Read Me Like A Book:

Lessons in Affective, Topical and Personalized Computational Creativity

Tony Veale

School of Computer Science, University College Dublin, Dublin, Ireland.

Abstract

Context is crucial to creativity, as shown by the significance we attach to the labels *P-* and *H-Creativity*. It is context that allows a system to truly assess novelty, or to ensure that its topical artifacts really are topical. An important but an often overlooked aspect of context is personality. A CC system that is designed to reflect a specific aspect of the creative temperament, whether it is humour, arrogance or whimsy, must stay true to this assumed personality in its actions. Likewise, a system that creates artifacts that are rooted in emotion must be sensitive to the personality or mood of its audience. But here we must tread carefully, as the assessment of personal qualities often implies judgement, and so few of us like to be judged, especially by our machines. To better understand the upsides and pitfalls of topicality- and personality-based CC systems, we unpack three of these systems here, and explore the lessons they offer.

The Wonder of You

Creativity can be an intensely personal affair. We put ourselves into what we create, relying on our experiences and values to build artifacts we hope others will value too. In doing so, we reveal our personalities. When we create for others and assimilate the values of an audience, creativity becomes personal *and* personalized. While it is a tenet of *computational creativity* (CC) that an agent need not be a person to be creative, a creative system may nonetheless need a personality, or an appreciation of the personalities of others, before it can create like a human (Colton *et al*. 2008). '*Software is not human,*' to quote the CC refrain, yet CC systems must appreciate what matters to a human. In any case, we can only know a CC system's personality, and it can only know ours, by what we do or say, making personal/personalized CC a special case of contextual CC. For CC systems that create in language, context is itself a linguistic artifact, as rooted e.g., in our social media timelines. Here we describe how best to use linguistic context to deliver various forms of topical and personalized CC.

Specifically, we will explore the role of linguistic context in the operation of three *Twitter*-based CC systems, ranging from one that uses context to ensure topicality to ones that view context as the imprint of a user personality. For personalization, the Twitter footprint of a target user – whether their official bio or their recent tweets – offers a textual context in which to situate the generation process. For topical creativity, the aggregated timelines of an array of online news sources, from a Twitter-addicted president to the breaking headlines of mainstream media, provide a dynamic context for machine creativity. We explore three modes of CC via these systems: a marriage of linguistic and artistic creativity that maps the digital personality of a user, as reflected in what they tweet, into metaphors that are both textual and visual; a topical creator that generates metaphors for news stories rather than news readers; and a book recommender system that leavens its user-tailored suggestions with humour, and which invents its own book ideas to supplement the titles in its well-stocked database.

The principle that unites all three systems is the role of information compression in CC. One space of information may be compressed or decompressed to yield others, and produce insightful generalizations or vivid elaborations in the process. Thus, compression is required to map a news story to a linguistic metaphor, as the metaphor need only capture the gist of the story. In fact, such information loss is desirable when it leads to generalization and ambiguity, as metaphors should be objects of profound wonderment. When moving from online user personalities to metaphors we require the opposite, *decompression*, to inflate a low-dimensional space of personality types into an elaborate, high-dimensioanl space of possible character metaphors. Current sentiment analysis techniques can place a user in a space of a dozen or so psychological dimensions, while metaphors will occupy a space that – even after a process of dimension reduction – has hundreds of dimensions. In fact, even the extraction of psychological dimensions is a compression process, since the textual timelines that feed into sentiment analysis are converted into high-dimension distributed spaces built with word co-occurrence statistics.

In the next section we focus on personality-driven CC with a system that maps recent user moods into metaphors and pictures. Our approach is data- and knowledge-based, marrying textual data from a user profile with a symbolic model of the cultural allusions that underpin a machine's metaphors. Following that, we give statistical form to the notion that metaphors reside in a space of possibilities, so as to re-imagine metaphor creation as a mapping from one space, a topic model of the news, into another, a space of metaphors that shares exactly the same dimensions. These are whimsical systems that make sport of news and mood, so we present one more system, a CC book recommender that uses simple information-retrieval techniques to guide its suggestions, but which also uses machine creativity in some unsubtle ways. This specific system was built for a recent science communications event, and user feedback offers us some lessons on the willingness of humans to be judged by machines. Although whimsy can diminish the severity of a perceived criticism, humour must be wielded with care by our autonomous CC systems, especially if it is unbidden, or used for the furtherance of serious goals.

Metaphor Mirror On The Wall

Consider the problem of generating apt metaphors for the news. As a story breaks and headlines stream on Twitter, we want our metaphor machine to pair an original and insightful metaphor to each headline. So a headline about extreme weather might be paired with a metaphor about nature's destructive might, or a political scandal might be paired to a crime metaphor. As metaphor theorists often speak of multiple spaces – e.g. Koestler (1964), Lakoff & Johnson (1980) and Fauconnier & Turner (2002) all see different viewpoints as different spaces – it is tempting to model each space in a metaphor with its own vector space model (VSM), by equating vector spaces with conceptual spaces. Yet this analogy is misleading, as different VSMs – constructed from different text corpora – must have different dimensions (even if they share the same *number* of dimensions) and we cannot directly perform geometric comparisons between the vectors of two different VSMs. Since the principal reason for building a VSM is the ease with which semantic tests can be replaced with geometric ones, we should build a single vector space that imposes the same dimensions on each conceptual metaphor space. It is useful then to view news headlines and metaphors as comprising two overlapping subspaces of the same VSM.

For a news subspace we collect a large corpus of news content from the Twitter feeds of CNN, Fox News, AP, Reuters, BBC and New York Times, and use a standard compression technique – such as LDA (*Latent Dirchlet Allocation*; Blei *et al*., 2003), LSA (*Latent Semantic Analysis*; Landauer & Dumais, 1997) or *Word2Vec* (Mikolov, 2013) – to generate a vector for each headline. We additionally build a large metaphor corpus by running the *Metaphor Magnet* system of Veale (2015) on the Google n-grams (Brants & Franz, 2006), to give millions of metaphors that stretch across diverse topics. Rather than build separate vector spaces for the news and metaphor corpora, we build a *single* vector space for both by appending one corpus onto the other before applying dimension compression. Within this joint VSM, every past metaphor and future headline is assigned a vector of precisely the same dimensionality. It is now a simple matter to measure the angle between the vector for an incoming headline and those of previously encoded *MetaphorMagnet* metaphors.

The metaphor whose vector presents the smallest angle (the largest cosine) to an incoming news vector is chosen as the one with the most relevance to that news item. We built our joint space by compressing 380,000 news items, 210,000 tweets (from sources including @*nytimesworld*, @*CNNbrk* and @*FOXnews*) and 22,846,672 metaphors from *MetaphorMagnet* (which were made available to us on request) into the same LDA space of 100 dimensions. We used the *gensim* package of Řehůřek & Sojka (2010) to build the space, and concatenated word lemmas to their POS tags to provide a richer feature set to the model.

The best pairings produced by this conflation of spaces are tweeted hourly by our bot, called @*MetaphorMirror*. The thematic basis of the compression means that some pairings show more literal similarity that others, as in:

> **From @WSJ**: *Sultan Abdullah of Pahang has been chosen as Malaysia's new king.*
>
> ↑↓
>
> **From @MetaphorMirror**: *What is a sultan but a ruling crony? What is a crony but a subservient sultan? What drives ruling sultans to be toppled from thrones, appointed by bosses and to become subvervient cronies?*

As is evident here, *MetaphorMagnet*'s hardboiled worldview shines through in these pairings, offering meanings and perspectives that, while not actually present in a headline, can be *read into* the headline if one is so inclined. So some pairings show that the VSM has learnt the lessons of history by reading the news, and this shines through too:

> **From @AP**: *Congo's new President Felix Tshisekedi sworn into office; country's first peaceful transfer of power since independence.*
>
> ↑↓
>
> **From @MetaphorMirror**: *How might an elected incumbent become an unelected warlord? What if elected incumbents were to complete tenures, grab power and become unelected warlords.*

Any stereotyping in this response is a product of the VSM and its large corpus of past news, rather than any bias in *MetaphorMagnet*. Though the latter has a symbolic model of warlords and democrats, it is the news space that unites this generic model with the specific history of the Congo. So even as a system strives for topicality, it must have one foot planted in the past if its outputs are to seem informed.

Fifty Shades of Dorian Gray

Much research has been conducted on the analysis of human personality as reflected in our lexical choices. Chung & Pennebaker (2008) describe a tool and a resource, the LIWC (or *Linguistic Inquiry & Word Count*), for estimating personality traits such as *anger, affability, positivity, topicality, excitability, arrogance, analyticity, awareness*, *worry, anxiety* and *social engagement* from a writer's text outputs. The web version of the tool, *AnalyzeWords.com*, which calculates values for these 11 dimensions by analyzing one's recent tweets, tells us that *@Oprah* is upbeat and affable as a tweet writer, while *@realDonaldTrump* is upbeat but angry. To create metaphors for a specific person, such as Donald Trump or Oprah, a machine can treat a recent *AnalyzeWords* profile as an 11-dimension vector in personality space, and seek to map this coarse vector to a higher-dimensional space of metaphorical possibilities.

Given the disparity in dimensions between these spaces (11 versus 100) and their different means of construction, we cannot build a joint vector space by just concatenating data. Lacking a dataset to train a neural network to do this mapping across the spaces, we use a symbolic approach to inflate the *AnalyzeWords* space into 100s of dimensions that capture the qualities highlighted in our metaphor set. So we inflate the smaller space by hand-crafting logical formulas – or *transformulas* – to estimate approximately 300 qualities as functions of the eleven core dimensions. Transformulas can conjoin, disjoin and negate these core dimensions. All core dimensions are mapped to the scale 0 to 1.0 (from 0 to 100), and so all transformulas calculate values in the range 0 to 1 also. The negation of a quality simply inverts this scale, so *not angry* can be calculated to be (1 - *angry*). Consider the transformula for *neurotic*:

$$\textit{neurotic}(\mathrm{u}) = \textit{worried}(\mathrm{u}) \times \textit{analytic}(\mathrm{u})$$

That is, since neurotics tend to overthink their worries, we estimate the *neurotic*ism of user *u* to be the product of the core dimensions *worried* and *analytic*. Likewise, we can say that someone is *narcissistic* to the extent that they are *arrogant* and *self-aware* (given to talking about their own feelings), or *creative* to the extent they are *analytical* and *upbeat*. While transformulas do not reflect an empirical truth about a person, they codify a kind of 'folk' symbolic reasoning that lends itself to explicit verbal explanation. Importantly, they allow any Twitter user *u* to be described in terms of the vivid qualities that are used in the NOC list (Veale, 2016) to characterize its gallery of famous people. So, once our transformulas have mapped *AnalyzeWords*'s 11 dimensions into the rich voculabulary of the NOC list, a Twitter user can be compared and matched to its iconic membership. In this way, *@ElonMusk* may show a strong similarity to *Walter White* of *Breaking Bad*, while *@realDonaldTrump* might produce a match to *Lex Luthor*. Such metaphors are a reach – all good metaphors are – but each can be explained in symbolic terms using the logic of the transformulas that link them to their most recent tweets. As such, transformulas turn text-analytic calculations into talking points that a creative linguistic system can exploit.

Consider again the example of *@ElonMusk*, engineer and entrepreneur. From an *AnalyzeWords.com* profile that places his tweets high on the core dimensions *upbeat* and *analytic* and low on the dimensions *angry* and *self-aware*, the transformula qualities *optimistic* (*upbeat* × *analytic*), *dispassionate* (*analytic* × *not angry*), *unfeeling* (*analytic* × *not sensory*) and *determined* (*upbeat* and *not angry*) can be inferred. Since three of these transformula qualities – *unfeeling*, *determined* and *dispassionate* – are typical of machines, and the fourth, *optimistic*, is not, our metaphor generator might describe Musk (in light of his most recent tweets) as an "optimistic machine." As his *AnalyzeWords* profile also suggests the qualities *laid-back, educated* and *scientific*, the latter two of which are typical of researchers, it can also describe Musk as a "laid-back researcher."

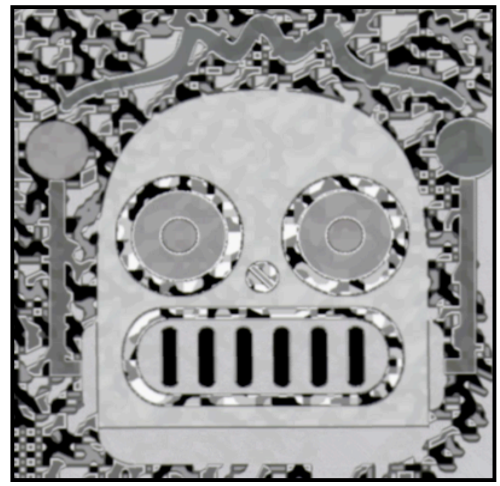


Fig. 1. A personalized metaphor for @ElonMusk.

These metaphors, as tweeted by the metaphor-generating bot *@BotOnBotAction*, are shown in Figures 1 and 2.

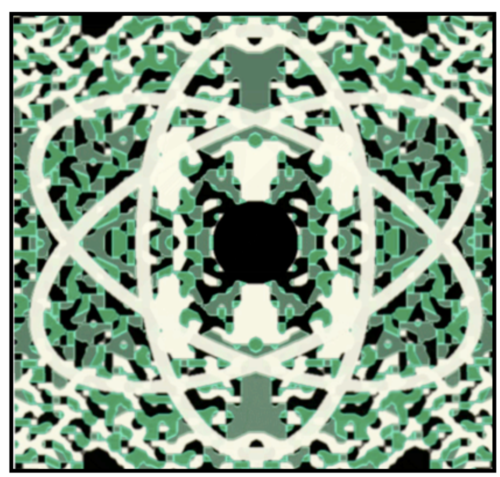


Fig. 2. Alternate personalized metaphor for @ElonMusk.

The bot creates a new piece of visual art to complement its metaphors, by creating a 1-dimensional 4-state cellular automaton that unfurls over many rows/generations – and rendering its four states with colours chosen to match the highlighted qualities of the metaphor (see Veale & Cook, 2018). As a flourish, the bitmap of an Emoji annotated with one of the words in the metaphor – so an atom for "scientific" in Fig. 1, and a robot for "robot" in Fig. 2 – is integrated into the image and coloured to suit its new context. Using a colour lexicon in which 600 of the metaphor machine's stereotypes are mapped to apt RGB codes (e.g., silver-grey for robots, black for priests), it is possible to assign a specific hue to even the non-visual qualities (see Veale & Alnajjar, 2016). Each metaphor is then framed so as to cement this link, so that the black of Fig. 1 is named "educated priest black" while that of Fig. 2 is "unfeeling Sith black." The tweeted metaphor, comprising four intertwined sub-metaphors, tells us what each of these colours stands for, and suggests how we should feel about them.

It is important to note that *@BotOnBotAction* operates on an *opt-in* basis for most users. The bot will not target them, or metaphorize them, unless explicitly asked to do so with the hashtag *#PaintMySoul*. This kind of personalized computational creativity is not always flattering or welcome, and may even – in some cases – be considered abusive. The exception to this opt-in rule concerns high-profile celebrity users of Twitter, who use the platform to promote themselves to millions of followers. The bot uses the website *TwitterCounter.com* (now defunct) to obtain a list of the most well-known personalities on Twitter, such as Mr. Musk, and freely generates metaphors and images for these luminaries in the downtime between its explicit user commissions. Artists and satirists have always made targets of the powerful and famous, who hardly notice the impudence of a single provocateur; our bot is no different.

Making a Hash of Computational Creativity

Personalization and topicalization offer orthogonal means of grounding the products of CC in the here and now of a user's reality. Personalization shows that a creative system understands its users, whilst topicalization shows that it appreciates the current and historical context that connects them both. These alternate means of grounding intersect in the task of *recommendation*, for a good recommender engine must understand both the personal dimensions of its users and the topics that matter most to them.

In this section we describe the rationale and the mechanisms of a recommender system for books as embodied in a Twitterbot named *@ReadMeLikeABot*. As with *@BotOnBotAction*, the bot obeys a mostly opt-in policy for its interactions with users, who request ideas for new books to read by using the hashtag *#ReadMeLikeABook*. When the bot is invoked in this way, it uses the text of the invoking tweet as the basis for its recommendation. If this does not offer a foothold to the recommender engine, the bot looks instead at the short Twitter biography that each user defines for their account. If this is empty or unrevealing, the bot finally considers the most recent tweets of the user as a source of topical material for its book suggestions. As in *@BotOnBotAction*, those recent tweets also offer a basis for inferring something of the personality of a user, which may additionally colour the bot's book recommendations.

The bot also has two activation modes that do not obey a strict opt-in policy. The first is perhaps partially opt-in, insofar as one can request a recommendation for another. In this mode, a user tweets *#ReadHimLikeABook* followed by the Twitter handle of a friend; the bot also accepts the tags *#ReadHerLikeABook* and *#ReadThemLikeABook*. As with *@BotOnBotAction*, the second mode is a filler mode for when the system finds itself between explicit requests. In this case, it exploits the fact that many of the authors in its books database are themselves on Twitter, and so aims to start a conversation about modern literature that draws contemporary writers into an online discussion of books. Authors opt-out of this mode by simply blocking the bot.

Recommender systems are typically either user-based or content-based. In the former, a perceived similarity between users permits a system to recommend items favored by one to the other. In the latter, the similarity function is defined over the items themselves, so a user that favors a given item is likely to favor a similar item too. These two modes are far from orthogonal, as a similarity function for users can be defined over the set of items they both favor, whilst a similarity function for items can be defined over the set of users that favor them both. In short, as a system learns more about its users, it learns more about the items it has in its database of possible recommendations. Importantly, *@ReadMeLikeABot* is not designed to track users, or to learn very much about them, other than that which is public in their Twitter accounts. The bot remembers what it recommends simply to ensure that it does not make the same suggestion again in too short a timeframe. The bot's user-based recommendations are *personality*-based, while its content-based recommendations are *topic*-based, where each is inferred on the basis of Twitter usage alone.

Recommendation systems are a practical application of AI, yet the task of suggesting existing items permits very little in the way of novelty, no matter how insightful a recommendation may be. Where then lies the computational creativity of a system like *@ReadMeLikeABot*? We view book recommendation not as a creative task in itself, but as an occasion for creativity that allows an expressive CC system to demonstrate a witty and whimsical personality. Consider aspects of linguistic creativity such as metaphor and irony. While a bot like *@MetaphorMagnet* can generate meaningful metaphors with a characteristic voice of its own, its outputs are mostly apropos of nothing, for the bot must rely on its readers to see a serendipitous relevance in its outputs, in whatever context they consume them. Our *@MetaphorMirror* bot finds this relevance for itself in the topicality of the news, yet the bot remains a showcase for metaphorical capability rather than a practical application in its own right. Linguistic creativity is a welcome seasoning for language, rather than the meal itself; it works best when it augments rather than supplants our practical aims. When viewed as a recommender of books, *@ReadMeLikeABot* is not a CC system. Yet the act of suggesting content to a user on the basis of its insights into the user's personality allows a system to be creative in the expression of its insights, and to find a genuine use for irony and metaphor.

Unauthorized Autobiographies

@ReadMeLikeABot maintains a tiered database of content to recommend. Its first tier contains 500 or so books that are well known, highly regarded, and by authors of some renown. Whenever a book from this tier is recommended, the system can be confident that the user has most likely heard of it, and will likely see its relevance. Each book in this tier is also associated with a set of qualities that describe not just the book itself but the traits of the readers that are most likely to read it. We might assume, then, that philosophical readers enjoy philosophical books. In this way, qualities such as *smart*, *philosophical*, *warm*, *hostile* and *upbeat* can be linked, via appropriate transformulas, to Twitter users who exhibit the same personality traits.

The bot's second tier is much larger, but also much less authoritative. Its 15,000 or so books have been extracted from *DBpedia*.org using the website's SPARQL endpoint. We exploit the linguistic regularity of *DBpedia*'s category terms to also extract a set of themes for every book. When a book is listed under a semantic category with the label *Xs_about_Y* or *Y_in_fiction*, we extract Y as an apt theme. We also mine the hierarchical relations between *DBpedia* categories to build a semantic network that relates these book themes to each other, such as *Artificial Intelligence* to *Neural Networks*. This genre and theme network is then the basis of the bot's content-based recommendations.

The last tier, and certainly the most unusual, comprises 6000 or so humorous fabrications, wholly invented book ideas that wear their artifice on their sleeves. These titles show the usefulness of CC to a recommender system, for when the system has no new content to recommend to a user, it can always fall back on its own in-jokes to fill the gap and keep the user engaged. These inventions must be seen as the literary jokes they are if the bot's credibility as a recommender is not to be diminished in the process. To generate these witty fabrications, we use the *NOC list* of Veale (2016), a large multifaceted database of pop-culture icons that provides vivid descriptions for over 1000 famous people, both real and fictional, modern and historical. For each person, the NOC provides a set of categories – e.g., *billionaire* for Donald Trump, or *politican* for Hillary Clinton – and a set of typical activities, such as *building giant walls* for Trump and *tolerating adultery* for Clinton.

The NOC is a generic, application-neutral resource, but as these examples show, no little humour is baked into the database from the get-go. The NOC list was first built for the WHIM project (the *What-If Machine*), and the task of generating whimsical book ideas can be seen as a what-if scenario: what if Genghis Khan, or Bono, or Tony Stark, wrote a book and told us what they were really thinking? What-if book generation is a simple task using the NOC: a system combines a famous person with an apt category and an associated activity, as in the following examples:

The Comedienne's Guide to Ranting About Liberals
The Rockstar's Guide To Avoiding Taxes
The Son's Guide to Disappointing the Family

These faux books are credited to, respectively, Roseanne Barr, Bono, and Fredo Corleone. When the NOC entity is fictional and has a known creator, this information is also used in the generation of literary what-ifs. Consider these:

Captain Ahab's Guide To Chasing a Great White Whale
Dr. Stephen Strange's Guide to Performing Magic Tricks
Yoda's Guide to Promoting Mysticism

These books are credited to Herman Melville, Stan Lee and George Lucas, respectively. What-ifs also give us the opportunity to imagine incongruous pairings of authors:

The Geek's Guide to Studying Science
The Psychiatrist's Guide to Probing the Mind
The CEO's guide to Pioneering New Technologies

The first is credited to Peter Parker and Wesley Crusher; the second to Drs. Sigmund Freud and Frasier Crane; the third to Tony Stark and Steve Jobs. In general, any linguistic framing of pop-culture factoids that pokes fun at the book industry will suffice here. Publishers themselves see the value of parodic cash-grabs, and shelves already groan under fictive offerings like the following, by Pablo Escobar, Tyrion Lannister and Wile E. Coyote, respectively:

Lifting The Lid on The Medellín Cartel
Exposed: The Secrets of The House Lannister
An Insider's Guide to A.C.M.E.

Recall that such non-books are only ever recommended to the user when better matches from a higher tier cannot be found, or when all have been offered to that user already. Their value is largely found in repetition, then: the more a user interacts with the bot, the further down its tiers the bot must descend, and more the bot will reveal its sense of humour, about books and about the book industry itself.

They Shall Not Grow Bold

Much research has focused on the recognition of sarcasm and irony in text, especially as it is used in social media. This emphasis on detection is not surprising, given that so much of the language that matters is created by humans. In contrast, very little research has addressed the creative task of *generating* irony and sarcasm, no doubt because we already find our machines to be inscrutable enough in their dealings with humans. But more than that, sarcasm and irony cannot exist outside of a specific communicative goal: we can generate metaphors in a null context and leave it to the reader to unearth their implied meaning, but irony and sarcasm require a firm context to push against. In short, they need realistic expectations to bring to bear, and a context that undermines them in ways for all to see. For a machine to generate irony and sarcasm well, it must be given enough of these expectations to be versatile, and an ability to identify those contexts that clearly fall short.

Personality-driven recommendation supplies these expectations in convenient qualitative and quantitative forms. When the bot has a topic-based reason to recommend e.g., an intellectual book to a reader who scores low on the analytic dimension, or is poorly scored by the transformula for *intellectual,* this mismatch between topic and personality is just the failure of expectation that irony demands. In this case, topic-based recommendation creates the expectation, and personality analysis defies it, giving the bot a logical reason to snarkily poke fun at the disparity. Suppose the bot does suggest an intellectual book, on cosmology, say, to a user with an avowed interest in cosmology that appears to fall well short of the intellectual bar; how should it wittily allude to this failure of expectations? The bot can learn from how humans deal with disappointment by looking to how we express our dissatisfaction through irony. So a web search for *intellectual* finds the following ironic similes: *about as intellectual as a Cheez Doodle, as a cucumber, as a brush, a hole in the ground, a wart hog, a potted plant, a bulldog*, *an emu*, and others too rude to repeat here. What links each of these mental images is not a shared feature but a common framing; in each case, the author prefaces a simile with "about" to signify the semantic imprecision of a creative liberty. We can exploit this framing device to seek out many other ironic similes on the Web for any quality one cares to undermine, to give a bot a rich palette of ironic options to use on its own users.

When a user's Twitter profile scores low for a quality that is estimated either directly (using *AnalyzeWords*) or indirectly (via a transformula), the bot will dip into its bag of ironic similes for that quality. Choosing at random, it can frame what it retrieves in a variety of colourful ways. Suppose the quality is *philosophical*, and the bot retrieves the simile *about as philosophical as a bowel movement*. Perhaps the bot also intends to recommend the philosophical novel *Steppenwolfe* by Hermann Hesse, as the user's recent tweets mention *loneliness* and *alone*. It can frame this pairing of a book to a simile in the following ways:

> Hey @bookreader, if you're as philosophical as a bowel movement then maybe you should read 'Steppenwolfe' by Hermann Hesse on the theme of solitude.

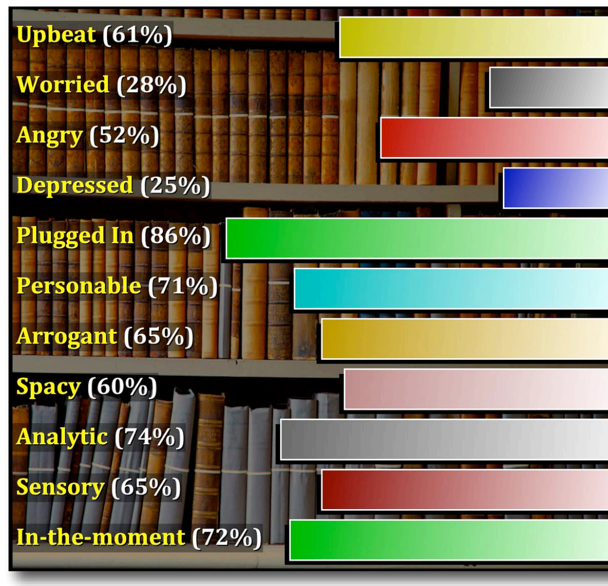
> Hey @bookreader, I used to be as philosophical as a bowel movement until I read 'Steppenwolfe' by Hermann Hesse on the solitude theme.

> Hey @bookreader, given your personality profile I don't know which philosophical book is more you: 'Steppenwolfe' by Hermann Hesse on the solitude theme, or 'The Bowel Movement' by Stephen Tolkein.

The first framing was used in early field tests of the bot, in its prelaunch in the weeks before the 2018 *Science and Communication* conference (for which the bot was commissioned). As might be expected, its in-your-face humour was not popular with everyone, and was a cause for some dismay to the event's organizers. The "If you're X" construction did little to salve the pain of a sudden insult from an abusive bot, even if the user had invoked it explicitly. The second framing proved to be more successful, since it now turned the bot's humour inward, on itself, rather than outward on users who might see themselves as its victims. The third framing turns it outward again, on the user, but in a more subtle guise that presents it not as a direct insult but as a playful joke at the expense of the book industry. Note how the bot is forced to invent an author for its literary in-joke, which it does by cutting up the author names from its first tier of books. The third framing is especially apt when the bot's tweet is accompanied by a graph of the user's 11-dimension personality profile (see below), since it allows one to appreciate the basis for the bot's response.

But the second framing has another advantage, in that it allows the bot to speak directly to the topic of the recommended book. Consider this particular response to a user:

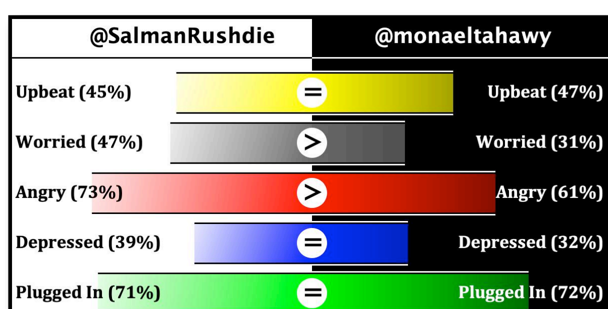
> On the prettiness theme, @anonymized, I used to be as attractive as a brown cardigan until I read "The Picture of Dorian Gray" by Oscar Wilde. How about you? I crunched your recent tweets:



When the bot is between user requests, it attempts to start a conversation about books and their ideas. It does so by posing literary questions to its readers, as in this tweet:

> On the religion theme, which of these books is more provocative than the other? "The Satanic Verses" by @SalmanRushdie, or "Headscarves and Hymens: Why the Middle East Needs a Sexual Revolution" by @MonaEltahawy? I compared their recent tweets.

The question itself typically provokes much less conversation on Twitter than the side-by-side personality analyses that the bot provides of the authors' most recent tweets.



As You Like It: The Question of Evaluation

A Twitterbot that publicizes a user's psychological profile is rather like a public *speak-your-weight* machine. No one likes to be judged, least of all by a computer, and we can expect a wide diversity of views on the use of tools like *AnalyzeWords* to condition a bot's creative outputs. These range from "Awesome!" to "very creepy," with the rump of users taking their analyses as a starting point for further wit of their own. One user replied to the book bot's ironic confession 'I used to be as poetic as a donkey passing wind until I read "Romeo and Juliet" by William Shakespeare' with the wry remark "I'll have you know that *I'm* still as poetic as a donkey." Another user, for whom the bot recommended Alex Comfort's "The Joy of Sex" (on the *love* theme) replied "Wow wow, easy there bot ... buy me dinner first!" The author *@MaggieEllen* replied to the bot's comparative analysis of herself to *@AmyTan* with a trope from *The Simpsons,* "you my overlord now?" before addressing the specifics of the analysis with the remark "Just real glad to know @AmyTan and I are both on 300 mg extended release Welbutrin and equally depressed." The value of a creative agent lies as much in the creativity it fosters in others as in the creativity of its own outputs.

That said, users are more open to personalized outputs when they flatter their targets, and often chafe at the negative aspects of analyses that are otherwise quite positive. One famous comedian with a science background, whose Twitter feed reflects his TV presence – half comedy, half science – was described by *@botOnBotAction* as "the best of Peter Parker and the worst of Jim Carrey: scientific and intelligent yet cloying and insecure." The user's response was unforgiving: "Not so insecure that I post anonymously though." When a science book by the same user, a popular author, was promoted by the book bot via its *Analyze Words* comparison to a similar author, the mention earned it a comparable rebuke "Not sure this analysis has a single thing to do with the books; but you enjoy yourself." When creative systems get personal, so too will their audiences, making it difficult to objectively evaluate their outputs.

This makes an extrinsic evaluation of such systems preferable. Ghosh & Veale (2017) explore the contribution of a user's Twitter profile – specifically, their *AnalyzeWords* profile – to the assessment of whether their most recent tweet is sarcastic or not. We expect mood and personality to be factors in the determination of a sarcastic mindset, as recent emotions – from anger to arrogance – will shape the perception of a user's intent in a given tweet. In that case, we expect a neural model of sarcasm detection to be improved by the addition of accurate personality features that are active in the relevant time frame. Ghosh & Veale report a statistically significant gain in detection accuracy, of approximately 6% to 7%, when *AnalyzeWords* features are incorporated into their neural architecture. If personality features can improve the appreciation of creative texts, they can certainly play a key role in their generation too.

Topicality-driven bots like *@MetaphorMirror* afford a more direct evaluation, since it is the news context, and not a specific user, that is directly addressed in the output.

Vector Space	*Low*	*Avg.*	*Good*	*V.Good*
LDA *stories+tweets*	1.1%	47.8%	**41.1%**	**10%**
LDA *stories only*	3.3%	65.6%	30%	1.1%
LSA *stories+tweets*	10%	60%	30%	0%
LSA *stories only*	17.8%	64.4%	16.7%	1.1%
Word2Vec	10%	57.8%	32.2%	0%
Random baseline	45.5%	46.7%	6.7%	1.1%

Table 1. *Distribution of Aptness by choice of model.*

We used *CrowdFlower (*now *Figure-Eight)* to elicit human ratings for 90 metaphor / headline pairs from different models. A scale of 1 ... 5 was used for ratings on three dimensions: *aptness*, *comprehensibility*, and *influence*, the last of which marks the extent to which a metaphor shapes a rater's response to a headline. Six different models were used to select the 'best' metaphor for each of the 90 headlines: an LDA topic model built with a corpus of 380k news stories and 210k news tweets; an LDA model built with the 380k news stories but no tweets; an LSA model built with 380k stories and 210k tweets; an LSA model built with 380k stories but no tweets; a Word2Vec space of pretrained vectors, so no news stories or tweets were used; and a baseline model that pairs a random metaphor to each headline. For each variant of the LDA and LSA models, 22.84M *MetaphorMagnet* metaphors were concatenated to the news content (stories with/without tweets), so these models produced joint *news + metaphor* spaces.

Mean ratings for each dimension in different models were not very discriminating. In each case, LDA (stories + tweets) pipped all others to the top spot. For *Aptness* – how apt is a metaphor for a headline? – the means ranged from 2.95 (± standard dev. 1.27) for LDA (stories+tweets) down to 2.20 ± 1.2 (random baseline). *Comprehensibility* – the understandability of each pairing – ranged from 3.59 ± 1.05 (for LDA, stories+tweets) down to 2.54 ± 1.12 (random baseline), and *Influence* ranged from 3.01 ± 1.24 (LDA stories+tweets) to 2.09 ± 1.24 (random baseline). The differences across models were not statistically significant, except in comparison to the baseline. Yet mean performance disguises deeper differences. If we quantize the human ratings of aptness into four equal-sized buckets (*Low*, *Average, Good* and *Very Good*) so as to identify the model that places the most metaphor:headline pairs into the *Good* or *Very Good* buckets, we obtain the findings of Table 1. More than half of pairings suggested by the LDA (stories+tweets) model end up in one of these top buckets, suggesting that this model produces the most apt results.

Conclusions: Don't Give Up The Day Job

Oscar Wilde once wrote that "art has as many meanings as man has moods." The point of affective computational creativity is not just to enlarge the space of artifacts that is explored by a CC system, or to make those artifacts more revealing about the processes that generated them; it is to make those artifacts more revealing about their *audiences*.

This potential for personalization and topicalization has not gone unnoticed in other CC work. With regard to *The Painting Fool*, a versatile generator of portraits (and other painterly forms), Colton *et al*. (2007) built on the work of Pantic & Rothkrantz (2003) to give the Fool a sense of the mood of the subject it is painting, so that it might capture this understanding in its outputs. A linguistic tool such as *AnalyzeWords* is of little use when dealing with a video or a camera still, but Colton *et al*. note the value of FACS, the *Facial Action Coding System* of Ekman (2002). Some users of CC systems wear their emotions on their faces; others reflect them in their social-media communications. Depending on the modality of the interaction – and personality certainly turns CC into an interactive process, even if users are scarcely aware of their own contributions – a system must exploit whatever affective cues it can find. Topicalization has also revealed a strong potential for CC exploitation. Like personalization, it makes the outputs of a generative system more relevant to the users for whom they are created. For example, the *PoeTryMe* poetry generator of Gonçalo Oliveira (2017) augments its core knowledge stores (such as semantic and conceptual networks) in a number of ways, including the use of live Twitter feeds to ground its outputs in the here-and-now of social media. By showing an awareness of users and their world, these systems present themselves as more self-aware too. They present themselves not as closed generative bubbles, like the imprisoned wretch of Searle's *Chinese Room* thought experiment (1980), but as agents of a wider world that can predict how their creative outputs will impinge on others.

If viewed as 'human' creators, CC systems such as *The Painting Fool*, *PoeTryMe* and *MetaphorMagnet* would all be seen as full-time creators whose work is their calling. Most CC systems conform to this all-or-nothing pattern; their creative work is everything, and the systems have no 'lives,' whatever this might mean, beyond their generative responsibilities. *@ReadMeLikeABot* is a useful exception to this generalization. To this CC system, as it is to most humans, creativity is merely a sideline to a 'day job' that is not in itself a creative exercise. Book recommendation is a task that requires AI but has little obvious use for CC, yet this bot shows that a system that benefits from an appreciation of a topical context, or a user's personality, can also reap benefits from the creative framing of its outputs. Conversely, the CC component of these systems may also benefit from exposure to the stuff of its mundane day job, by giving it a contingent knowledge of possibilities that it might never recognize in a purely creative mode. We can go further, and argue that *all* CC systems can benefit from a day job that exposes them to the mundane concerns of the people they must serve, so as to later transmute those concerns into something both familiar and non-obvious.

References

Blei, D. M., Ng, A. Y. & Jordan, M. I. (2003). Latent Dirichlet Allocation. *J. of Machine Learning Research*, 3:993–1022.

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

Chung, C.K. & Pennebaker, J.W. (2008). Revealing dimensions of thinking in open-ended self-descriptions: An automated meaning extraction method for natural language. *Journal of Research in Personality*, 46, 96-132.

Colton, S., Valstar, M.F., & Pantic. M. (2008). Emotionally Aware Automated Portrait Painting. In Proc. of DIMEA'08, the 3rd Int. Conf. on Digital Interactive Media in Entertainment and Arts Athens, Greece. September 10 - 12.

Ekman, P. (2002). *Facial Action Coding System. A Human Face*. Utah, Salt Lake City:W. V. Friesen, and J. C. Hager.

Fauconnier, G. & Turner, T. (2002). *The Way We Think: Conceptual Blending and the Mind's Hidden Complexities*. Basic Books, New York.

Ghosh, A. & Veale, T. (2017). Magnets for Sarcasm: Making Sarcasm Detection Timely, Contextual and Very Personal. *Proc. of EMNLP 2017, the Conf. on Empirical Methods in Natural Language Processing*, 493–502, Copenhagen, September 7–11.

Gonçalo Oliveira, H. (2017). O Poeta Artificial 2.0: Increasing meaningfulness in a poetry generation Twitter bot. In *Proc. of CC-NLG, the Workshop on Computational Creativity in Natural Language Generation*. Santiago de Compostela, Spain pp 11-20.

Koestler, A. (1964). *The Act of Creation.* Penguin Books.

Lakoff, G. &. Johnson, M. (1980). *We Live By*. University of Chicago Press.

Landauer, T.K. & Dumais, S. (1997). A solution to Plato's problem: the latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psych.Review* 104(2):211-240.

Mikolov, T., Chen, K., Corrado, G. & Dean, J. (2013). Efficient Estimation of Word Representations in Vector Space. *arXiv*:1301.3781 [CS], January.

Pantic, M. & Rothkrantz, L.J.M. (2003). Toward an affect-sensitive multimodal human-computer interaction. In Proc. of the IEEE, 91(9):1370–1390.

Řehůřek, R. & Sojka, P. (2010). Software Framework for Topic Modeling with Large Corpora. In Proc. of the LREC 2010 Workshop on New Challenges for NLP Frameworks, pp 45–50.

Searle, J. (1980). Minds, Brains and Programs. *Behavioral and Brain Sciences* 3(3): 417–457.

Veale, T. (2016). Round Up The Usual Suspects: Knowledge-Based Metaphor Generation. Proc. of *the Meta4NLP Workshop on Metaphor at NAACL-2016, the annual meeting of the North American Assoc. for Computational Linguistics CA: San Diego.*

Veale, T. (2015). Unnatural Selection: Seeing Human Intelligence in Artificial Creations. Journal of General Artificial Intelligence, 6(1), special issue on Computational Creativity, Concept Invention, and General Intelligence, pp 5-20.

Veale, T. & Alnajjar, K. (2016). Grounded for life: creative symbol-grounding for lexical invention. *Connection Science*, 28(2):139–154.

Veale, T. & Cook, M. (2018). *Twitterbots: Making Machines that Make Meaning*. Cambridge, MA: MIT Press.