealed distinct patterns in f_o (Anolli et al., 2002). Cheang and Pell (2009) compared verbal irony production in English and Cantonese and found that Cantonese speakers tended to use relatively higher f_o than English speakers. A similar effect of higher f_o in ironic speech has been found in French (González-Fuente et al., 2016). Kreuz and Roberts (1995) suggested that the supposed ironic tone was really just the vocal signaling of exaggeration and that it was relatively unusual in ordinary discourse. Ultimately, there are good reasons to suspect that in many cases prosodic consistencies across actors are the products of stereotyped performances capturing elements of folk notions of what irony sounds like, and do not adequately reflect the diversity of vocal signals in spontaneous verbal irony that manifest in typical conversations (Bryant & Fox Tree, 2005).

Nevertheless, the literature on vocal correlates of irony has informed researchers in other areas of psychology, such as developmental psychologists exploring children's language understanding, or neuropsychologists examining pragmatic deficits in brain-damaged individuals (for a review, see Bryant, 2012). These investigators generally use acted materials incorporating what many describe as a "sarcastic intonation" that presumably includes acoustic features like those described by Cutler (1974) and Rockwell (2000) among others: lowered pitch, slowed down, and perhaps sneering or nasal (i.e., spectral features affecting voice quality). Sarcasm can be thought of as a subtype of verbal irony with particular elements (e.g., biting ironic criticism) that contains a narrowed suite of linguistic and paralinguistic features (e.g., Cheang & Pell, 2008). Overall, perceptual studies using acted stimuli typically find that listeners rely on vocal features when judging ironic tokens, but strong contextual cues can cause these features to be ignored and/or unnecessary.

Vocal Strategies

Conversationalists can use verbal irony for many social communicative reasons. Thus, we should expect speakers to adopt a variety of delivery strategies. The ways speakers use the voice to convey irony relate to more general principles of prosodic production. For example, emphasizing a particular

Table 12.1 *A simple categorization of vocal strategies common in ironic speech*

Vocal strategy	Description
Local prosody	Pitch, loudness, speech rate, or voice quality features shifting individually or concurrently on segmental units (e.g., prominence on syllables or individual words)
Global prosody	Pitch, loudness, speech rate, or voice quality features shifting across suprasegmental units. Musical prosody in speech (tunes), often via vowel prolongation, rhythmic features, and suprasegmental prosodic changes (e.g., sing in high voice)
Voice impression	Shift spectral (i.e., perceptual voice quality) features of the voice to depict a specific person or imagined agent
Laughter and other nonverbal vocalizations	Generate laughter before, during, or after ironic utterance, setting up a play frame, or signaling ironic intent. Other vocalizations include snorts, fricatives, sighs, and gasps

word or phrase that sets it apart in the speech stream can guide listeners' inference processes. But some strategies might be relatively more common in ironic use, such as slowing down speech precipitously as described earlier. Table 12.1 presents a list of proposed categories of vocal strategies that speakers use to convey ironic meaning.

This categorization scheme is intended to provide a basic organization of the kinds of vocal approaches speakers adopt when using irony. In a given instance of ironic speech, speakers could potentially use just a single prosodic strategy, several from multiple categories, or none. As will be described in the section "Vocal Strategies and Signal Design," different prosodic features afford different types of effects in listeners. Moreover, multiple control mechanisms likely underlie prosodic production, allowing for the concurrent operation of two or more prosodic features at once (Cole, 2015; Fujisaki, 1983; McRoberts et al., 1995). It is also worth noting that significant cross-linguistic differences should be expected, especially to the extent that prosodic resources more generally pattern differently across languages. Base rate expectations are important since deviations from ordinary production can have illocutionary signaling value (Bryant, 2010). The four proposed categories are described in the following text.

Local prosody allows for intra-sentential focus and can dramatically affect communicated meaning. For instance, as we saw in Figure 12.1, the way speakers mark lexical items during questions affect what exactly is being asked. One common way to focus a segment is to alter the pitch (up or down) on a word or syllable, and possibly concurrently alter the duration, typically by lengthening (e.g., Chen & Boves, 2018). By focusing on an item prosodically,

a speaker highlights it for communicative effect. Endless examples are possible, but could include key words (marking a word to be noticed specifically); contrastive focus (emphasizing a lexical item to contrast it from other possible items in the slot, or to mark alternative meaning); and echoing (repeating a previously used word for rhetorical effect, often contrary to its original use). Of course, all of these prosodic functions occur in nonironic language as well, illustrating the ubiquitous importance of context and inference in all language use. These focus phenomena can be acoustically subtle, but communicatively powerful – listeners are quite sensitive to variations in prosodic signaling and information structure.

The ways global prosodic features are used in irony are similarly continuous with nonironic use. Generally, prosody that operates across multiple segments is used to convey intentional and emotional information that can be superimposed on linguistic meaning. A semantically neutral utterance can be produced with variations of pitch, loudness, and duration to express many nuanced emotional meanings recognizable across disparate cultures (e.g., Bryant & Barrett, 2007; Cowen et al., 2019; Pell et al., 2009). When using verbal irony, speakers can incorporate intonation patterns associated with emotions as part of staged performances (Bryant, 2011). A performance could be an exaggeration (e.g., using positive language in reference to something negative, and adding an overly positive prosodic form such as increased overall pitch, loudness, and speech rate), or it could run counter to the expected pattern (e.g., saying something happy, but produce a prosodic form communicating sadness such as lowered pitch, loudness, and slowed speech rate). Speakers have many performative options that can interact in complicated ways with language. Global prosodic forms can also turn musical, with features like vowel prolongation (e.g., fixed pitch), rhythmic elements, and other melodic aspects. Overall, global features help convey emotional attitudes toward attributed propositions that are separate from the information in the language itself.

Similarly, vocal impression often involves producing vocal changes suprasegmentally, but in these cases, the changes involve spectral shifts that modify voice quality (analogous to musical timbre). Voice quality can be altered for simple effects such as increased nasality, harshness, or apparent gender shifts, but can also be used in more complex ways such as imitating a specific individual, or adopting the persona of a fictional character. Simple voice impression effects are highly similar to global prosodic effects (and could be plausibly grouped as such), but the meanings contained in the acoustic information are far less likely to map onto emotional categories without context (emotion in voices can often be recognized without context or words). Vocal impressions for imitating a specific individual allow speakers to perform their ironic speech and emphasize the pretense behind their meaning (Clark & Gerrig, 1984). By playing a targeted individual, a speaker has a platform to vocalize at length and be understood as portraying that person.

The aforementioned vocal strategies all involve acoustic manipulations of the ironic speech. But other vocal strategies involve nonlinguistic utterances that

accompany verbal irony. Among the nonlinguistic vocal possibilities in irony, laughter is by far the most common, as will be described. But other vocalizations can be used of course, each contributing subtle meanings with affective impact. For example, speakers can use sighs, gasps, cries, coughs, snorts, moans, or fricatives (e.g., /f/ or /s/). These kinds of vocalizations could occur just prior to an ironic utterance, during one, or just after. Imagine a person, after hearing what they consider to be a bad idea from another speaker, first producing a sigh, and then immediately exclaim "great idea!" In such a case, the sigh could be understood as reflecting a genuine negative reaction, but it aids listeners in understanding a proceeding ironic remark. Of course, these vocalizations can also be used ironically. For instance, a person might report an event that is not surprising in any way (but perhaps implies it is), and a responder produces a large gasp as an ironic portrayal of shock. Other examples include portrayed crying ("*waah*, your life is so hard"), moaning ("*ughh*, I can't wait another second"), or coughing ("**cough sound** never mind me").

Given the preceding four categories of strategies, how might one generalize any particular vocal pattern to a given communicative goal? Unfortunately, it is not quite possible. Given context, any intentional speech act could be potentially linked to multiple vocal approaches. This underlies the main problem with the notion of an ironic tone of voice (Bryant & Fox Tree, 2005) – variation in production possibilities interacts with rich social contexts, creating a landscape of options with limited regularities. Rather than expecting consistent global prosodic patterning in ironic speech (i.e., an ironic tone), Bryant (2010) examined the role of prosodic contrasts in irony production (first proposed by Attardo et al., 2003). Speakers can signal indirect intent by altering one or more vocal dimensions in a way that stands out in its given context (e.g., the speech immediately preceding it, or a speaker's ordinary tone). In this study, spontaneously produced instances of verbal irony were extracted from natural conversations recorded in the lab, along with speech in the preceding context. Utterances were categorized as ironic targets, baseline utterances (occurring just before the target was uttered), and pre-baseline (occurring just before the baseline speech). All utterances were analyzed acoustically on basic perceptible dimensions (f_0 [mean and SD], dB [mean and SD], and speech rate [mean syllable duration]). Comparisons between pre-baseline and baseline utterances provided a base rate of prosodic contrasts not associated with verbal irony – there were some. But the rate of contrasts between baseline speech and ironic targets was considerably higher. Moreover, speakers used more prosodic contrasts simultaneously when speaking ironically.

These findings undermine the ironic tone of voice concept, and instead demonstrate that in spontaneous speech, conversationalists tailor their ironic utterances for the situation. Speakers were quite variable in the prosodic movements they made, although one consistent pattern emerged across many instances of ironic speech – speakers often slowed down. This has been noted with actors, and constitutes one element of a proposed ironic tone. Slowed speech is the most commonly reported prosodic correlate of irony, appearing across irony

subtypes within and across languages (e.g., Aguert, 2022; Bryant, 2010; Cheang & Pell, 2008, 2009; Chen & Boves, 2018).

As a mechanism for marking indirectness, prosodic contrasts make empirical sense for a number of reasons: (1) they exploit basic principles of auditory processing (e.g., change detection); (2) they easily afford a form–function analysis; (3) they explain a major source of variation in vocal signals of irony; (4) they are consistent with the literature on play signals in humans and nonhumans; and (5) they are consistent with a central theme running through different theoretical approaches to verbal irony (i.e., irony as a contrast between what is stated and what is expected) (e.g., Colston, 2000).

Communicative Intent in Verbal Irony

When using verbal irony, speakers can seek to fulfill a variety of communicative goals, often more than one at any time. For example, a speaker might attempt to be funny and could use vocal tools to help achieve this – perhaps they use a special voice, sing a particular melody, or hyper-articulate a word or phrase at a relevant moment in the speech. And, for instance, these same tactics could be used as a vehicle for mocking, which could be humorous for some audience members, but particularly hurtful for a target, or other audience members. At the same time, the speaker's attempt to be funny, with targeted antagonism, might be in the service of gaining an upper hand in a specific social situation. By using humorous, indirect language with a strategic performance (i.e., vocal or other nonverbal approaches such as bodily and/or facial gestures), speakers might enhance that social effort. The various ostensive forces in an interaction will shape how speakers use a particular strategy. For instance, more playful and less critical intent could affect the nature of a parodic tone in a voice impression. The point being, social communicative goals are multifaceted, hierarchical, context sensitive, and in deep interplay with multimodal signaling variables, including vocal behavior.

At a fundamental level, irony can be construed as a form of verbal play, and associated nonlinguistic signals can be understood, in many cases, as play signals (Bryant, 2011). Play vocalizations occur widely across mammal species, and in many primates manifest as what is arguably labeled "laughter" (for a review, see Winkler & Bryant, 2021). Social play often occurs as a means to calibrate skills, such as chase games emulating predator–prey interactions, or rough and tumble wrestling assisting the development of fighting ability (Burghardt, 2005). Verbal play in people could be potentially functioning at multiple levels, including calibrating social and language skills, and developing social bonds. But when animals (including humans) are engaged in play, they need a way to communicate the nonserious nature of the activity, and signal benign intent. Many scholars now agree that the origins of human laughter stem from a ritualized indicator of exertion during physical play (i.e., heavy breathing), and play vocalizations across many current species, including humans, are evolutionarily

related (Bryant, 2020; Provine, 2000; Winkler & Bryant, 2021). That is, human laughter is a species-typical version of a homologous play vocalization, and we have incorporated it into our complex communicative repertoire that includes language and other behaviors. Laughter is an evolved mechanism to reduce the likelihood of escalations to aggression, and achieves this effect by inducing positive affect and signaling cooperative intent.

Verbal irony, teasing, parody, and many other language games often have a strong play element, and they are experienced as humorous. Laughter occurs quite frequently in proximity to spontaneous occurrences of verbal irony (Bryant, 2011; Clift, 1999; Eisterhold et al., 2006; Gibbs, 2000; Kotthoff, 2003). Of course, verbal irony is often funny – humor emerges from interactions that include irony and it proximately facilitates the pragmatic effects, including being playful (Gibbs et al., 2014). Recent work has also shown that the presence of laughter causes participants to judge ironic speech as more indirect, and judged playfulness in isolated laughter is associated with the magnitude of its impact on ratings of indirectness of adjacent ironic utterances (Bryant & Nagy, in prep). Thus, laughter is a vocal strategy that increases the probability that interlocutors understand one another when using irony and is likely to be important especially when indirect language use could plausibly be misconstrued as aggressive or threatening. That said, laughter can be strategically used to enhance the threatening intent in sarcastic utterances as well (e.g., the veiled threat, "it would be a shame if… [laughter]").

This last point reveals an important distinction regarding the nature of how people "cooperate" that has consequences for how different communicative strategies might arise during social interaction. Neo-Gricean pragmatic approaches such as Relevance Theory incorporate a version of the cooperative principle introduced by Grice (1975). Cooperation, in this sense, means coordination, where interlocutors engage in joint action for mutual communicative and cognitive benefit. But cooperation, in the biological sense of the term, involves the costs and benefits of interaction shaped by the conflict of interest between interactants – a dynamic best understood from the perspective of evolutionary game theory. Indirect language, by this view, is one strategic possibility for maintaining plausible deniability of some attributed view or suggestion (minimizing potential costs), while still allowing for the transmission of an idea (maximizing potential benefits). Pinker et al. (2008) used the example of the rational briber to illustrate this point: speakers can indirectly make risky offers, such as a bribe, and if they have a receptive audience, a mutually beneficial transaction can occur. If the audience is not receptive, there is no direct evidence that a crime was committed (low risk). In game-theoretic terms, indirect speech constitutes an optimal strategy in such a scenario – we can exploit shared individual knowledge without establishing common knowledge. The same logic can apply to veiled threats, sexual come-ons, and other forms of indirectness including verbal irony.

Research examining the social impacts of using irony as a device for criticism illustrates this principle well. Speakers using sarcasm can convey intended

critical meanings while potentially avoiding the costs of issuing direct criticism, or enhance a critical blow while seemingly being playful, for example (there is evidence for both functions; see Colston, 1997; Dews & Winner, 1995). Vocal signaling can facilitate the varied possibilities. Moreover, prosodic forms on their own can communicate plausibly deniable meaning, often beyond the awareness of the speaker. A classic argument between intimate partners involves the dispute of what a speaker meant by the way they said it. Imagine a married couple preparing to leave their house for a party, and during small talk about previous parties at the same location the husband remarks to his wife, "Well, you always have a good time," but he states it in a way that indirectly refers to the wife's history of getting embarrassingly drunk on these occasions. The vocal options are numerous, and different forms will have different impacts. For instance, singing the phrase could be playful and critical at the same time, whereas a local prosodic focus on "you" could be construed as relatively more critical and not funny.

The comment is interpreted critically and makes the wife angry. The husband can plausibly deny this implicature, or at worst claim he was just playing and not seriously criticizing her, but both he and his wife ultimately know what he means. That is, they have individual shared knowledge of his indirect meaning, but since the implication is communicated largely through his vocal delivery (though also somewhat by his choice of words), common knowledge is not established (as opposed to if he said directly, "You always get drunk at these parties and it's embarrassing for both of us"). The wife is upset because the illocutionary force is as blunt (or possibly more so) as an explicit remark, and perhaps seems sneaky and/or mean-spirited. Denying such an intention is often not deceptive – speakers can behave in self-deceptive ways as a design feature of effective manipulation. There are few better ways to deny one's malicious intent than to not be consciously aware of it in the first place (Trivers, 2000). We unconsciously generate strategic utterances all the time, and then sometimes deny what we meant. This can explain why many people view sarcasm as a passive-aggressive strategy, and find it distasteful.

One dimension of utterance interpretation closely related to verbal irony is sincerity. How do we know if people really mean what they say, and what role do vocal signals play? The primary difference here, however, is that insincerity often amounts to a white lie as opposed to purposeful irony. But in both cases people generally do not mean what they literally say. Fish et al. (2017) created dialogs recorded by actors who produced responses to questions, mostly in the form of issuing solicited compliments (e.g., a question "How do I look in my bikini?" in which the responder either believes the person looks great or terrible). Actors were guided by the contexts to generate sincere or insincere responses that were either confident or uncertain, and the question/answer pairs were presented to listeners who rated respondents' sincerity. Listeners were able to reliably distinguish between sincere and insincere responses, and vocal features predicted judgment patterns. Consistent with work on verbal irony and the voice, slowed speech and lowered pitch were associated with insincerity. One

interpretation of these findings is that vocal signals of ironic speech exaggerate features that are associated with insincerity and uncertainty in direct speech. Many forms of indirect speech incorporate vocal features to enhance (or create) their effects. For instance, Cutler (1974) provided the example of a "doubting" intonation in the expression "it's not bad" with a fall–rise intonation on the last word that communicates the meaning of "it" being "not good at all."

Vocal Strategies and Signal Design

So how do people choose the way they modulate their voice to help convey their intended meaning when using indirect speech? As we have seen, the prosodic possibilities are vast, but there are some basic principles we can use to understand typical production strategies. Again, by strategy, I do not mean to suggest people are consciously or deliberating planning their vocal approach. But speakers attempt to maximize their communicative effectiveness in a given social context, and the efforts are shaped by principles of signal design that include the form–fit relationship between acoustic properties of vocalizations and communicative functions. Relevance Theory and a form–function approach together provide a useful integrative framework in this regard.

According to Relevance Theory, speakers aim to optimize their communicative behaviors to achieve maximum cognitive effects with the least possible effort on the part of hearers. The theory was developed with linguistic communication in mind, but the cooperative principle extends beyond language use. We can conceive of this mutual effort as a kind of strategic principle shared across language users. Relevance Theory dovetails nicely with a form–function account of vocal signaling as speakers strive to optimize relevance by following principled connections between sound and meaning. The form–function approach has proven very useful for understanding the relationships between vocal acoustics and vocalizers' intent in humans and nonhuman animals (Owren & Rendall, 2001; Pisanski et al., 2022). Acoustic structure (i.e., the form) in a vocal signal can be explained to a great extent by its communicative function(s). One classic example in animal signaling is alerting components (i.e., high-energy, noticeable onsets in certain signals) that grab receivers' attention. For a human example, consider a loud prohibitive yell to rapidly interrupt a target's behavior (e.g., "NO!! …stop that"). A good deal of infant-directed (ID) speech can be understood in this framework, such as how soothing vocal sounds directly downregulate infants' affective state, or energetic, melodic speech can encourage and incite positive emotions and behavior (Bryant & Barrett, 2007; Fernald, 1992). We can conceptualize ID speech as a strategic means by which caretakers manage dependents' emotions and behaviors. Similarly, infants can engage in strategic manipulation of caretakers through their vocal modulations, such as using the aversive sound of crying to induce crying cessation (Bryant, 2021). These vocal strategies by caretakers and infants, shaped by evolution, constitute a means by which the respective vocalizers make their communicative

intent relevant for their target audiences, and particular acoustic deliveries optimize their effectiveness (Wilson & Wharton, 2006).

This overall approach is powerful for understanding how prosody in indirect speech such as verbal irony can manifest itself – it is highly variable, but certainly not random. Ironic speech acts contain linguistic surface structure (i.e., the actual words used) and an attitude toward an attributed proposition contained in those words. The prosodic form will often be shaped by the expressed attitude or emotion superimposed onto the surface speech. For example, when Herb informs his guest that he is completely out of alcohol, but his friend wants more to drink, she might exclaim "Great!" with a contradictory, angry prosody. The pretense of anger drives the acoustic form, explicable through a form–function account of why anger has rather specific and universal acoustic manifestations (i.e., high arousal causing increased pitch and loudness, and possible vocal nonlinearities resulting in a harsh, distorted sound). Research using actors, as well as spontaneous speakers, has shown that the negative affect typical in vocal emotions like anger is common in sarcastic portrayals (Bryant & Fox Tree, 2005; Cheang & Pell, 2008). But alternatively, a response could be "Greaaaat!" with an exaggerated positive affect added through the prosody (e.g., increased pitch with an exaggerated pitch melody, combined with a prolongation) that could enhance the pretended positive message contained in the words. In either case, the prosodic form is shaped by the emotional function.

Consider another example. Imagine a person is told that they will have to endure some event that they would rather not experience (e.g., a visit with disliked in-laws). They can express their negative attitude by exclaiming ironically "I can't wait!" In many contexts, of course, vocal signaling is not necessary for an audience to understand ironic intent. But aside from producing verbal irony in a way that increases the likelihood of it being understood, it can also be produced with flair to enhance its effect, which can include humor and play (Gibbs et al., 2014). The conventionalized expression "I can't wait" conveys eagerness (i.e., a positive stance). There are many possible strategies that a speaker can implement when using this expression ironically. A genuinely excited pronouncement of "I can't wait" would likely have some combination of the preceding acoustic correlates of positive affect, but none in great excess. By overdoing one or more of these positively associated acoustic variables, the utterance will start to sound sarcastic, but the conveyed affect maintains the same valence as the literal meaning of the utterance. A speaker could, however, easily alter the vocalized valence, and superimpose various prosodic features of negative affect, including a disinterested, drawn-out contour (e.g., lowered pitch and pitch variability, lowered loudness, and slowed articulation), disinterested and terse (fast speech rate), or overacted, dark enthusiasm (lowered pitch with increased pitch variability, increased loudness, slowed, and rough voice quality). Figure 12.2 shows spectrograms of a single speaker producing a relatively neutral version of "I can't wait" along with seven alternative versions all conveying ironic intent.

As we have discussed, the interaction of a vocal performance with a verbal expression, or even one lexical item in an utterance, can strategically convey

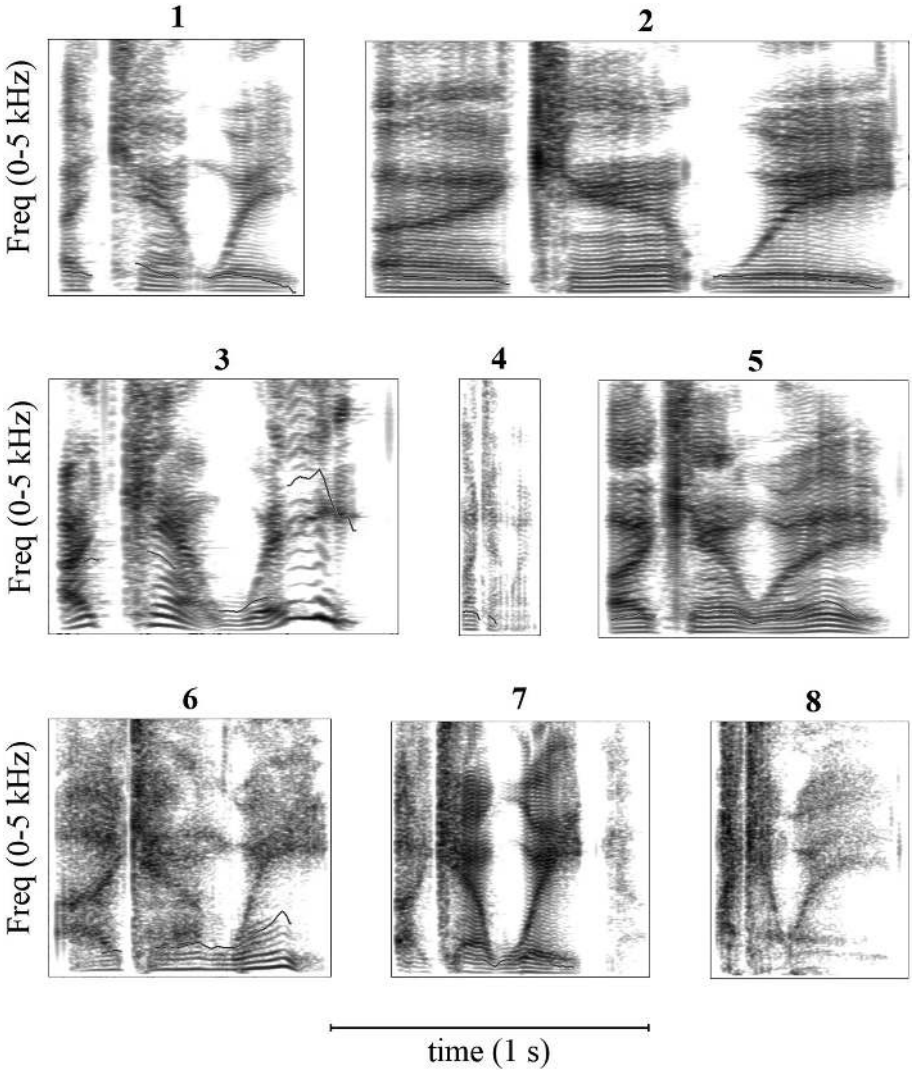

Figure 12.2 *Narrowband fast Fourier transform (FFT) spectrograms of eight versions of the ironic remark "I can't wait" produced with varying prosodic features conveying different affective portrayals of pretended enthusiasm: (1) baseline, (2) nonenthusiastic, (3) over-enthusiasm, (4) nonenthusiastic, (5) over-enthusiasm, (6) over-enthusiasm, (7) over-enthusiasm, and (8) over-enthusiasm. Solid lines indicate fundamental frequency contour (f_o). Sound files of these utterances available online at https://osf.io/wevygl.*

indirect meaning such as irony. But overall, communicative strategies manifest themselves at multiple levels, including procedural and conceptual meaning (Blakemore, 1987). Generally, vocal signals constitute procedural meaning in that they direct the inferential process toward particular semantic messages in language, but prosody can also communicate specific messages without

language (conceptual meaning), such as in nonlinguistic vocalizations (Anikin et al., 2018). A recent analysis of laughter in conversation makes a similar point (Mazzocconi et al., 2020).

Conclusion

Conversationalists use language to engage in a dynamical, interactive, multimodal coordination that involves ostensive communication at its core. This kind of interaction occurs in a complex social milieu where people have awareness of others' immediate intentions and long-term overarching goals. Interlocutors mutually aim to be optimally relevant for their audiences, and enact various strategies to achieve this in the service of their own agenda. Rather than think of the many indirect communicative practices we use as ornaments or secondary effects, we should consider them as unavoidable emergent properties of language and cognition (Bryant, 2012; Gibbs, 1994). Language evolved to work this way (Scott-Phillips, 2014). Verbal irony comprises a culturally evolved subset of possible pragmatic strategies that incorporates social cognition – a type of indirect speech that relies crucially on mindreading and complex intention communication.

The voice operates at multiple production levels, each with associated constraints. Communicative clarity at one level can constrain production at another, with multiple systems aligning their efforts in the service of making utterances relevant in highly context-specific ways (Bryant & Fox Tree, 2005; Fujisaki, 1983). Prosodic production involves the patterning of acoustic information at multiple levels, with local prosody operating on segmental information and global prosody working suprasegmentally. A speaker can use prosody to focus on a single word, or produce an emotional tune that signals specific indirect meaning. Speakers can also manipulate their voices in many ways to achieve communicative effects, such as altering the voice quality to imitate other speakers, or adopt fictional personas. Additionally, people use a variety of nonlinguistic vocalizations such as laughter, sighs, moans, and gasps to help convey ironic meaning. Laughter in particular is a frequently used vocal strategy that is likely derived from an evolved play function related to play vocalizations present in many extant nonhuman species.

Because speech is the predominant medium of conversational interaction, vocal behavior plays a crucial role in how interactions unfold. Relevance Theory and a form–function perspective on vocal signaling together provide a potent theoretical framework to help us understand how people use their voices to communicate in general, including in specific cases of indirect speech like verbal irony. Traditionally, there has been a focus on an ironic tone of voice that has neglected the bigger picture of how language, speech, and social behavior function. I have argued that speakers engage in strategic practices of simultaneously conveying intended meanings, at some level with selfish motives such as maintaining plausible deniability, while operating in a context of cooperative coordination with co-interlocutors for efficient communication.

Like any communication device rooted in language, verbal irony has deep connections to underlying cognitive structure. The strategies that speakers adopt to achieve their communicative goals have culturally evolved, and these evolutionary innovations are constrained and shaped by fundamental aspects of thought that exist independently of particular languages or linguistic abilities. One can consider verbal irony as a culturally evolved phenomenon that is motivated by a cognitive sensitivity to aspects of the environment where actions and event outcomes are in apparent contradiction and might be described as ironic (Lucariello, 1994). That is, situational irony constitutes a cultural attractor for pragmatic mechanisms that employ verbal irony as a rhetorical device. Sperber (1996) introduced the notion of cultural attractors, which has been developed extensively in many ways since (see Heintz et al., 2019). Culturally evolved phenomena are selected, in part, on the basis of a form–fit relationship between environmental demands and aspects of our cognition. For example, music exploits a number of perceptual and behavioral traits that evolved for other reasons (e.g., auditory scene analysis, vocal emotion, tool making), and its continued evolutionary refinement and transformation is made possible by its reliance on these underlying mechanisms (Bryant, 2013). We are attracted to certain phenomena in the environment due to our predispositions, and this attraction has consequences for whether particular cultural artifacts are (or are not) transformed and ultimately maintained at the population level.

By this view, verbal irony exploits a variety of cognitive phenomena that cause it to proliferate culturally. For instance, the capacity for metarepresentation and mindreading allows speakers to separate a direct perspective from an implied perspective, affording an ability to deliver commentary on an idea without stating it outright. This aspect of thought likely provides an attractor space for many pragmatic mechanisms, including verbal irony, making many pragmatic communicative mechanisms emergent properties of the interface between mindreading and language. Moreover, causal-reasoning mechanisms that sensitize people to violations of expected outcomes prime us for ironic reasoning. Mindreading and ironic logic thus emerge in language use, providing the fodder for the cultural evolution of tropes exploiting these systems. Verbal irony, consequently, has features that require metarepresentational ability to understand, and incorporates, at least minimally, elements of situational irony that pervade our everyday experiences in the world.

Listeners must integrate multimodal information to access meaning, and formal models of pragmatic reasoning represent recent strides toward developing coherent computational accounts of how this unfolds. The complete understanding of specific connections of voice and meaning can only be achieved by recognizing overarching principles that transcend any particular phenomenon. Here, I have tried to lay out how vocal strategies of verbal irony relate to the larger problem of how people use their voices to solve myriad communicative problems of everyday interaction. Through an interdisciplinary approach, we will one day be able to properly characterize the nature of human communication, including the minutiae of irony and the voice. I can't wait!

Greg Bryant

References

Aguert, M. (2022). Paraverbal expression of verbal irony: Vocal cues matter and facial cues even more. *Journal of Nonverbal Behavior*, 46(1), 45–70.

Anikin, A., Bååth, R., & Persson, T. (2018). Human nonlinguistic vocal repertoire: Call types and their meaning. *Journal of Nonverbal Behavior*, 42(1), 53–80.

Anolli, L., Ciceri, R., & Infantino, M. G. (2002). From "blame by praise" to "praise by blame": Analysis of vocal patterns in ironic communication. *International Journal of Psychology*, *37*(5), 266–276.

Attardo, S., Eisterhold, J., Hay, J., & Poggi, I. (2003). Multimodal markers of irony and sarcasm. *Humor: International Journal of Humor Research*, 16, 243–260.

Blakemore, D. (1987). *Semantic constraints on relevance*. Blackwell.

Bryant, G. A. (2010). Prosodic contrasts in ironic speech. *Discourse Processes*, *47*(7), 545–566.

Bryant, G. A. (2011). Verbal irony in the wild. *Pragmatics and Cognition*, *19*(2), 291–309.

Bryant, G. A. (2012). Is verbal irony special? *Language and Linguistics Compass*, *6*(11), 673–685.

Bryant, G. A. (2013). Animal signals and emotion in music: Coordinating affect across groups. *Frontiers in Psychology*, 4, 990.

Bryant, G. A. (2020). Evolution, structure, and functions of human laughter. In K. Floyd & R. Weber (Eds.), *Handbook of communication science and biology* (pp. 63–77). Routledge.

Bryant, G. A. (2021). The evolution of human vocal emotion. *Emotion Review*, *13*(1), 25–33.

Bryant, G. A., & Barrett, H. C. (2007). Recognizing intentions in infant-directed speech: Evidence for universals. *Psychological Science*, *18*(8), 746–751.

Bryant, G. A., & Fox Tree, J. E. (2002). Recognizing verbal irony in spontaneous speech. *Metaphor and Symbol*, 17(2), 99–117.

Bryant, G. A., & Fox Tree, J. E. (2005). Is there an ironic tone of voice? *Language and Speech*, 48(3), 257–277.

Bryant, G. A., & Nagy, M. (in preparation). *Vocal signals of indirect speech*.

Burgers, C., van Mulken, M., and Schellens, P. J. (2011). Finding irony: An introduction of the Verbal Irony Procedure (VIP). *Metaphor and Symbol*, 26(3), 186–205.

Burghardt, G. (2005). *The genesis of animal play*. MIT Press.

Cheang, H. S., & Pell, M. D. (2008). The sound of sarcasm. *Speech Communication*, *50*(5), 366–381.

Cheang, H. S., & Pell, M. D. (2009). Acoustic markers of sarcasm in Cantonese and English. *Journal of the Acoustical Society of America*, 126(3), 1394–1405.

Chen, A., & Boves, L. (2018). What's in a word: Sounding sarcastic in British English. *Journal of the International Phonetic Association*, 48(1), 57–76.

Clark, H. H., & Gerrig, R. J. (1984). On the pretense theory of irony. *Journal of Experimental Psychology: General*, 113(1), 121–126.

Clift, R. (1999). Irony in conversation. *Language in Society*, *28*(4), 523–553.

Cole, J. (2015). Prosody in context: A review. *Language, Cognition and Neuroscience*, *30*(1–2), 1–31.

Colston, H. L. (1997). Salting a wound or sugaring a pill: The pragmatic functions of ironic criticism. *Discourse Processes*, 23(1), 25–45.

Colston, H. L. (2000). On necessary conditions for verbal irony comprehension. *Pragmatics & Cognition*, *8*(2), 277–324.

Cowen, A. S., Laukka, P., Elfenbein, H. A., Liu, R., & Keltner, D. (2019). The primacy of categories in the recognition of 12 emotions in speech prosody across two cultures. *Nature Human Behaviour*, *3*(4), 369–382.

Cutler, A. (1974). On saying what you mean without meaning what you say. In M. Lagaly, R. Fox, & A. Bruck (Eds.), *Papers from the tenth regional meeting of the Chicago Linguistics Society* (pp. 117–127). Department of Linguistics, University of Chicago.

Dews, S., & Winner, E. (1995). Muting the meaning: A social function of irony. *Metaphor and Symbol*, *10*(1), 3–19.

Eady, S. J., & Cooper, W. E. (1986). Speech intonation and focus location in matched statements and questions. *Journal of the Acoustical Society of America*, 80(2), 402–415.

Eisterhold, J., Attardo, S., & Boxer, D. (2006). Reactions to irony in discourse: Evidence for the least disruption principle. *Journal of Pragmatics*, *38*(8), 1239–1256.

Fernald, A. (1992). Human maternal vocalizations to infants as biologically relevant signals. In J. Barkow, L. Cosmides, & J. Tooby (Eds.), *The adapted mind: Evolutionary psychology and the generation of culture* (pp. 391–428). Oxford University Press.

Féry, C. (2013). Focus as prosodic alignment. *Natural Language & Linguistic Theory*, *31*(3), 683–734.

Fish, K., Rothermich, K., & Pell, M. D. (2017). The sound of (in)sincerity. *Journal of Pragmatics*, *121*, 147–161.

Fujisaki, H. (1983). Dynamic characteristics of voice fundamental frequency in speech and singing. In P. MacNeilage (Ed.), *The production of speech* (pp. 39–55). Springer-Verlag.

Gibbs, R. W. (1994). *The poetics of mind: Figurative thought, language, and understanding*. Cambridge University Press.

Gibbs, R. W. (2000). Irony in talk among friends. *Metaphor and Symbol*, *15*(1–2), 5–27.

Gibbs, R. W., Bryant, G. A., & Colston, H. L. (2014). Where is the humor in verbal irony? *Humor: International Journal of Humor Research*, 27(4), 575–595.

González-Fuente, S., Prieto, P., & Noveck, I. (2016). A fine-grained analysis of the acoustic cues involved in verbal irony recognition in French. *Speech Prosody*, *2016*, 902–906.

Goodman, N. D., & Frank, M. C. (2016). Pragmatic language interpretation as probabilistic inference. *Trends in Cognitive Sciences*, 20(11), 818–829.

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. Morgan (Eds.), *Syntax and semantics. Vol. 3: Speech acts* (pp. 41–58). Academic Press.

Heintz, C., Blancke, S., & Scott-Phillips, T. (2019). Methods for studying cultural attraction. *Evolutionary Anthropology: Issues, News, and Reviews*, 28(1), 18–20.

Jun, S. (2019). (Ed.). *Prosodic typology II: The phonology of intonation and phrasing*. Oxford University Press

Kotthoff, H. (2003). Responding to irony in different contexts: On cognition in conversation. *Journal of Pragmatics*, 35(9), 1387–1411.

Kreuz, R. J., & Roberts, R. M. (1995). Two cues for verbal irony: Hyperbole and the ironic tone of voice. *Metaphor and Symbol*, 10(1), 21–31.

Ladd, D. R. (2008). *Intonational phonology*. Cambridge University Press.

Lieberman, M. D. (2013). *Social: Why our brains are wired to connect*. Oxford University Press.

Lucariello, J. (1994). Situational irony: A concept of events gone awry. *Journal of Experimental Psychology: General*, 123(2), 129.

Mazzocconi, C., Tian, Y., & Ginzburg, J. (2020). What's your laughter doing there? A taxonomy of the pragmatic function of laughter. *IEEE Trans. Affective Computing*, 1–19.

McRoberts, G. W., Studdert-Kennedy, M., & Shankweiler, D. P. (1995). The role of fundamental frequency in signaling linguistic stress and affect: Evidence for a dissociation. *Perception & Psychophysics*, *57*(2), 159–174.

Muecke, D. C. (1978). Irony markers. *Poetics*, *7*(4), 363–375.

Owren, M. J., & Rendall, D. (2001). Sound on the rebound: Bringing form and function back to the forefront in understanding nonhuman primate vocal signaling. *Evolutionary Anthropology: Issues, News, and Reviews: Issues, News, and Reviews*, 10(2), 58–71.

Pell, M. D. (2001). Influence of emotion and focus location on prosody in matched statements and questions. *Journal of the Acoustical Society of America*, 109(4), 1668–1680.

Pell, M. D., Monetta, L., Paulmann, S., & Kotz, S. A. (2009). Recognizing emotions in a foreign language. *Journal of Nonverbal Behavior*, 33(2), 107–120.

Pinker, S., Nowak, M. A., & Lee, J. J. (2008). The logic of indirect speech. *Proceedings of the National Academy of Sciences*, 105(3), 833–838.

Pisanski, K., & Bryant, G. A. (2019). The evolution of voice perception. In N. S. Eidsheim & K. L. Meizel (Eds.), *Oxford handbook of voice studies* (pp. 269–300). Oxford University Press.

Pisanski, K., Bryant, G. A., Cornec, C., Anikin, A., & Reby, D. (2022). Form follows function in human nonverbal vocalisations. *Ethology Ecology & Evolution*, *34*(3), 303–321.

Provine, R. R. (2000). *Laughter: A scientific investigation*. Penguin Books.

Rockwell, P. (2000). Lower, slower, louder: Vocal cues of sarcasm. *Journal of Psycholinguistic Research*, *29*(5), 483–495.

Scott-Phillips, T. (2014). *Speaking our minds: Why human communication is different, and how language evolved to make it special*. Macmillan International Higher Education.

Sperber, D. (1984). Verbal irony: Pretense or echoic mention? *Journal of Experimental Psychology: General*, 113, 130–136.

Sperber, D. (1996). *Explaining culture: A naturalistic approach*. Cambridge University Press.

Sperber, D., & Wilson, D. (1981). Irony and the use-mention distinction. In P. Cole (Ed.), *Radical pragmatics* (pp. 295–318). Academic Press.

Sperber, D., & Wilson, D. (1995). *Relevance: Communication and cognition* (2nd ed.). Blackwell.

Trivers, R. (2000). The elements of a scientific theory of self-deception. *Annals of the New York Academy of Sciences*, 907(1), 114–131.

Wilson, D., & Wharton, T. (2006). Relevance and prosody. *Journal of Pragmatics*, *38*(10), 1559–1579.

Winkler, S. L., & Bryant, G. A. (2021). Play vocalisations and human laughter: A comparative review. *Bioacoustics*, *30*(5), 499–526.