A Massive Sarcastic Robot: What a Great Idea!

Two Approaches to the Computational Generation of Irony

Tony Veale
School of Computer Science and Informatics
University College Dublin, Belfield D4, Ireland.
Tony.Veale@UCD.ie

Abstract

Irony is a versatile seasoning for language that is just as useful for sugaring insults as for salting compliments. It is a flavoring that adds bite to much of our online interaction, making its computational analysis – recognition *and* understanding – a necessity for affective language processing. If the computational *generation* of irony is a harder sell, it is perhaps because this mode of communication is so often vexing for humans too. However, an artificial fluency with irony is as desirable as fluency in *any* form of creative language, from metaphor and analogy to humour and persuasive argumentation. We explore two distinct approaches to irony generation in this paper: knowledge-based generation *ab initio*, and a more shallow approach we name 'mere *re*-generation.' We consider the relative merits of each, conduct a user evaluation, and demonstrate some practical uses.

The Devil's Seasoning

To communicate with irony is to talk with a forked tongue. Yet while the need for the computational analysis of irony is strong, since so much of our online language is playfully or scornfully sardonic (Hao & Veale, 2010; Reyes *et al*., 2013; Ghosh & Veale, 2017), the case for computationally *generating* irony is much less compelling. However, there are practical reasons for granting our machines a fluency in this challenging mode of communication. Irony requires a delicate blending of mental spaces to concisely express a double-edged attitude to a failed expectation: we highlight an expectation, act as though it were fulfilled, and criticize its lack of fulfilment, all in a single breath. So a generative model of irony is also, by necessity, a model of conceptual blending (Fauconnier & Turner, 2002), one that lends not just concision, but creativity too, to a machine's outputs.

An ironic statement can be a most charming disguise for a conceptual conflict, though the charm of the disguise and the profundity of the conflict will vary from one speaker to another – skill is a factor, after all – and with the obviousness of the context. Irony is a subtle form of sarcasm (conversely, sarcasm is a vulgarized form of irony) in which we believe in what we say but not *how* we say it. So when we use the norms of flattery to sweeten an insult, or the norms of criticism to sour a compliment, we intend an audience to appreciate the seasoning but to look beyond it too, to grasp our deeper meaning and our ambivalent attitude towards it. Ironies, like metaphors, are allusive (if sometimes *elusive*) products of the imagination that appeal to the imaginations of others. But these products are not built in a vacuum; an ironic worldview makes a critical claim about something in the world that others can see and evaluate for themselves. For a machine to generate an ironic observation, it needs knowledge of how things are and how they should be, and it needs the ability to frame the gap between the two in a pithily suggestive fashion. So, as befits the duality of irony, we explore *two* approaches to its generation in this paper.

The first is a knowledge-based approach that explicitly models that which is expected and that which is observed, so as to juxtapose them both in the same tweet-sized text. We use *dis*analogy as the unifying conceit to maximize the dissonance between the two perspectives, but the resulting text also supports other markup strategies to signal ironic intent to an audience. While the most obvious applications of machine-generated irony may well be human-computer interfaces, a less obvious, but no less useful, application is the generation of well-controlled test-data for experiments into the human appreciation of irony. Our evaluation of this parameterized approach shows how linguists can better understand how we as humans process irony, by presenting human subjects with the products of machine creativity.

How do we humans acquire our sense of the ironic? This sense is not something we are born with, but something we cultivate over time, such as via continuous exposure to the ironic stylings of others. A machine can also be exposed to ironic language to learn its signature qualities, with social media platforms such as Twitter making it easy to retrieve large amounts of texts that are self-annotated with the hash tags *#irony*, *#sarcasm* and *#yeahright*. The overt tagging of verbal pretence is not a new phenomenon, even if Twitter elevates the practice to new levels of explicitness. Speakers have always used subtle cues to signal their ironic intent. By harvesting a broad spectrum of cued ironic texts from the web, a machine can build a large case-base of attested examples to be reused wholesale, or recycled with novel variations, as parts of its own future utterances. Our second approach does not analyse these web examples in any great depth, but pursues a philosophy we dub *mere regeneration* to find new uses and framings for old word combinations. The qualities that make one juxtaposition of words or ideas seem more poetic, more beautiful, more ridiculous or more hilarious than others may not always defy logical scrutiny, but for all practical intents they remain ineffable for now. Our machines should thus do what most humans do at one time or another: reuse the resonant combinations of words that have worked for others and claim them for themselves.

We present each of these approaches in the sections to follow, beginning with a review of related work and ideas in the next. After an empirical evaluation of the first, and a discussion on how the second supports the rapid development of humorous CC systems, the paper concludes with a discussion of the relative merits of each approach to irony.

Related Work and Ideas

To judge by the diversity of tweets that Twitter users tag with *#irony*, the general public operates with a somewhat diffuse understanding of what irony is. The popular view, and the most oversimplified, is that to speak ironically is to say one thing but to mean its opposite (Kierkegaard, 1841; Grice, 1978). But irony is a nuanced idea that demands just as nuanced a definition, and *irony-as-opposite* disappoints on several fronts: it is meaningful for just a subset of the utterances that speakers intuitively grasp as ironic; even in such cases, opposition accounts for just one aspect of the intended meaning; and even then, it is not always obvious how one can arrive at the opposite of an ironic statement. Recall a scene in the film *Amadeus* in which the composer Salieri has just premiered his new opera. When put on the spot for a positive response, Mozart shrewdly replies that "*When one hears such music, one can only think 'Salieri'!*" Mozart's words are criticism masked as flattery and Salieri suspects as much, though he cannot know for sure. But what is the opposite of Mozart's reply here – that one does *not* think of Salieri when hearing such music? No, what is inverted here is not what Mozart says but what he implies, "When one hears such [*lovely*] music, one can only think 'Salieri'!" His ironic meaning thus becomes "When one hears such [*unlovely*] music, one can only think 'Salieri'!"

To speak ironically then is to say one thing, insincerely imply the obvious, and intend something so different that it often amounts to the opposite of what is implied. This constitutes an act of *verbal pretence* (Clark & Gerrig, 1984) and *pragmatic insincerity* (Kumon-Nakamura *et al*., 1995) that is designed to be penetrated by audiences. But if ironic statements are meant to be understood as such, they spark a conflict of implications between the default and the non-default (Giora *et al.,* 2015), or the obvious and the creative, that audiences must somehow resolve for themselves. The context often determines how fraught this conflict will be, with contexts that are strongly supportive of an ironic interpretation nudging audiences to look past the obvious. In contexts that are equally supportive of the default and non-default interpretations, the audience is left – like Salieri – in a rather uncomfortable superposition of affective states.

In such cases, authors have a number of ways to nudge an audience toward the creative. Many ironic utterances are context-*external*, which is to say that the information one needs to discern ironic from non-ironic is found *outside* the utterance itself. Mozart's reply is a context-external irony. Many more are context-*internal*, insofar as an author bakes the necessary context into the utterance itself. For example, consider this classic image from *Farewell My Lovely* by Raymond Chandler: "He looked about as inconspicuous as a tarantula on a slice of angel food cake." The *He* is Moose Malloy, a hulking white brute who is newly-released from prison and in search of his faithless wife Velma in a black neighborhood of the city. Throwing discretion to the wind, the white Malloy stomps about town, terrifying the locals while sticking out like a large black spider on a white cake. The comparison is enough to alert readers that Malloy is the very opposite of inconspicuous. Yet note that the simile means more than "Malloy was very conspicuous indeed." It means "Malloy should have tried to be inconspicuous, but the dumb brute could not be subtle if he tried." We use irony to conflate perspectives *and* to criticize, all at once.

Besides building useful context into his simile, Chandler also prefaces the comparison with "about", a marker of imprecision that alerts readers to his insincere use of words. Hao & Veale (2010) use these markers of imprecision to harvest creative similes in bulk from the web. While the marker "about" is not reserved for ironic comparisons – that depends on whether the simile is intended as flattery or criticism – it *is* a reliable marker of linguistic creativity. As analyzed in Veale (2013), about-similes tend to use longer descriptions (or "vehicles") that constitute what Fishelov (1992) deems PS (poetic similes). He contrasts these with the NS (non-poetic) similes that pervade language, such as "clear as mud", "light as a feather" and "dry as a bone." PS similes use many of the same words as NS similes, but use them in striking juxtapositions that are memorable and sometimes hilarious, as in "as sophisticated as a zombie at a dinner party" and "as quiet as a cat in a blender." Taylor (1954) laboriously compiled a large corpus of PS similes that had become the stuff of proverb in California – such as "as useful as teats on a boar" – but markers such as "about" allow our machines to amass such corpora automatically.

Despite helpful markers such as "about", "almost" and "not exactly", irony tends much less to the formulaic than sarcasm. The latter does not sustain its verbal pretence for very long, nor does it leave its audience in much doubt as to the true intentions of a speaker. Sarcastic tweets such as the following are thus a commonplace on Twitter: "I love it when my 'friends' forget my birthday" and "Don't you just love it when your boss throws you under the bus?" Without the safety-net of face-to-face interaction, Twitter users are careful to signal their insincerity openly, as misunderstandings on social media can lead to public shaming. The fear of public rebuke is so strong that users routinely tag even the most formulaic sarcastic tweets with *#sarcasm*. So it is an easy matter to harvest large amounts of apt training data from Twitter, to train our machines to recognize a sarcastic attitude using supervised machine-learning techniques.

Statistical classifiers can use everything from the words themselves to their POS tags, bigram / trigram collocations and sentiment scores to discriminate sarcastic from non-sarcastic texts. Riloff *et al.* (2013) used the mixed emotions of sarcasm as a characteristic signature, and obtain good results on the short texts that are typical of Twitter. Reyes *at al*. (2013) used a broader basket of features, including symmetry, to identify both irony and wit more generally. Ghosh *et al*. (2013) focused not on irony or sarcasm detection but on the estimation of sentiment in figurative tweets that comprise ironic, sarcastic and metaphorical examples; their annotated corpus is frequently used as a training and test set for sarcasm detection. Ghosh & Veale (2015) first trained a neural network to recognize sarcasm, and later (Ghosh & Veale 2017) extended this network to integrate a model of sarcastic mood. When working with tweets, a machine has access to a timestamp for each, to author data, and to the timeline in which each was posted. Using the web service *AnalyzeWords* (Tausczik & Pennebaker, 2010) to perform a mood analysis of the prior 100 posts leading up to a given tweet, an extra 11 dimensions – including *anger*, *positivity*, *remoteness*, *worry* and *analyticity* – can be added as network inputs. Ghosh & Veale show that this personal context is as useful as the usage context of a tweet (i.e. the text to which it was posted in reply) in recognizing the user's pragmatic intent. Those authors also introduced another innovation to the detection of sarcasm: rather than use independent raters to annotate the training and test sets for sarcasm, they used a Twitterbot, @*SarcasmMagnet*, to contact the owner of each tweet directly, to obtain in real time the author's own statement of pragmatic intent.

These approaches still bring with them a concern about over-fitting. Are the features that prove to be most useful at detecting sarcasm and irony truly generic, or do they just happen to be the words that best separate the positive from the negative instances in a particular testset? These systems perform detection without ever striving for understanding, but we humans take a very different approach: to recognize an act of pragmatic insincerity, we first analyse its intent in terms of the meaning it might communicate to others. This analysis is crucial for the generation of irony, for a system cannot be ironic if it does not know what it intends to say or cannot know if it has faithfully conveyed that intention. Our best statistical models of detection are too shallow to be reversed to serve as models of generation, so we are still quite some way from a CC system that can accept "Your music sucks, Salieri!" as an input and generate as its output "When one hears such music, one can only think 'Salieri'!" With this dour prognosis in mind, we limit ourselves in the next section to a highly-structured expression of irony that machines can both generate and appreciate for themselves.

EPIC Fails

To give a machine a capacity for generating irony, we must first break its heart. For whatever else irony might be, and regardless of whether it is used to criticize (its main use) or to praise (a minority pastime), every ironic statement is an expression of disappointment. A machine whose sole job is to be ironic is a machine that must always be disappointed. Since disappointment results from a failed expectation, our ironic machine must thus possess a model of expectation.

We propose a simple model of property-oriented expectation, named EPIC, in which an expectation (E) predicts a property (P) of an instance (I) of a concept (C). Take the concept of a "party." An instance I – my birthday party, say – of this concept C carries with it one or more expectations (E) of the typical properties (P) of parties: so we expect I to be fun, to be entertaining and to be social. An expectation E fails if the expected property P cannot be asserted of I, and fails ostentatiously if we can instead assert its opposite, not-P. Even if E fails in a more subtle way, the task of the ironist is to exaggerate the truth for humour's sake. In this way, a failed expectation E_1 of I_1 concerning P can match an expectation E_2 of a non-salient concept C_2 that predicts not-P. So just as parties should to be fun and entertaining, we often expect lectures to be dull and boring. In failing to be fun, a party I_1 fulfils an expectation of lectures that few guests actually bring to a party. But by matching a failed expectation for P to an non-salient expectation for not-P, an ironist can dramatize the non-P of I_1 (an instance of C_1) by pretending that I_1 is an instance of C_2 that entails the expectation not-P, or can perhaps feign mock surprise to have attended I_2 (an instance of C_2) instead of I_1.

Even the most committed ironists spend only a tiny part of their lives being ironic. The rest of their time is dedicated to the stuff of everyday life: working, reading, shopping and interacting with others. A machine whose sole task is to be ironic is an oddity indeed, but how is it to acquire the expectations that we humans spend a lifetime developing? The answer, as in many NLP tasks, is the web. To acquire the expectations E that people bring to instances I of the concepts C, a machine can consider the adjectives "P" that adorn the word "C" in common usage. The Google ngrams database (Brants & Franz, 2006) provides a large inventory of frequent web collocations. Consider these 3-grams:

W1	W2	W3	*Web Count*
a	fun	party	*10060*
a	dull	party	*772*
an	entertaining	party	*161*
a	boring	lecture	*1882*
a	dull	lecture	*267*

Notice how "dull" is more used often, in absolute terms, to describe parties than lectures, so frequency alone is not a reliable indicator of expectation strength. Machines can use n-gram data to suggest apt candidates for property-oriented expectations, but they must look elsewhere to confirm their hypotheses. Since similes are linguistic constructions that take full advantage of conceptual expectations, a machine can determine whether P is a widely-held expectation for instances of C by looking for similes of the form "as P as a C." Veale (2012) shows how expectations that conform to the EPIC structure are harvested in bulk from the web by retrieving all matches for the wildcard query "as * as *."

Our machine shall also need some relational knowledge, to understand how others typically relate to the concepts C that are so-described with a property P. For instance, how do people relate to parties or lectures, and can we relate to instances of each in the same way? If so, an analogy can be constructed from the shared relationships. A relation is any triple <C_2 R C_1> linking two concepts in the abstract and two instances of those concepts in the specific. The query logs of web search engines are a good source of common-sense triples, since users expose their expectations of the world in the questions that they pose online. So when a pet owner asks "why do dogs chase cars" or "why do cats arch their backs" these questions assume that everyone else believes that <dogs chase cars> and <cats arch backs> too. Veale & Li (2011b) show how a large database of question-derived triples can be "milked" from the query continuations offered by Google. While its query log is private, when the engine suggests popular completions for partial queries it is effectively exposing recurring entries on that log.

The thwarted expectation in which an ironic utterance in rooted can take many forms. EPIC assumes that the expectation concerns a property P for concept C_1, but it can be extended to a concept C_2 by the relation <C_2 R C_1>. To highlight a failure to observe P of C_1, an ironist can compare C_1 to a C_3 for which not-P is expected, on the basis of a parallel relation <C_4 R C_3> and the analogy C_1:C_2::C_3:C_4. Since C_1 and C_3 are not so much compared as contrasted on the basis of a conflict between P and not-P, the juxtaposition is more disanalogy than analogy. In our example of parties and lectures. the disappointment of a failed event can be conveyed with irony with the following disanalogy:

Some hosts arrange "entertaining" parties the way presenters arrange boring lectures.

We can now appreciate the function of the shared relation R (in this case, *arrange*): it focuses the ironic charge of the disanalogy toward those who arrange the parties that fall so short of our expectations, in the same way that explosives experts shape their charges to explode in a given direction. By wrapping the expected property "entertaining" in ostentatious scare-quotes, the charge appears to be echoing a lie, a failed prediction that a speaker now mimics with ridicule. The "echoic mention" of an unwise prediction (Sperber & Wilson, 1981; Kreuz & Glucksberg, 1989) offers a way of elevating the veiled criticism of irony into open mockery. If the criticism were expressed on Twitter, a speaker might go so far as to append the hashtag *#irony*, as if to say "Isn't it ironic when …" The relative merits of these strategies – disanalogy, scare-quotes and overt tagging – for conveying an ironic worldview will be evaluated in the next section.

EPIC Succeeds

The success of an ironic utterance hinges on its capacity to highlight the failure of a reasonable expectation. As some are more successful in this regard than others, we need a gradated yardstick of success that goes beyond the binary. Notice that while EPIC predicates success on the inference of not-P in a context that implies P, it does not subscribe to an unnuanced irony-as-opposition view. Instead, it assumes that irony is successful when audiences shift their expectations of C from P toward not-P either in whole or in part. A successful ironic utterance may leave audiences with the mixed feeling that instances of C occupy a middle-ground between P and not-P that conforms to neither extreme; for example, that "many parties that promise entertainment are only ever entertaining to the people that host them." While we cannot measure nuanced feelings like this, Valitutti & Veale (2017) propose a convenient proxy: if P is a positive property and not-P is a negative property, then an ironic statement in the EPIC mold is successful to the extent that audiences downshift their mean rating of P's positivity in the context of the irony. We can expect, for instance, that the mean positivity of the property "entertaining" in a null context is higher than its mean rating in the context of a disanalogy that lends the word a halo of disappointment.

This gradated downshifting view permits us to measure success for irony generation overall, as well as the relative contribution of our different strategies – disanalogy, scare-quotes and overt tagging – to this success. We conduct a crowd-sourced evaluation using the platform *CrowdFlower* in which anonymous judges are each paid a small sum to rate the positivity of focal words in the ironic utterances constructed using EPIC. The focal word in each case is the property P, or in other words the adjective that is placed in scare-quotes. We use our generative system to generate 80 distinct ironic utterances, with the same structure as our party/lecture example, around a different focal property in each case. Each test instance exploits an expectation E for a positive property P that, we expect, is shifted toward a negative evaluation by the use of a disanalogy with another expectation E' for not-P. Here is one such test instance:

#irony: *When "cultured" gentlemen pursue ladies the way feral predators pursue prey*.

This is the fully-loaded version of the output, including the disanalogy, the scare-quotes and the overt tag. A number of other variants can be generated by ablating one or more features, and by asking judges to rate alternate variants of the same observation about a focal property, we can tease out the relative impact of each feature to the downshift. We label each variant as shown in the following examples:

BASE:
Cultured gentlemen pursue ladies

BASE+QUOTE:
"*Cultured" gentlemen pursue ladies*

BASE+COMP (disanalogy)
Cultured gentlemen pursue ladies the way feral predators pursue prey

BASE+QUOTE+COMP:
"*Cultured" gentlemen pursue ladies the way feral predators pursue prey*

BASE+QUOTE+COMP+HASH:
#Irony: "*cultured" gentlemen pursue ladies the way feral predators pursue prey*

We provide alternate variants of the same utterance to different judges, and ask each to estimate the positivity of the focal word on a scale from +1.0 (most positive) to -1.0 (most negative). We elicit ten ratings per utterance variant and then calculate the mean positivity rating for each. But to appreciate the extent of the ironic shift, we need to know how judges would rate these focal words in a null context, free of the baleful influence of the ironic utterance.

In another CrowdFlower experiment, one that is actually conducted prior to the one above, we do precisely this. We provide the 80 focal properties from the 80 automatically-generated utterances – words such as "entertaining" and "civilized" and "smart" and "creative" – and ask judges to rate their overall positivity on the same +1.0 to -1.0 scale. The mean ratings provide an estimate of the positivity of the words in their primary dictionary senses. We can now calculate the mean shift in positivity caused by an ironic utterance; the means are displayed in Table 1 below, with standard deviations in parentheses.

Structural Variant	*Mean Positivity*
BASE	0.51 (*SD* 0.38)
BASE+QUOTE	0.41 (*SD* 0.46)
BASE+COMP	0.29 (*SD* 0.49)
BASE+QUOTE+COMP	0.20 (*SD* 0.54)

Table 1. *Mean positivity of the focal words in ironic utterances with different structural variants. All differences between conditions are significant at the p < .001 level.*

As shown in Table 1, each successive feature increases the mean downshift in perceived positivity of a focal word P and its associated expectation E, with disanalogy offering the most forceful shift into negative territory. We can ask how often an utterance succeeds in not just diminishing the positivity of a focal word but in making it appear negative to an audience. Table 2 reports how likely the focal word is to be seen as positive overall by raters.

Structural Variant	*Positive Likelihood*
BASE	0.91 (*SD* 0.15)
BASE+QUOTE	0.82 (*SD* 0.13)
BASE+COMP	0.75 (*SD* 0.15)
BASE+QUOTE+COMP	0.64 (*SD* 0.16)

Table 2. *Likelihood that a focal word is viewed as positive rather than negative in different structural conditions.*

Again these show that disanalogy has a greater impact than scare-quotes on the upending of perceived sentiment, while combining *both* features yields a larger impact still. But the experiments also point to a negative finding not shown in Tables 1 and 2: overt marking with *#irony* has no discernible impact on utterances that already use scare-quotes and disanalogy, and has far less impact than either of those variants when it is used without them. It is one thing to explicitly announce an ironic mindset, and quite another to seed it effectively (and affectively) in the minds of an audience.

Mere Re-Generation

These tightly-organized utterances add nuance to the irony-as-opposition debate by effectively creating an ambivalent middle ground between an expected property P of C and its negation, not-P. But they do this by appealing to the experience of an audience rather than to its imagination. It takes experience – of parties and lectures, for example – to appreciate how our expectations can be fulfilled or thwarted. But no strange new concepts are introduced in these utterances, and no category boundaries are challenged. Instead, concepts are used in their simplest guise. In contrast, the "about" similes harvested by Hao & Veale (2010) and analyzed in Veale (2013) offer a surfeit of vivid detail to help us visualize a concept. In these similes we are told not just of parties but of *grunge* parties, *frat* parties, *hen* parties, *stag* parties, *beach* parties and *tea* parties. These events attract equally vivid guests: *wasps* at a tea party, a *skunk* at a lawn party, a *zombie* at a dinner party, a *teetotaller* at a frat party and even, absurdly, a *jackboot* at a testicle party. We also hear not just of lectures, but of six hour lectures on mahogany, advanced lectures on theoretical physics by Stephen Hawking and 40-minute lectures on pockets! The humour resides as much in the detail of C_1 as it does in any juxtaposition between C_1 and another concept C_2. Or, rather, the concept C_1 is already a vivid mix of ideas.

Veale (2011a) presented a system, the *Jigsaw Bard*, that repurposes n-gram collocations as descriptive vehicles for novel similes. For instance, the Google 2-gram "robot fish" names a family of aquatic drones, but, as evidenced by the stock similes "as cold as a fish" and "as cold as a robot," it might also describe a person who is emotionally cold. Our words carry a myriad unspoken constraints that are grasped only by fluent speakers, so the *Bard* sidesteps the challenges of building its imaginative word combinations *ab initio*. Rather, it uses a simple rule for locating its *objets trouvés* in web ngrams: a bigram "$W_1 W_2$" suggests how a concept combination C_1: C_2 for which the system already possesses the stock similes "as P as a W_1" and "as P as a W_2" can be repurposed as the novel simile "as P as a W_1 W_2." Shared expectations of P thus yield unified similes for P. Now, W_1 and W_2 are part of the *Bard*'s creative vocabulary by virtue of already serving as vehicles in its library of stock similes. But what if we reuse the vehicles from our *about* similes, which tend to be longer and more vivid, in the same way?

Many of those vehicles are inherently ridiculous. Just as irony is more than mere opposition, the ridiculous is more than mere absurdity. It occupies a place between the absurd and the impossible where our reaction is one of laughter or horror rather than puzzlement or stupefaction. It marks out another possible world that is out of joint with this one. So the similes of our *about* corpus speak of a dog in a sweater (insightful), pants in a nudist colony (necessary), a nun at a Reggae festival (inconspicuous) or a 10-ton rock in a canoe (useful). Each is the product of a personal sense of humour that can be as tasteless as leopard skin pants at a funeral or as welcome as a fart in a spacesuit. Each composition is a vivid conceptual blend (Fauconnier & Turner, 2002) that unites disparate ideas to spark emergent inferences. Since a dog in a sweater offers a surface imitation of human intelligence, the blend ironically undercuts – with some inference – a *pseudo*-intellectual who merely dresses for the part.

By harvesting a large corpus of *about* similes and their ridiculous mental images from the web, we can provide our machine with the rudiments of a composite sense of humour, a singular comedic voice formed out of the multitude. This corpus of ironic blends is also a comprehensive database of EPIC fails, which is to say, failures of expectations about properties that are vividly painted on a grand scale. We may try to dissect these failures into their individual parts, to take the dog out of its sweater and the nun out of her Reggae festival, so that an ironic machine can recombine the parts in new ways; perhaps by putting the nun in the spacesuit, the dog in the festival and the fart in the sweater. But the unspoken logic of the ridiculous that dictates how irony and humour emerge is unlikely to carry across to the new combinations. Those leopard skin pants may not seem so tasteless at a Reggae festival, nor a dog so conspicuous, and it is the air seal on a full-body spacesuit that makes the smell so much more unwelcome there than in a sweater. As Veale (2015) argues, jokes are compressed thought experiments, and humorous blends such as these can rely just as much on our physical intuitions as a conundrum in physics. We can no more chop up these blends and recombine their parts to generate a new one that is just as witty than we can chop up a science text to propose new theories in science.

The language may be robust at times, but the humour of these EPIC fails is fragile indeed. If an ironic machine is to exploit it reliably, it must make as few changes to possible. So the simplest and most reliable strategy is to reuse each blend in its entirety. Suppose we want to ridicule the lack of insight of a scientist or a reporter or of any kind of critic that speaks with authority. Our machine can retrieve from its database a blend that is ironically associated with the property *insightful*, such as our dog in a sweater, a plastic cap, a shaving foam commercial, a myopic mole in a sack, a college freshman essay, gravel, a rock, a fortune cookie, or a child writing home from summer camp, and attach this blend to our target. This mere *re*-generation is more effective than it is creative, yet it enables the rapid construction of CC systems like chatbots and interactive Twitterbots. Consider the political satire bot *@TrumpScuttleBot*, which offers a knowledge-based parody of a person whose tweets regularly flirt with the vulgar and the ridiculous. Rather than slice and dice the man's own tweets into a statistical gumbo of prejudice and provocation, as done by bots such as *@DeepDrumpf*, our parody bot works from first principles to create analogies and metaphors and whimsical political comparisons. For example, the bot often compares itself to other leaders, using EPIC to suggest the kinds of people that we expect to show leadership, as in this tweet:

> I come from a long line of HUGE demagogues who led a long line of TREMENDOUS mobs. The lying media don't tell you THAT! #MAGA

The satirical humour of the bot goes hand in hand with its own ineptitude, which is gleefully projected onto its target. Demagogues are leaders, yes, but not good leaders, and it is poor manners indeed to paint one's voters as a mob. But the bot also paints savage caricatures of its stock enemies:

> Those REGRETTABLE NBC hacks are as friendly as a rabid pit bull challenging all comers to a choice cut of steak! #MAGA

As the bot does not expect enemies to be *friendly*, it pulls out an ironic *about* simile. Other expectations are packaged as self-promotion, to highlight some key recurring themes:

> TRUMP University will re-open to offer TERRIFIC Degrees in Virtue to AMERICAN philosophers. Foreign philosophers are about as ethical as the tobacco industry! #MAGA

Echoing its target's distinctive language patterns, the bot makes frequent use of generic descriptors such as HUGE, TERRIFIC, SAD, and so on. When it talks about a specific concept C, such as *philosopher*, it can also bring specific EPIC expectations to bear. As the philosophers who study virtue are expected to be ethical, it looks to its database of blends to suggest an ironic comparison with Big Tobacco. The mere-regeneration of ironic similes here plays second-fiddle to a main act that revolves largely around analogy.

Mere re-generation is especially useful in deciding how a bot should respond to the tweets addressed to it by other, human users on Twitter. *@TrumpScuttleBot* has three strategies for replying to humans: in the first, it matches words in the human's tweet to the entries in a dictionary of dream symbols, rewriting the text of the interpretation in Twitter terms so that "to dream about" becomes "to tweet about"; in the second, it identifies a property P in the human tweet for which it can find an ironic simile in its database; and in the third, it thanks the human for a comment that is one of *{topical, useful, fascinating, deep, insightful, intelligent}*, and attaches an aptly ironic simile to subvert this gratitude. Here is a real exchange between the bot and a human user:

From @Anonymized, replying to @TrumpScuttleBot
Ties? He couldn't be further away.

From @TrumpScuttleBot, replying to @Anonymized
Thanks for the useful feedback. It's as useful as a bag full of farts.

From @Anonymized, replying to @TrumpScuttleBot
Open it up and get a whiff … aaaaahhhhhhhh

From @TrumpScuttleBot, replying to @Anonymized
To tweet of being or moving up suggests you are emerging from a depressing or negative situation.

From @Anonymized, replying to @TrumpScuttleBot
Yes Dr. Wizard

From @TrumpScuttleBot, replying to @Anonymized
To mention a wizard in your tweet suggests you are trying to hone your skills and exercize your power

From @Anonymized, replying to @TrumpScuttleBot
OK Goofy, if it makes you feel better.

From @TrumpScuttleBot, replying to @Anonymized
You make Jar Jar Binks seem about as goofy as doomsday

From @Anonymized, replying to @TrumpScuttleBot
Good one, you got me.

From @TrumpScuttleBot, replying to @Anonymized
You make others seem about as good as a movie about a leprechaun in space

From @Anonymized, replying to @TrumpScuttleBot
At least Trump is president.

From @TrumpScuttleBot, replying to @Anonymized
Thanks for the useful feedback. It is as useful as a knock on the head.

From @Anonymized, replying to @TrumpScuttleBot
You keep responding, so who's head is being knocked?

From @TrumpScuttleBot, replying to @Anonymized
To mention a head in your tweet signifies wisdom, intellect, understanding and rationality.

A rapid-response bot like *@TrumpScuttleBot* always gets the last word, but the creative challenge is in sustaining an engaging, if combative, banter with sporadic flashes of wit. Notice how the bot enriches a blend from its database with an extra flourish of its own. If a tweet contains a property P associated with an entity, fictional or real, in its database of familiar faces, the bot mixes that character into the blend too, as when it compares a "goofy" user to *Jar Jar Binks*.

Comedy Gold

Our machine should not meddle with the distinctive mix of images in a prebaked blend, but it might rework its syntax. Consider our ironic bot, *@OldSkoolFunBot*, which aims to generate witty banter by repackaging the ironic blends in its database of *about* similes, as in the following tweets:

Question: Where are you most likely to find a wine stain? Well, in my world, how about on a white shirt?

If you're like me you'll absolutely despise cooking spaghetti in the washing machine – What's that all about?

Kids nowadays have their iTunes but in MY day we had to make do with being strapped to the rack

I gave the mother-in-law the south end of a north-bound spiny lobster for Valentine's day but the grouch said my gift wasn't GOOD-LOOKING enough

The bot derives its sense of the ridiculous from the ways in which images are juxtaposed in ironic *about* similes. These juxtapositions are transformed via mere re-generation into quips that are just as ridiculous, even if no longer similes. New comic forms for prebaked combinations can be added quickly to adapt the bot to a new social trend. Consider the world of microbrew gastropubs, craft beers, and the quirky names that draw trend-setters to them. English pub names are famous for their naming conventions, in pairings such as "The Duke and Pony" and "The White Hart." To generate novelty pub names, *@OldSkoolFunBot* ekes out a pair of juxtaposed images from an *about* simile, to invent new names such as "The Porkchop and Synagogue", "The Dog and Sweater" and "The Fart and Spacesuit." To invent eye-catching new brands of craft beer, it reframes similes such as "friendly as a rabid dog" and "firm as a wobbly jelly" as "Rabid Dog IPA" and "Wobbly Jelly ale," in the hope that the humour of the juxtapositions persists in the new syntax. Here are some sample tweets from the bot in this vein:

I'm off down to my local microbrew pub, The Drop And Bucket, for a pint of Mediaeval Ordeal ale.

Fancy going down to the new microbrewery, The Bug And Rug, for a pint of Hungry Snake weizenbeer?

I'm off down to my local microbrew pub, The Elephant And Tutu, for a pint of Triple Espresso lager.

These confections are more than random but less than fully appreciated by the bot itself. However, the random aspect does allow for unplanned resonances to emerge, as in:

Fancy going down to the new microbrewery, The Dog And Wheelbarrow, for a pint of Golden Retriever lager?

Fancy going down to the new microbrewery, The Fart And Car, for a pint of Ford Corsair ale?

Many other re-generation opportunities present themselves for rapid development in this way, such as in the naming of movie sequels by *@InterCableBot* or the naming of books by *@BotOnBotAction*. Re-generation bots like these aim to invent something new that has the magic of something old.

Stranger Then Friction

On encountering a robot with a humour setting, Cooper, an astronaut on a daring mission in the 2014 film *Interstellar*, says: "A massive, sarcastic robot. What a great idea." The robot in question, TARS, is blessed with a sense of humour that encompasses the sarcastic and the ironic, and proves to be a most excellent partner in the execution of the mission. KIPP and CASE, the film's other two robots, are said to be newer and faster than TARS, yet their senses of humour are less developed, presumably because TARS has more of what it takes to be witty: experience of people in the world.

Ironists use this experience to play one kind of friction against another: to exploit a friction between ideas, and the gap between expectation and reality, to lessen the tension between two people or between people and a machine. An ironic machine is attuned to disappointment yet knows how to repackage failure as amusement. Two approaches to this transformation have been presented here: a tightly controlled form of disanalogy that conflates an expectation and its failure in a single affect-shifting utterance, and a form of creative quotation that reuses attested examples of irony in new descriptive contexts that make them relevant again. Each approach can be modulated at the level of presentation to achieve effects that are more pointed or more subtle, but each works with different materials. In comedic terms, the first is a *straight* man that uses propositions that are not humorous in themselves to explain a failure of reasonable expectations; the second is a *funny* man whose material resonates with our own negative experiences of the world, but who encourages us to laugh at those experiences.

As with any successful comedy partnership, the perfect ironist is a marriage of both approaches, in which the first brings buttoned-down control and the second brings manic energy. This tight integration of structure and imagination has yet to be achieved in a single computational generator, even if each approach can be implemented side-by-side in a single system, such as a Twitterbot. Generative grammars (in a format named *Tracery*; see Compton *et al*., 2015) for the bots presented in this paper are available for download from *github.com/prosecconetwork*. This forced marriage of approaches will remain a chaste two-bed affair until we can better appreciate the magic of a screwball juxtaposition in computational terms. Until then, that "massive, sarcastic robot" will remain "a great idea," but a fictional one too.

References

Brants, T. & Franz, A. (2006). Web 1T 5-gram Version 1. *Linguistic Data Consortium*.

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

Compton, K., Kybartas, B. & Mateas, M. (2015). *Tracery: An Author-Focused Generative Text Tool*. In Proc. of International Conference on Interactive Digital Storytelling, Denmark:154-161.

Fauconnier, G. & Turner, M. (2002). *The Way We Think. Conceptual Blending and the Mind's Hidden Complexities.* Basic Books.

Fishelov, D. (1992). Poetic and Non-Poetic Simile: Structure, Semantics, Rhetoric. *Poetics Today*, 14(1), 1-23.

Ghosh, A., Li, G., Veale, T., Rosso, P., Shutova, E., Barnden, J., & Reyes, A. (2015). SemEval-2015 Task 11: Sentiment Analysis of Figurative Language in Twitter. *In the Proc. of SemEval-2015, the 9th International Workshop on Semantic Evaluations,* Denver, Colorado, June 4-5, 2015.

Ghosh, A. & Veale, T. (2016). Fracking Sarcasm with Neural Networks. In Proceedings of *WASSA'2016, 7th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, at NAACL-2016, San Diego, USA

Ghosh, A., & Veale, T. (2017). Magnets for Sarcasm: Making sarcasm detection timely, contextual and very personal. In *Proc. of EMNLP'2017, the Conf. on Empirical Methods in Natural Language Processing*. Copenhagen, Denmark, September 2017.

Giora, R., Givoni, S., & Fein, O. (2015). Defaultness reigns: The case of sarcasm. *Metaphor and Symbol, 30/4*, 290-313.

Grice, H. P. (1978). Further Notes on Logic and Conversation, In P. Cole (ed.), *Syntax and Semantics, Vol. 9: Pragmatics* (pp. 41-57). New York: Academic Press.

Hao, Y. & Veale, T. (2010). An Ironic Fist in a Velvet Glove: Creative Mis-Representation in the Construction of Ironic Similes. *Minds and Machines 20*(4), 635–650.

Kierkegaard, S. (1841). *The concept of irony with continuous reference to Socrates*. Harper & Row, 1966 reprint, p. 278.

Kreuz, R. J., & Glucksberg, S. (1989). How to be sarcastic: The echoic reminder theory of verbal irony. *Journal of Experimental Psychology: General*, *118*(4), 374-386.

Kumon-Nakamura, S., Glucksberg, S., & Brown, M. (1995). How about another piece of pie: the allusional pretense theory of discourse rrony. *J. of Experimental Psychology: General*, *124*, 3-21.

Reyes, A., Rosso, P. & Veale, T. (2013). A multidimensional approach for detecting irony in twitter. *Language Resources and Evaluation 47:239--268.*

Riloff, E., Qadir, A., Surve, P., De Silva, L., Gilbert, N. & Huang, R.. **(**2013). Sarcasm as contrast between a positive senti- ment and negative situation. In *EMNLP*, vol. 13:704–714

Sperber, D., & Wilson, D. (1981). Irony and the use-mention distinction. In P. Cole (ed.), *Radical Pragmatics* (pp. 295–318). New York, NY: Academic Press.

Taylor, A. (1954). Proverbial Comparisons and Similes from California. *Folklore Studies* 3. Berkeley: Uni. of California Press.

Valitutti, A. & Veale, T. (2015). Inducing an Ironic Effect in Automated Tweets. In Proceedings of the Sixth International Conference on Affective Computing and Intelligent Interaction (ACII2015), September 21-24, Xi'an, China

Veale, T. (2011a). We Can Re-Use It For You Wholesale Serendipity and Objets Trouvés in Linguistic Creativity. In Proc. of the 2nd International Conf. on Computational Creativity, Mexico.

Veale, T. (2011b). Creative Introspection and Knowledge Acquisition: Learning about the world thru introspective questions and exploratory metaphors. In Proc. of AAAI'2011, the 25th Conf. of the Assoc. for Advancement of AI, San Francisco, CA.

Veale, T. (2013). Humorous Similes. *HUMOR: The International Journal of Humor Research,* ***21****(1):3-22.*

Veale, T. (2015). The Humour of Exceptional Cases: Jokes as Compressed Thought Experiments. *In Brône, G., Feyaerts, K.. and Veale, T. (eds.), Cognitive Linguistics and Humor Research: Applications of Cognitive Linguistics*, Mouton de Gruyter, 69-90.

Yla R Tausczik & J W Pennebaker. (2010). The psychological meaning of words: Liwc and computerized text analysis methods. *Journal of language and social psychology*, 29(1):24–54.