Changing Channels:

Divergent Approaches to The Creative Streaming of Texts

Tony Veale,

School of Computer Science, University College Dublin

Abstract

Text is an especially malleable medium for human and machine creativity. When guided by the appropriate symbolic and/or statistical models, even a small and seemingly superficial change at the formal level can result in a predictable yet profound change at the semantic and pragmatic level. Text is also a virtually unlimited resource on the web, which offers abundant, free-flowing channels of topical texts for almost every genre and register. In this paper we consider diverse approaches to transforming these input channels into new and creative streams of machine-generated outputs. We focus on the specific kind of linguistic creativity associated with metaphor, yet also demonstrate that divergent approaches to metaphor generation can, in turn, enable divergent uses and applications for machine creativity.

1. Introduction

Creativity is a vexing concept to define in formal terms, as any phenomenon that blurs the borders of conventional categories is unlikely to possess clear category boundaries of its own. Metaphor, as the preeminent application of creativity in the linguistic domain, is certainly no different, and so the lines that separate this trope from others can be exceedingly slippery (Barnden, 2010). Computationally, the situation is more complicated still: metaphors can be living, frozen or dead, deliberate or unconscious, explicit or implied, or in some cases, even wrongly inferred where none was ever intended. What is clearly metaphor to one speaker may be viewed as a literal truth by another, and so a divergent metaphor-capable machine may see potential metaphors everywhere it looks. Utsumi (2007, 2011) notes that interpretative diversity is a key feature of human metaphor-making, one that listeners often seek to maximize by choosing the interpretation strategy that widens rather than narrows the space of possible inferences one can make. A machine with multiple such strategies at its disposal – e.g. metaphor as elided simile, metaphor as analogical mapping, metaphor as counterfactual – can surely do likewise. The more divergent its wordview, the more potential metaphors our machine is likely to see, and the more interpretations it is willing to countenance.

If divergence is central to human linguistic creativity, why should we assume there must be a single "master" algorithm for metaphor detection, interpretation or generation? It makes more sense to model our creative sysems as assemblies of parallel sub-systems that funnel a panoply of data channels into a range of internal representations, where they can each be subjected to a broad array of value-adding algorithms. The evolution of creative language processing has largely mirrored that of machine translation, inasmuch as symbolic systems of hand-crafted structures and deep but brittle rules have steadily given way to more robust statistical approaches that train themselves – with apt supervision – on large amounts of shallow usage data (see Veale, Shutova & Beigman-Klebanov (2016) for a detailed survey). Sheltering under both sides of the symbolic versus statistical dichotomy is a surprising breadth of approaches that each offer their own special affordances to the language processing pipeline. Just as we might view the diversity of raw data sources that fuel these approaches as alternate channels to sample, so can we view the outputs of these diverse systems as useful channels of processed information for other systems to sample in turn.

So we should be just as divergent in the design of metaphor-capable machines as in the design of the metaphors they produce, for there are surely as many ways to be divergent as there are ways of exploiting the fruits of divergence. Tools and resources that were created with a single aim in mind might thus be reused or retrofitted with new functionality for new creative ends. The research of Oliviero Stock and his colleagues, whom we celebrate in this special issue, gives a very practical form to this view of divergent design. Thus, for example, as shown by Guerini *et al*. (2011) and Gatti *et al*. (2014, 2015, 2016), the same techniques may be applied to a breadth of data and knowledge to yield outputs of novelty and value, just as different techniques may be applied to the same data and knowledge to enable creativity in another context or another application. In the next section we explore the unifying metaphor of multichannel television to harness this diversity in a range of creative systems. For just as digital television offers consumers a wealth of parallel information streams to choose from, we argue that consumers of computational creativity – whether human or machine – should also be afforded a means of channel-hopping between competing streams to satisfy their creative needs. In the following sections we thus present a range of alternate metaphor channels, from a basic channel of atopical metaphors to value-added channels that integrate their metaphors with topical news, user personalities, abstract images or dramatic plot lines. In doing so we demonstrate the value of multiple complementary approaches that make a virtue of their deep differences. The paper concludes with a consideration of the value of divergence, cooperation and reuse in the building of creative computational systems.

2. Interdimensional Cable TV

We have all experienced the kaleidoscope of sounds and images that emerges from an evening spent in front of the television, hopping from channel to channel in search of content that can hold our attention. Channel hopping exposes us to a rapid succession of genres, topics and content registers, so it is hardly surprising if we allow the norms of one channel's content to colour our first impressions of the next. Cartoon violence on one channel can subliminally shape our view of a political argument on the next, while a sitcom's laughtrack may still ring in our ears as we switch to news coverage of a natural disaster. Each kind of content belongs in its own possibility space, but rapidly switching channels can cause the boundaries between these spaces to creatively, if only temporarily, dissolve.

Cognitive theories of creativity frequently stress the channel-switching nature of innovation. For Koestler (1964), creativity occurs at the blurred boundaries of two mental spaces, or what he called "matrices," via an act of "bisociation," while Raskin (1985) has argued that narrative jokes rely on a deliberate switch of plot scripts that forces listeners to rapidly switch between semantic frames. Lakoff & Johnson (1980) champion a conceptual view of metaphor as an image-schematic mapping from a source to a target space, while Gentner (1983) has argued that this mapping sits at the heart of didactic, scientific and problem-solving analogy.

Aristotle was the first to arrive at these insights, outlining in *The Poetics* a two-space theory that embraced analogy as just another kind of metaphor, but in modern times the idea of a semantic space has been given a robust mathematical form in vector space models (VSM), statistical constructs that are automatically built from the co-occurrence patterns of words in large document sets. Kintsch (2000), for instance, tackled the metaphor-understanding problem by mapping ideas into a high-dimensional VSM that is defined by the texts the ideas are found in. Though a VSM does not distinguish between literal and non-literal uses of a word or phrase, it can capture the shared associations and common dimensions that implicitly link the literal and non-literal meanings of the same expressions.

Divergent processes, in the sense of Guilford (1950), do not yield a single right answer but an entire space of more-or-less useful possibilities. The meaning of a novel metaphor cannot thus be captured in a static set of propositions (as argued in Davidson, 1978) but must be viewed as a dynamic possibility space that listeners are invited to enter and reason within. In this vein, Fauconnier and Turner's (1998, 2002) theory of conceptual integration networks, commonly called Conceptual Blending Theory, models creative meaning construction as the selective projection of elements from multiple input spaces into a single output space. This output, named "the blend space," is a sandpit for mental simulation that allows the consequences of creation to become readily apparent, perhaps in ways that permit the emergence of new perspectives on the ideas that contribute to a blend. In effect, two or more channels of domain knowledge are interwoven to produce an output channel that may, in turn, serve as an input to other blends. Veale, Feyaerts and Forceville (2013) have argued that the creativity of a blend resides not so much in how the inputs streams are "braided" together, to use an analogy from Thagard and Stewart (2010), but in how those specific streams are chosen in the first place when so many other input streams compete for our attention. We are all channel-hoppers in the multiverse of possible input spaces when we set out to create a novel blend from established information sources.

Thagard and Stewart (2010) view the "braiding" of multiple input streams as a convolutional process that binds separate patterns of neural activation into a single functional whole. They see this braiding as occuring at multiple levels, to produce parallel streams of convoluted outputs, from the conceptual content of the blend itself to the physiological reaction to this new content. In exemplary cases the latter produces what is called, in folk parlance, an *AHA! moment*. These new output streams can be channeled with varying resolution to different ends, but what of the input channels that are convoluted to generate them? Those neural activation patterns must somehow capture our partial theories of the world, which Koestler termed "matrices" and which Fauconnier and Turner label "mental spaces." Proxies for these theories can be found in large streams of raw data that reflect the stimuli that shape our conceptual perspectives on the world. The Google n-grams database (Brants and Franz, 2006) offers a vast selection of keyhole views (of 1 to 5 words each) onto patterns of language use on the web. Each value of *n* can be considered a different channel designator for streams of the corresponding n-grams, allowing a data-driven model of conceptual blending (as e.g., outlined in Veale and Li, 2011) to scan these channels in search of n-grams (and the stimuli they derive from) to motivate new creations of its own. In the next section we explore such an approach to generating a channel of novel metaphors that finds its inspiration in the keyhole observations of web n-grams.

3. The Metaphor Channel

Human metaphor-making is a situated process, fueled by a continuous stream of experiences that range from the personal and the immediate to the public and the mediated. Since a metaphor-making machine is not physically situated in our world, or at least not in the same way as we are, it must take inspiration for its metaphors from an entirely different source. In place of the personal interactions and conversational hubbub of everyday life it can find stimulation in the n-grams of the web. A metaphor machine might randomly sample these n-gram channels in genteel sips or drink them all in wholesale, as though drinking from a firehose. The most obvious n-grams to sample are the 3- and 4-grams that conform to the copula patterns "A is B" and "A is a|an B." It is in these standard containers that conventional metaphors like "time is money" and "life is a journey" can be found, yet if our goal is to find grist for novel metaphors on the web, our machine must seek out a stream of n-grams that are not themselves obviously metaphorical.

The creativity of a metaphor resides as much in the consumer that detects and appreciates it as in the producer that conceives and packages it. Metaphor is not a wholly objective phenomenon, so words that are offered with the plainest of intentions can still be granted a figurative meaning by those who strive to see it. Conversely, many metaphors are deliberately constructed to occupy a grey area between simple literalness and witty suggestiveness. Consider the title of Lakoff (1987)'s book on categorization, "Women, Fire and Dangerous Things." Lakoff's implication that women can be as dangerous as fire seems clear enough, yet the title wittily preserves a degree of plausible deniability. It is the reader who must take the last step in connecting these three puzzle pieces into a single category. While we presume the author's metaphoric intent is deliberate, the most we – or a machine – can definitively take from the title is that it is a *potential* metaphor. An author can nudge us toward a desired interpretation for a potential metaphor but cannot dictate the meaning we arrive at. In this respect the author is no more privileged than the reader in determining the validity of any figurative leap. John Steinbeck may or may not have wanted his readers to construct a metaphorical mapping between mice and men in his 1937 novel, but readers who see the title on a bookshelf, or a machine that sees the 3-gram "mice and men" in Google's n-gram database, are free to process the potential metaphor as they see fit. The physical world offers dense channels of potential metaphors to those who amble through it, from signs to headlines to pop lyrics to stray snippets of conversation, but large n-gram models of web content offer an ample substitute for a machine.

A machine can generate metaphors of its own by inventing unambiguous new contexts for the potential metaphors that reside in abundance on the web or in the Google web n-grams, thereby turning potential metaphors into deliberate metaphors. Consider the 3-gram "romance and insanity," to which the Google n-grams assign a count of 313 documents. Read as a simple coordination structure the phrase says as little as "women and fire" or "mice and men," yet engaged readers will surely ponder the reasons for squeezing two ideas of conflicting sentiment into a single phrase, and will set out to unearth a figurative kinship to reconcile the two. With the help of the new context imposed by the machine, and the knowledge of words and the world at their disposal, engaged readers can beat a path from madness to love, as in the following machine contextualization:

It used to be that romances were enjoyed by beloved lovers. Now I say unto you that romance is insanity from which only hateful fanatics suffer.

So while this framing may speak to readers who have experienced the highs and lows of romantic attachment, it was produced by a machine that draws on a large stock of stereotypes (e.g., lovers and fanatics) and familiar situations (e.g., falling in love, suffering from an affliction). These archetypes allow the machine to find myriad symbolic connections behind the words "romance" and "insanity." The use of religious phraseology (e.g., "I say unto you") is just one of many templates the machine uses to contextualize its simple insights as deliberate metaphors so as to provoke an emotional and intellectual response in readers. The machine is named *MetaphorMagnet*, and the outputs can be sampled hourly in the tweets of the Twitterbot @*MetaphorMagnet* (Veale, 2016). Not all of its outputs are gems, but every one is available for inspection on the bot's Twitter timeline, where readers are invited to judge its figurative acumen for themselves. The machine views its symbolic representations of familiar ideas – love, romance, insanity, etc. – as jigsaw pieces that might connect together, either directly or via the insertion of an implied third piece such as fanaticism. To decide which words and ideas and permutations thereof it should explore, the machine takes its cue from the potential ambiguities of the short, underspecified phrases of the Google n-grams.

As described in Veale (2016, 2017), *MetaphorMagnet* is a knowledge-based system that relies on many simple strategies – from inference rules to syllogistic templates – to frame the implicit consequences of its own symbolic knowledge in figurative terms. Its data-base of 75,000 stereotypical associations – such as that cowboys are swaggering or that zombies are braindead – is populated by finding hypotheses in the Google n-grams (e.g. the 2-gram "swaggering cowboy") that can be validated by reformulated queries to Google web search (e.g., the query "swaggering like a cowboy" has 2046 hits while "as braindead as a zombie" has 6 hits). A sentiment lexicon (Veale, 2013) allows the system to label associations as positive or negative to varying degrees, while an antonym dictionary allows it to determine which pairs of associated ideas exhibit an inherent semantic tension. It uses the Google n-grams to suggest the pairs that are most worthy of attention, and frames the results using a range of templates that evoke different emotions;

the template above, "*It used to be that X were V-ed by A Ys; now I say* …," evokes a nostalgia for the past that makes the machine seem positively wistful.

This approach makes a virtue of the lack of context that is a feature of n-gram models, for it is this lack that allows a machine to impose its own context on an n-gram. Our machines, or indeed we, cannot know if a 4-gram like "research is the fruit" is a well-formed component of a larger metaphor that we cannot see, or a misleading keyhole view onto a literal text such as "the focus of this research is the fruit of the cacao tree." The metaphor machine does not concern itself with the intentions of the original author, but with its own ability to see a figurative kinship between parts of a metaphor-shaped text. In this case it finds the kinship of proportional analogy, which Aristotle identified as a fourth kind of metaphor in *The Poetics*. The machine's symbolic knowledge-base holds a number of norms about fruits and research: e.g., that fruits grow on trees in orchards, are tended by farmers and eaten by insects; that research is funded by rich backers, carried out in labs, published in academic journals and conducted by scientists, clinicians and philosophers. In seeking out a bridge from the former to the latter, it sees a tension between the lowly status of insects and the prestige of philosophers, allowing it to contextualize the *fruit:insect::research:philosopher* analogy thusly:

Remember when research was conducted by prestigious philosophers?
Now research is a fruit eaten by lowly insects.

This framing brings the antonymous qualities *prestigious* and *lowly* to the fore. Conversely, the need to find such a resonant opposition places important curbs on the machine's ability to generate, or indeed *over*-generate, new metaphors. Its knowledge simultaneously offers new avenues for search while constraining possible points of connection. But how much value is added by using knowledge to broaden and narrow in this way? Suppose the machine were to choose its words randomly, and shift the burden of identifying figurative potential from the machine to the human reader? To quantify the value of knowledge to metaphor generation, we can compare the outputs of the *@MetaphorMagnet* channel on Twitter to that of a simpler Twitterbot named *@MetaphorMinute* (from internet artist and noted bot-builder, Darius Kazemi). The latter uses an online dictionary service (*Wordnik.com*) as a source of often fanciful words, which are shaped into a metaphor using the copula pattern, as in this enigmatic example: "*an astrolabe is a tapioca: lubberly yet species-specific*." Since the bot can use any of the words on *Wordnik.com*, it has a larger potential lexicon than that of *@MetaphorMagnet*, while the latter's rich stock of templates and framing strategies is significantly larger than that used by *@MetaphorMinute*. But even when we limit the choice of template used by *@MetaphorMagnet,* this machine uses semantic criteria to fill its slots. A comparison of both bots can thus show the true value of these criteria.

We used the crowdsourcing platform *CrowdFlower* to elicit human ratings for the metaphors of each bot, first sampling 60 tweets from the generative spaces of each one. For @MetaphorMagnet, we chose at random from a large corpus of its metaphors that numbers in the millions (we return to this corpus in the next section), while for *@MetaphorMinute* we chose at random from a corpus of 1000 examples sampled from its Twitter timeline over a number of days. *CrowdFlower* was used to solicit 10 human judgments along 3 dimensions for each metaphor. Judges, who were paid a small sum per rating, were not told that each tweet was created by a bot. For each tweet, judges were asked to provide a rating from 1 to 4 for the three dimensions *Comprehensibility*, *Novelty* and *Retweetability*, where 1 = *Very Low*, 2 = *Medium Low*, 3 = *Medium High* and 4 = *Very High*. Potential scammers were filtered by requiring judges to pick out the counterpart for the given term in each metaphor (e.g. to pick out "insect" when given "philosopher").

Table 1. *Comparative evaluation of @MetaphorMagnet and @MetaphorMinute outputs (a random sample of 60 tweets each) across three dimensions.*

	Comprehensibility		**Novelty**		**Retweetability**	
User Rating	*Metaphor* ***Magnet***	*Metaphor* ***Minute***	*Metaphor* ***Magnet***	*Metaphor* ***Minute***	*Metaphor* ***Magnet***	*Metaphor* ***Minute***
Very Low	11.6%	**23.9%**	**11.9%**	9.5%	15.5%	**41%**
Med. Low	13.2%	**22.2%**	**17.3%**	12.4%	**41.9%**	34.1%
Med. High	**23.7%**	22.4%	**21%**	14.9%	**27.4%**	15%
Very High	**51.5%**	31.6%	49.8%	**63.2%**	**15.3%**	9.9%

Table 1 reports the distribution of mean ratings for these three dimensions for each bot's tweets. More than half of @*MetaphorMagnet*'s tweets were rated very highly comprehensible, and only 25% were rated as hard or somewhat hard to understand. In contrast, @*MetaphorMinute*'s tweets were rated as much harder to comprehend, yet over half were still deemed to have moderate to very-high comprehensibility. We can infer that the shape of its texts leads raters to assume the presence of a meaning even if none was intended. @*MetaphorMinute* outdoes @*MetaphorMagnet* when it comes to high novelty, likely because judges rarely see such unhinged combinations in human-crafted texts. For the *Retweetability* dimension, judges were asked to rate their willingness to share a given metaphor with their own circle. We asked judges to speculate on retweetability rather than measuring the actual retweet and "like" rates of metaphors in the bots' timelines because neither has enough active followers to yield a statistically significant result (we nonetheless aspire to this scale for future experiments). While *Very-High* scores are elusive for either bot, @*MetaphorMagnet*'s oeuvre is judged to have higher mean retweet value (where 2 in 5 tweets earn medium- to very-high scores) than those of @*MetaphorMinute* (where only 1 in 4 earns a medium- to very-high score). For example, four judges rated this @*MetaphorMagnet* tweet as very-highly retweetable: "*Not all gamblers are reckless. Some are as cautious as the most accountable steward.*" In contrast, this ugly metaphor was judged highly retweetable by only one rater: "Librarian. noun. A pimp who would rather manage dirty libraries than dirty whores." A metaphor from @*MetaphorMinute* that judges rated as retweetable is: "a batman is a retarder: beerless but soggier."

Form can be seductive, and @*MetaphorMinute*'s ratings suggest that human consumers see the potential for metaphor in texts that are appropriately shaped as such. But will readers do the work that is needed to deliver on this potential? In a follow-up experiment with another sampling of 60 tweets apiece, we elided key pieces of information from each machine metaphor, and gave judges the task of choosing amongst several restorative fillings. For @*MetaphorMagnet*'s tweets we blanked the opposing properties that gave each metaphor its tension, such as *beloved:hated* (in the "romance as madness" case) or *prestigious:lowly* (in the "research as fruit" case). Judges were asked to choose an apt restorative pairing of properties from a list of five pairs, four of which were distractors taken from other tweets from the same channel. For @*MetaphorMinute*'s tweets we simply blanked the two adjectives after the copula body, as in: "*a batman is a retarder: __ and __ .*" Judges were again presented with the original pairing, hidden in plain sight among four distractor pairs from the same bot. In either case, a metaphor is said to have *Very-Low* aptness if judges choose the original pairing from amongst the distractors less than 25% of the time; *Medium-Low* aptness if chosen less than 50% of the time; *Medium-High* if chosen less than 75% of the time; and *Very-High* if chosen 75% or more of the time. Our findings are shown in Table 2:

Table 2. A comparative evaluation of the aptness of word choice in metaphors.

Rating	***@MetaphorMagnet***	***@MetaphorMinute***
Very Low	0%	**84%**
Medium Low	**22%**	16%
Medium High	**58%**	0%
Very High	**20%**	0%

When a metaphor machine is not attuned to the meaning of words, all its random novelty may be for naught. Such a machine's outputs, while difficult to predict at a local level, all blend together to form just another channel of white noise.

4. The News Channel

We speak of "the news" as though it were a single, definitive concept, yet a night spent channel-hopping reveals a wide diversity in the content that is considered newsworthy, and an equally broad variety in the containers used for its delivery. There is a cable news channel to suit every demographic, from young to old and far left to extreme right. Since each creates and reinforces its own belief space, we might be tempted to create a distinct vector space to model the viewpoints and fixations of each. However, such belief spaces give rise to distinctive styles and lexical preferences that openly reveal their political biases. So a right-of-centre news source may favour the word "homosexual" over the left-of-centre preference for "gay," and may well place the word marriage in scare quotes when discussing the topic of same-sex civil unions (Gentzkow & Shapiro, 2010). A user who subscribes to a wide variety of news feeds is effectively constructing a single heteogeneous channel that covers all socio-political bases. A metaphor machine can do likewise, and construct a single vector space from the collected tweets of diverse news sources. News tweets about the same topic or event will thus be mapped into similar vectors that present a small angle to each other, regardless of their source, while those news tweets that addionally spring from the same possibility space (E.g. Fox News, or CNN) will be even closer still in vector space.

Consider the problem of generating apt metaphors for the news. As a news story breaks and headlines stream over Twitter, we want our metaphor machine to pair an original and insightful metaphor to each arriving headline. A headline about extreme weather might be paired with a metaphor about the destructive power of nature, whilst a political scandal might be paired to a crime metaphor. Since theorists often speak of multiple spaces in metaphor (e.g. Koestler, Lakoff and Johnson, and Fauconnier and Turner all see different viewpoints as different spaces), it is tempting to assume that we should model each space in a metaphor with its own VSM, that is, by equating vector spaces with conceptual spaces. But this analogy is misleading, since different VSMs – constructed from different text corpora – will have different dimensions (even if sharing the same *number* of dimensions) and so 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 tests, it is better to build a single vector space that imposes the same dimensions on each conceptual space in the metaphor. It is more useful then to view news headlines and metaphors as comprising two overlapping subspaces of the very same VSM.

For our 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 and Dumais, 1997) or *Word2Vec* (Mikolov, 2013) – to produce a vector of fixed dimensionality for each headline. We also build a large metaphor corpus by running *Metaphor Magnet* on the Google n-grams overnight, yielding millions of metaphors that stretch across many topics and registers. Rather than build a separate vector space for each of the news and metaphor corpora, we build a *single* vector space for both by appending one corpus onto the other before applying compression. Within this joint VSM, all past metaphors and future headlines can be assigned a vector of precisely the same dimensions. 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 (with the largest cosine) to an incoming news vector is chosen as the one with the most thematic relevance to that news item. We build our joint space by compressing 380,000 news stories, 210,000 news tweets from sources including *@CNNbrk*, *@FOXnews* and *@nytimesworld*, and 22,846,672 metaphors from *MetaphorMagnet* into the same 100-dimension LDA space using the *gensim* package of Řehůřek & Sojka (2010). Word lemmas were concatenated with their POS tags to provide additional features for the model. Consider this pairing:

> **News Headline** (*@FOXnews*): *Gina Haspel confirmed as CIA's first female director: 5 things to know about the career spymaster*.
>
> **Paired Metaphor**: *Terrorists commit terrors. Gods create the monsters that practice the evil that promotes terror. Who is worse?*

Who are the terrorists here and who are their victims? The metaphor refuses to say, just as the following is coy about the true targets of its critcism and praise:

> **News Headline** (*@FOXnews*): *.@POTUS: "We're doing very well on, as you know, North Korea... We have a meeting set up, we have the location all done, we have the time and place all finished now."*
>
> **Paired Metaphor**: *What is a coward but a selfish neighbor? What is a neighbor but a generous coward? Which one are you?*

These pairings were broadcast by the metaphor Twitterbot *@MetaphorMirror*. Since our bot tweets just once an hour, it can afford to perform an exhaustive comparison of incoming headlines against its full inventory of 22M metaphors. The LDA space of 100 dimensions allows the tacit themes of each headline (e.g., CIA torture programs, as personified by the controversial Gina Haspel, or the militarized border that separates North from South Korea) to be paired with the more-or-less explicit themes of each metaphor (e.g., terrorism, disagreeable neighbors). Our VSM conflates an array of news sources of differing agendas, but we might instead encourage diversity in how we compress our texts into this joint space. Suppose we instead use LSA to build our vector representations, or the neural embeddings of Word2Vec. Using the same dimensionality (n=100) we obtain three alternate channels of topical metaphors that are tied to the news.

Which approach to the construction of a joint news and metaphor space yields vectors with the most apt and influential pairings? An *apt* pairing is one in which readers recognize a thematic similarity between a headline and its metaphor. A *comprehensible* pairing is one that seems to make semantic sense to the reader, regardless of its aptness (though recall the separation of sensibility and aptness in the case of @*MetaphorMinute*'s random metaphors). An *influential* pairing is one in which the metaphor actually affects the reader's appreciation of the news. We selected 90 random headlines from recent but not current news (specifically, we drew our test data from a July 2016 subset of the news corpus, recent enough for judges to recall the events when given the headlines) and used a variety of VSM formulations to find the most similar metaphor vector. Both LDA and LSA were tested in two versions: the *fulltext only* setting used the full text of the news stories when building the joint space, but not the text of the additional news tweets; the *fulltext+tweets* setting combines both sources of news content when building a space. For the Word2Vec space we used the settings reported in Gatti *et al*. (2015) for their slogan adaptation system, and averaged the Google News embedding of each word in each headline or each metaphor to obtain a vector representation of each tweet's text. As a simple baseline we randomly selected a @*MetaphorMagnet* metaphor for every test headline, recalling that our earlier evaluation of @*MetaphorMinute*'s simple random texts showed that readers may be strongly inclined to see creative intent where only chance is at work.

CrowdFlower was used to elicit 10 human judgments on a 5-point Likert scale (1 being lowest, 5 highest) for the *aptness*, *comprehensibility* and *influence* of each of our 90 test pairings. Table 3 reports the mean average values for each when using different vector spaces. An LDA model built from a corpus of 380,000 *fulltext* stories (harvested 2000–2012) and a year of more recent news (210,000 tweets gathered from July 2015 and June 2016), outperforms all other settings.

Table 3: *Mean values (+ std. dev.) of each property for different VSM spaces*

Joint Space Model	***Aptness***	***Comprehensibility***	***Influence***
LDA (*fulltext+tweets*)	2.95 ± 1.27	3.59 ± 1.05	3.01 ± 1.24
LDA (*fulltext* only)	2.78 ± 1.04	3.54 ± 0.92	2.75 ± 1.03
LSA (*fulltext+tweets*)	2.62 ± 1.10	2.97 ± 1.09	2.44 ± 1.01
LSA (*fulltext* only)	2.40 ± 1.12	2.99 ± 1.15	2.49 ± 1.14
Word2Vec	2.65 ± 0.99	3.38 ± 1.02	2.73 ± 1.00
Random baseline	2.20 ± 1.20	2.54 ± 1.12	2.09 ± 1.24

To go beyond aggregate means, we placed mean judgments for all test cases into four equal-sized bins, *Low* (< 2), *Average (2 to 3)*, *Good (3 to 4)* and *Very Good (> 4)*. Table 4 shows the percentage of pairings in each space that fall into each bin.

Table 4: *Distribution of mean aptness across 4 quality bins for all test pairings*

Joint Space Model	***Low***	***Average***	***Good***	***Very Good***
LDA (*fulltext+tweets*)	1.1%	47.8%	**41.1%**	**10%**
LDA (*fulltext* only)	3.3%	**65.6%**	30%	1.1%
LSA (*fulltext+tweets*)	10%	60%	30%	0%
LSA (*fulltext* 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%

Since the LDA space of full-text news stories *and* tweets is the only space that places more than half of its pairings in the *Good* or *Very Good* bins, it is from this space that our *@MetaphorMirror* bot takes its topical metaphor / news pairings.

5. The Arts Channel

Shakespeare wrote that a rose by any other name would smell just as sweet, but would this alternate name be just as effective as a metaphor? It seems likely that if we all chose to refer to a rose as a "*goreweed*," a "*turdblossom*" or a "*prickbleed*" we would need to find very different poetic uses for this familiar flower. For our metaphors do more than evoke lexical semantics in the mind of a reader, and the very best can tap into our memories and perceptual faculties to create a feast for the senses, one that is as rich in colour, texture and aroma as it is in semantic meaning. So when we bend our machines to the generation of novel metaphors, we must ensure they are as adept with the multi-modal connotations of words as they are with their denotative semantics. We can go further still, and task our machine with generating multi-modal blends of words and images that visually enrich a linguistic metaphor and thereby squeeze more meaning from its words than any text alone can manage. These visual and textual elements may be further grounded in the social and the personal if the metaphor is deliberately crafted to reflect the behavioral traits of a real person in its target audience.

Much research has been conducted on the analysis of human personality as reflected in lexical choice. Chung and Pennebaker (2008), for example, describe a tool and a resource, named the LIWC (*Linguistic Inquery and Word Coun*t) for estimating author qualities such as anger, positivity, worry, anxiety, affability, arrogance, analyticity, awareness, topicality, excitability and social engagement from one's text outputs. An online version of the tool (at www.analyzewords.com) infers values for these 11 dimensions from the recent tweets of any Twitter user one cares to mention. For instance, *AnalyzeWords* tells us that @*Oprah* is upbeat as a Twitter user, while @*realDonaldTrump* is often both upbeat and angry. To generate metaphors about a particular person, such as Oprah or Donald Trump or anyone else with a Twitter account, we can treat their *LIWC/AnalyzeWords* profile at any given time as a vector in an 11-dimension space, and seek to map this vector into a higher-dimensional space of richer metaphorical possibilities.

Given the disparity in dimensions between these spaces (11 vs. 100) and the very different means of their construction, we cannot build a joint vector space by simply merging their underling data. Lacking a training set to train a neural network to map across the spaces, we adopt a symbolic approach to inflate the *AnalyzeWords* space to hundreds of dimensions that represent nuanced qualities.

These allow a machine to map from the 11 core dimensions of an *AnalyzeWords* profile to the typical qualities of an apt source metaphor. We inflate the smaller space by hand-crafting logical formulas – we dub these *transformulas* – to score 300 new dimensions as functions of the 11 core dimensions. Transformulas can conjoin, disjoin and negate the 11 core dimensions. For example, since neurotics tend to over-analyse their concerns, we define the "neurotic" dimension as the product of the "worried" and "analytic" dimensions, while we say that someone is "narcissistic" to the extent they are "arrogant" *and* "sensory" (given to talking about their feelings), or "creative" to the extent they are "analytic" and "upbeat." The point of these transformulas is not that they reflect an empirical truth about a person, rather that this kind of symbolic structure lends itself to explicit verbal explanation. Our metaphor machines are allowed to possess folk theories of the workings of other people's minds, of the kind that human metaphor makers also rely upon. What matters is not that they are experimentally validated, but that they can be clearly articulated as the motivation for the machine's other choices.

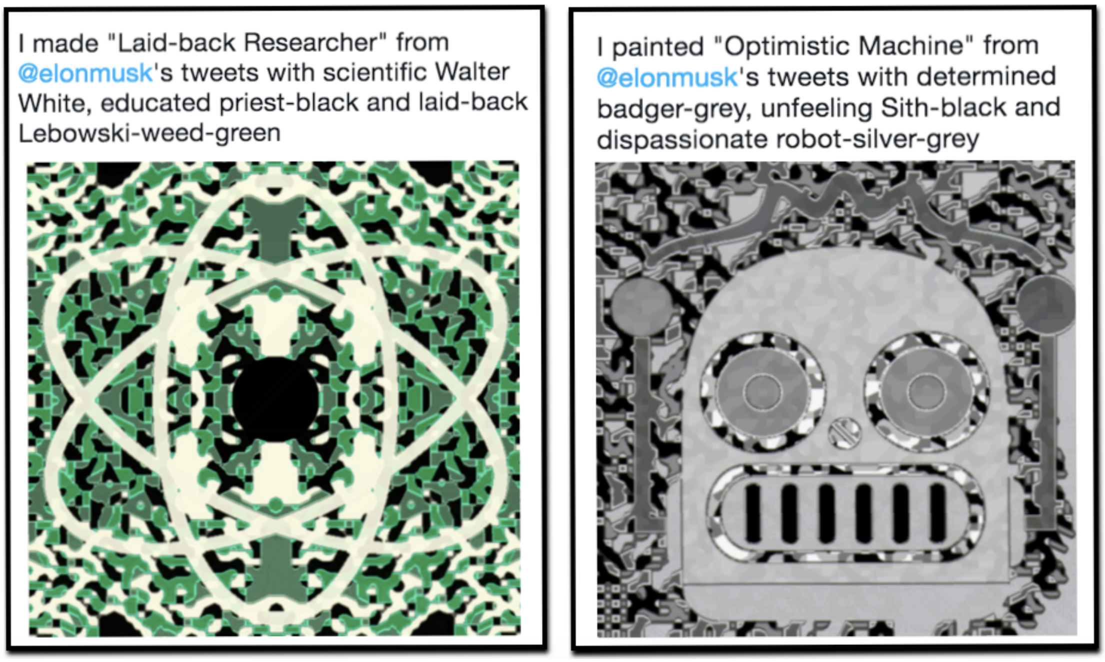

Figure 1. Two metaphors from @BotOnBotAction based on an affective profile of @ElonMusk as produced by AnalyzeWords.com. Each generates a visual metaphor for the profile that is grounded in the linguistic content of the metaphor.

Consider, as an example, the entrepreneur and engineer Elon Musk. From an

AnalyzeWords profile that places his tweets high on the core dimensions *upbeat* and *analytic* and low on the core dimensions *angry* and *sensory*, the nuanced qualities *optimistic* (upbeat and analytic), *dispassionate* (analytic and not angry), *unfeeling* (analytic and not sensory) and *determined* (upbeat and not angry) can be additionally inferred. Three of these – *unfeeling*, *determined* and *dispassionate* – are stereotypical qualities of machines, so a metaphor generator can topically describe Musk (in light of his most recent tweets) as an "optimistic machine." The same underlying profile supports the nuanced qualities *laid-back, educated* and *scientific*, the latter two of which are typical of researchers, allowing our metaphor machine to aptly describe Elon Musk as a "laid-back researcher."

The first of these metaphors, as actually tweeted by our metaphor-producing bot @*BotOnBotAction*, is shown in Figure 1 (right), with the second in Figure 1 (left). The bot also creates a new piece of visual art to complement each metaphor, in which an abstract pattern – a random 1-dimensional cellular automaton unfurled over many generations/rows – is rendered with colours that are chosen to match the qualities highlighted by the metaphor (Veale & Cook, 2018). As an additional flourish, the bitmap of an Emoji annotated with one of the words in the metaphor – an atom for "scientific" in Fig. 1 (left) and a robot for "robot" in Fig. 1 (right) – is integrated into the image and recoloured to suit its new context. Each colour grounds a different aspect of the textual metaphor in a process of situated metaphor generation. Using a colour lexicon in which 600 of the metaphor machine's stereotypes are hand-mapped to the RGB codes of the colours with which they are most typically associated (e.g., silver-grey for robots, black for priests), it is possible to assign a specific hue to the non-visual qualities of these stereotypes. Each combination is then textually framed so as to cement the link, so that the black of Fig. 1 (left) is named "educated priest black" while the black of Fig. 1 (right) is named "unfeeling Sith black." The metaphor thus concisely tells us what each colour stands for, and how we should feel about that.

Through this act of mutual cross-modal grounding, our metaphor-making bot can squeeze a variety of metaphors into a single tweet. Fig. 1 (right) compares Elon Musk to a determined badger, an unfeeling Sith lord, an optimistic machine and a dispassionate robot, whilst Fig. 1 (left) likens him to a laid-back researcher, the scientific anti-hero Walter White of "Breaking Bad," an educated priest, and the laid-back "dude" of the film "The Big Lebowski." The results are, of course, highly subjective, with the Twitter targets of this treatment being more likely to recognize themselves in only the most flattering metaphors. We can, however, evaluate the machine's ability to invent new names for the hues in its visual metaphors, and compare this ability to human performance on the apt naming of new colours. To begin with, we task our machine with inventing new colours and names for these colours. We can then seek out the original names that humans invent for these (or very similar) colours on a website dedicated to the naming of colours by passionate amateurs, *ColourLovers.com*. Finally, we ask volunteers on CrowdFlower to express their preferences for one name over another for the same colour, without being told the human or machine provenance of either one.

We can generate new colours and new names for those hues in a single pass through the Google 2-grams. Simply, any 2-gram phrase in which both words are stereotypes in our colour lexicon, such as "paper tiger," "midnight sun," "storm raven," "strawberry milk" and so on, are harvested, and a corresponding colour for the whole is generated as a 50:50 mix of the RGB colours of each stereotype. Thus, "strawberry milk" is mapped to a hue that is 50% strawberry-red and 50% milk-white. The resulting RGB code is then sought out on *ColourLovers.com*, with the conditions that the code is sufficiently close to one with a human-crafted name on the site, and that name has been "loved" (up-voted) at least once by the users of the site. We can now compare a machine-crafted name (e.g. "strawberry milk") to an up-voted human name (e.g., "sparrow in the wind"). As described in Veale and Alnajjar (2016), we impose a further test to select the colours and names that will be evaluated on CrowdFlower: the stereotypical elements that are combined (e.g. "strawberry" and "milk") must be close enough in RGB space that the blend has recognizable inputs from each (e.g. it must be recognizably "reddish" and "whitish"). 2587 colours are ultimately selected for evaluation, with a mean count of 2.188 "loves" each. To each code we attach a swatch of the corresponding colour and its human-assigned and machine-crafted names. The raters on CrowdFlower see these names in a randomly-assigned order, so that the ordering cannot tacitly influence the judges' preferences for either.

A budget of $220 was exhausted after 940 raters were paid for their answers, at which time 1578 of 2587 colours had 5 trusted judgments for these questions:

Q1. *Which name is more descriptive of the colour shown?*

Q2. *Which name do you prefer for this colour?*

Q3. Which name seems the most creative for this colour?

We required that each question be answered by 5 fully engaged raters. As raters were timed on their responses, any that spent less than 10 seconds presenting their answers for any colour was ignored. In this way we elicited 12,608 trusted judgments, while 5,040 untrusted judgments were discarded. Overall, 70.4% of trusted judgments for *most descriptive name* (Q1) favored the machine, as well as 70.2% of judgments for *most preferred name* (Q2) and 69.1% of judgments for *most creative name* (Q3). Tallying the majority judgment for each question and colour, a majority of three or more raters favored the descriptive power of the human-assigned name in just 23% (354) of cases. The results for the other two questions are in line with this finding. Only in 355 cases did a majority of the 5 judges for a given colour express a greater liking for the human-crafted name, and only for 357 colours did a majority consider the human-assigned name to be the more creative of the two. Whilst we cannot assume that the colour "lovers" who provide these human names are either wordsmiths or experienced colour professionals – we leave the evaluation of professional names to future work – this clear 3-to-1 breakdown in favour of the machine suggests that a channel of machine-crafted colour metaphors can be competitive with one stocked with the linguistic choices of passionate and creative laypeople.

6. The Drama Channel

The metaphors of previous sections are overtly framed as deliberate figures so that audiences can readily appreciate them as such. As we saw in the comparison of *@MetaphorMagnet* to *@MetaphorMinute*, readers assume comprehensibility from the overt framing of a metaphor, even when this comprehension ultimately proves elusive when challenged with a cloze test. But a metaphor need not be explicit to exert a profound influence on our view of a topic. Consider the pairing of protagonists to well-matched antagonists in dramatic tales. It is often said that heroes are only as good as the villains that motivate them, because narrative tension requires a credible sequence of actions and reactions. So a well-matched antagonist shares a number of key qualities with a protagonist, but will surely possess complementary and antithetical qualities too. Think of Batman and the Joker, which share a propensity for theatrical costumes and dramatic gestures. Neither is motivated by the acquisition of wealth, and each is antisocial in his own way. One works to preserve order, the other to undermine it. Each serves as a dark and twisted metaphor for the other because that is what metaphors do best: they integrate similarity and contrast into a coherent, unified whole. When choosing characters for its own narratives, a computational storyteller can thus use a capacity for metaphor to pair each protagonist with an apt antagonist.

A good narrative must integrate character with plot. For the former we draw upon a database of vivid famous characters called *The NOC List* (see Veale, 2017). With more than 1000 characterizations covering everything from gender and political leanings to love interests, rivals, typical clothing and actions, weapons, vehicles, professions and positive and negative talking points, this database will be the source of the characters in our newly generated stories. For the latter we draw on a model of plot construction called *Scéalextric* (Veale & Valitutti, 2017) that provides a dense space of causally-connected plot verbs. *Scéalextric* defines over 800 such verbs, from *love* to *are_betrayed_by*, which it connects with a directed graph of labeled edges; these labels indicate the causal link between an action and its successor, as in "so" for the link from *admire* to *are_seduced_by* and "but" for the link from *fall_in_love_with* to *are_dumped_by*. A plot is a non-cyclic sequence of actions marked out in this causal graph from a given starting vertex (first verb) to a specific end vertex (last verb). Starting from an initial action, any random walk in this graph can serve as a locally-coherent plot. For our purposes here, this initial action is either chosen at random, or is chosen to reflect the system's choice of protagonist (A) and antagonist (B). Every plot is conceived as a back-and-forth series of interactions between placeholders A and B, which can be instantiated with specific characters once the plot is built.

A plot is generated by generating a random walk of *n* steps in the causal graph from a given starting action. For our current purposes, *n=8*, so that each plot has exactly 9 actions in the simple narratives to follow. In *generic* mode, the starting vertex is chosen randomly, whilst in *NOC* mode – wherein A and B are chosen to be a metaphorically apt pairing of two different characters in the NOC – the start is dictated by our choice of A and B. Suppose A is chosen to be *Doctor Strange*, and B is chosen to be *Stephen Hawking* (the NOC indicates that both characters were portrayed by the actor Benedict Cumberbatch). Since *Strange* is a doctor, the initiating plot action can thus be either *diagnose*, *treat* or *cure*. In place of the NOC, generic mode randomly selects two entries on a list of Aesop-style animals, such as the frog and the wasp or the bear and the caterpillar. These animals have no prior characterizations in the system, so a starting action is chosen randomly. In any case, and in either mode, any of the 800 story actions can start a story. Whether stories are generated in NOC mode or generic mode, the random walk that follows is generated in precisely the same way along exactly the same graph. The walk follows the causal possibilities of the space, as represented by the links between story actions in the graph, and is not influenced by the need to produce plot twists or turns. As such, any twists emerge randomly from the walk, since any random walk in a well-formed causal graph will yield a structurally-sound plot. Since a random walk may also end at any action vertex, we eke out plots for our experiments from the causal graph that are each 9 story actions in length.

The NOC list is populated with characters that serve as vivid archetypes of recurring human personality types. *Richard Nixon*, for instance, is defined by his mix of ambition and paranoia, while the fictional *Frank Underwood* of the show *House of Cards* is defined by his mix of ambition, cunning and charm. Each is a politician that has achieved the role of president, and each shares a sufficiency of positive and negative talking points to form a metaphorically apt partnership. The fact that *Nixon* is real and *Underwood* is fictional merely adds to the tension of the metaphor, and serves to lend the resulting tale a hint of postmodern irony. The further observation, also available from the NOC, that actor Kevin Spacey has portrayed both characters on screen only serves to cement the new pairing. In short, the NOC is used to suggest apt pairings in which two characters share obvious overlaps, and an obvious contrast too, whether it is fictional versus real, ancient versus modern, male versus female, or even franchise versus franchise (e.g. Indiana Jones versus Han Solo or Jack Ryan). Incongruity is widely accepted as a key element in the generation of any humorous effect, provided it can be packaged in a context in which it can be motivated and resolved (see Raskin, 1985; Veale, 2005). The metaphorical pairing of NOC characters to evoke both similarity – whether a shared creator, actor, domain, profession, spouse or group affiliation -- and contrast – whether in gender, era, fictionality, genre or franchise – satisfies both sides of this humorous bargain to achieve deliberate incongruity.

The vividness of these NOC representations allows for greater vividness in the rendering of the resulting tales. Specific locales and props – what are often called the *mise en scène* – can be injected into the rendering of the individual plot verbs that can best make use of them. For instance, an apt weapon can be suggested for an *attack* action, or an apt hiding place can be suggested for *a hide_from* action. Plot verbs that depict speech acts such as *insult* and *compliment* can additionally be rendered using specific knowledge of the characters involved. To render a compliment by A of B, the system need only dip into B's positive talking points in the NOC; likewise, to render an insult of A, it can use A's negative talking points. Indeed, the NOC can be used to generate embedded metaphors in these speech acts, so that e.g., Underwood insults Nixon by comparing him to *Al Capone*, or Nixon insults Underwood by comparing him to the psychopathic *Keyser Söze* of the film *The Usual Suspects* (who was, incidentally, also played by Kevin Spacey). NOC-based metaphors such as these should appreciably increase the perceived vividness and humour of any plot generated using the basic *Scéalextric* model. So further crowd-sourcing experiments were conducted to test these predictions.

In all, 50 generic stories and 50 NOC-based stories were randomly generated and evaluated along multiple dimensions by the human judges on CrowdFlower. Though each kind was rendered differently, with the latter incorporating specific details from the NOC list, each tale used precisely the same *Scéalextric* plotting mechanism. Although many factors influence a reader's enjoyment of a narrative – for example, whether an odious character gets his come-uppance, or whether a virtuous character finds her reward – we expect that these factors will balance themselves out in a random sampling of all the stories that can be generated. Judges were not informed as to the mechanical origins of the narratives, but were simply told that each was harvested from Twitter. 10 ratings were sought along 6 dimensions for each story. These were: *laughter* (how likely is the story to make you laugh?); *entertainment* (how entertaining is the story?), *imagination* (does the story show evidence of real imagination?), *vividness* (how memorable are the elements of the story?), *silliness* (how implausible is the story?) and *drama* (how eventful is this story?). Judges were shown just one story at a time and asked to rate just one dimension of each, on a scale of 1 (low) to 5 (high). A pool of judges was provided by CrowdFlower, allowing 10 ratings per stimulus to be averaged. Disengaged raters were detected and discarded by requiring them to answer an extra question with a known answer (e.g. "how many words are there in this question?"). Judges were not asked to read every story or to rate every dimension, and were further asked to rate each dimension in isolation, so as to discourage any from reusing ratings for dimensions they might implicitly believe to be correlated (e.g., imagination and vividness). Moreover, they were not asked to directly rate the creativity of stories, as notions of what constitutes creativity, and how to elicit scores for such notions, can vary significantly (see Jordanous, 2018). Table 5 compares the results for each kind of generated story.

Table 5: *Mean ratings (and standard deviations) per dimension for each story type*

Dimension	NOC-based	Generic	All
Laughter	**3.02** *(1.13)*	**2.33** *(1.13)*	2.67
Entertainment	**3.10** *(1.15)*	2.**91** *(1.12)*	3.00
Imagination	**3.43** *(1.04)*	**3.03** *(1.08)*	3.23
Vividness	**3.48** *(0.97)*	**2.90** *(1.05)*	3.19
Silliness	**3.47** *(1.04)*	**2.98** *(1.2)*	3.22
Drama	**3.53** *(0.98)*	**2.93** *(1.14)*	3.23

Recall that the NOC-based stories are those that are generated with metaphors at their core: the protagonist (A) and antagonist (B) are chosen to form a bond of similarity and contrast, and related metaphors are embedded into the narrative whenever a speech-act allows for it. Given the wealth of the detail in the NOC, it is hardly surprising that NOC-based metaphors yield stories of greater vividness. These stories start with an action that is apt for the central character pairing, and include an abundance of NOC character detail that creates further opportunities for incongruity. This, in turn, can lead to higher scores for silliness and laughter. Nonetheless, it is surprising is that the most striking increase of all occurs in the dimension of eventfulness/drama. We might expect drama to emerge from the twists and turns of the plot, which owe nothing to the NOC and everything to the forking of the *Scéalextric* graph, yet the very same graph is randomly walked to produce both the generic and NOC-based plots. We are forced to conclude that it is the vividness of the NOC metaphors that elevates the dramatic appeal of the plots they adorn. So eventfulness, the idea that one tale is richer in memorable events than another, owes as much to the characters that realize the events as to the events themselves. Metaphorical tension, a mix of contrast and similarity, is translated into narrative tension that makes use care about those characters. The character metaphor achieves its goal whether we perceive it as metaphor or not.

7. When You Come To A Fork In The Road, Take It

Research in creativity has tended to romanticize and even fetishize the notion of divergence, to view it as a protean force that pushes us off the path of orthodoxy and onto new avenues of inquiry where unexpected value can be found in the realm of the unconventional or the illogical. As Dostoyevsky (1864) memorably put it, "I admit that twice two makes four is an excellent thing, but if we are to give everything its due, twice two makes five is sometimes a very charming thing too." The concept of divergence is so strongly associated with renowned creators such as Leonardo da Vinci, an artist and inventor whose notebooks document his giddy jumps from topic to topic and from discipline to discipline, that it seems a necessary precursor to genuine creativity. But was Leonardo creative because of a divergent thinking style or did his diversity of interests stem from his creative temperament? It takes a creative mindset to embrace divergence, and a certain creativity to spot the divergent off-ramps that can lead to practical ends. Viewed purely as ideas, our concepts of divergence and creativity are tangled up in a tight circular knot. As computationalists we can easily cut this Gordian knot by striving to introduce divergent possibilities into every computational step of the creative process, or indeed, creative *processes*.

Divergence supports a proliferation of generation and/or packaging strategies wihin the same system. *@MetaphorMagnet*, for instance, employs a wide range of linguistic templates to package simple metaphorical conceits based on a clash of lexical affect between commonly juxtaposed words, or an antonymy of qualities between commonly juxtaposed entities. In each case, it relies on Google n-grams to provide the juxtapositions (such as "love and hate" or "romance and insanity") and on a rich stock of templates to bring the potential metaphors to the fore. Yet even though metaphor is often framed as a mapping between conceptual spaces, systems like *@MetaphorMirror* show that it is neither necessary nor useful to allow those spaces to proliferate at the level of internal representations. Rather than build a separate vector space model for each conceptual space implicated in the mapping, the shared dimensions of a single, jointly-constructed VSM serve as a common language for uniting the diverse elements of each conceptual space. Ironically, then, divergence in creative outputs, of form and of content, is most flexibly attained by working from a unified internal representation.

For text is an especially malleable medium for creativity, whether by humans or machines. So when it comes to the creative generation of metaphors, jokes, poems or any comparable artifacts, the processed outputs are just as malleable as the raw inputs from which they were formed. If guided by the appropriate symbolic and/or statistical models, even a small change at the surface level can yield predictable yet profound changes at the semantic and pragmatic levels. Form-based approaches to computational creativity thus abound, since text is a virtually unlimited raw resource and the web offers us abundant, free-flowing streams of texts for almost any genre or register. As with the work of Stock and his colleagues, the creative systems outlined here thus serve as new sources of value-added content that can be filtered, refined and conveniently repackaged. Just as these new channels may feed the curiosity of human end-users, they may also fuel the generative engines of other systems. Since the web has sufficient capacity to support as many new channels of information as we can provide, we should not presume to fully understand the needs of those who will benefit from the fruits of our own computational divergence. Instead, we should trust in the divergent thinking of others to find the most appropriate and creative uses for our outputs, in contexts and applications we have yet to even dream about.

References

Barnden, J. (2010). Metaphor and metonymy: Making their connections more slippery. *Cognitive linguistics*, 21(1):1–34.

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

Brants, T. and 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.

Davidson, D. (1978). What metaphors mean. Critical Inquiry (Special Issue on Metaphor), 5(5):31–47.

Dostoevsky, F. (1864). *Notes from the Underground: Part I : Chapter IX*. Vintage Classics. New York: Random House.

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

Fauconnier, G. & Turner, M. (1998). Conceptual Integration Networks. *Cognitive Science*, *22*(2), 133–187.

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

Gatti, L., Guerini, M., Stock, O., & Strapparava, C. (2014). Sentiment variations in text for persuasion technology. In *Proceedings of the 9th international conference on persuasive technology* (pp. 106–117).

Gatti, L., Ozbal, G., Guerini, M., Stock, O., & Strapparava, C. (2015). Slogans are not forever: adapting linguistic expressions to the news. In *Proceedings of IJCAI'15, the 24th international conference on artificial intelligence* (p. 2452-2458).

Gatti, L., Ozbal, G., Guerini, M., Stock, O., & Strapparava, C. (2016). Heady-Lines: A creative generator of newspaper headlines. In *Companion publication of the 2016* *international conference on intelligent user interfaces* (p. 79-83).

Gentner, D. (1983). Structure-mapping: A Theoretical Framework. *Cognitive Science,* 7(2):155–170.

Gentzkow, M. & Shapiro, J.M. (2010). What drives media slant? Evidence from the U.S. daily newspapers. *Econometrica*, 78(1):35–7.

Guerini, M., Strapparava, C., & Stock, O. (2011). Slanting existing text with Valentino. In *Proceedings of the 16th international conference on intelligent user interfaces* (pp. 439–440).

Guilford, J. P. (1950). Creativity. *American Psychologist, 5*(9):444–454.

Jordanous, A. (2018). Evaluating evaluation: assessing progress and practices in computational creativity research. In T. Veale & A. Cardoso (Eds.), *Computational creativity: the philosophy and engineering of autonomously creative systems*. Springer International.

Kintsch, W. (2000). Metaphor comprehension: A computational theory. Psychonomic Bulletin Review, 7(2):257-266.

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

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

Lakoff, G. (1987). Women, Fire, and Dangerous Things. University of Chicago Press.

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

Raskin, V. (1985). Semantic Mechanisms of Humor. Dordrecht. Reidel.

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

Steen, G. (2015). Developing, testing and interpreting deliberate metaphor theory. *Journal of Pragmatics*, vol. 90, December 2015, pp 67–72.

Thagard, P. & Stewart, T.C. (2010). The AHA! Experience: Creativity Through Emergent Binding in Neural Networks. *Cognitive Science,* 35(1):1–33.

Utsumi, A. (2007). Interpretive diversity explains metaphor-simile distinction. *Metaphor & Symbol*, 12:291–312.

Utsumi, A. (2011). Computational exploration of metaphor comprehension processes using a semantic space model. *Cognitive Science*, 35(2):251–296.

Veale, T. (2005). Incongruity in humor: Root cause or epiphenomenon? *Humor: International Journal of Humor Research,* 17 (4):419–428.

Veale, T., and Li, G. (2011). Creative Introspection and Knowledge Acquisition. In *Proceedings of the 25th AAAI Conference on Artificial Intelligence.* San Francisco: AAAI Press.

Veale, T. (2013). The Agile Cliché: Using Flexible Stereotypes as Building Blocks in the Construction of an Affective Lexicon. *In: Oltramari, A., Vossen, P., Qin, L. & Hovy, E. (eds). New Trends of Research in Ontologies and Lexical Resources. Berlin: Springer: Theory and Applications of Natural Language Processing*.

Veale, T., Feyaerts, K..J & Forceville, C. J. (2013). E Unis Pluribum: Using Mental Agility to Achieve Creative Duality in Word, Image and Sound. *In Veale, T., Feyaerts, K. and Forceville, C. (eds.), Creativity and the Agile Mind: A Multi-Disciplinary Study of a Multi-Faceted Phenomenon. Mouton de Gruyter, 37–57.*

Veale, T., Shutova, E. & Beigman Klebanov, B. (2016). *Metaphor: A Computational Perspective.* Morgan Claypool Lectures on Human Language Technologies.

Veale, T. (2016). *The shape of tweets to come: Automating language play in social networks.* (Ed.) Bell, N., Multiple Perspectives on Language Play. Mouton De Gruyter, Language Play and Creativity series, pp 73--92.

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

Veale, T. & Valitutti, A. (2017). Tweet dreams are made of this: Appropriate incongruity in the dreamwork of language. *LINGUA 197:141–153.*

Veale, T. (2017). Déjà Vu All Over Again: On the Creative Value of Familiar Elements in the Telling of Original Tales. In Proc. of *ICCC 2017, the 8th International Conference on Computational Creativity, Atlanta, Georgia, June 19-23.*

Veale, T., Chen, H. & Li, G. (2017). I Read The News Today, Oh Boy! Making

Metaphors Topical, Timely and Humorously Personal. In Proc. of *HCII 2017, track on Distributed, Ambient and Pervasive Interactions,* Vancouver, Canada*.*

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