A World With or Without You*

*Terms and Conditions May Apply

Tony Veale, Alessandro Valitutti
School of Computer Science and Informatics, University College Dublin
tony.veale@UCD.ie, alessandro.valitutti@gmail.com

Abstract

We all share the same world, but are free to formulate and argue for our own interpretations of this shared reality. For different agents will grant differing degrees of importance to the same facts and norms. We cannot experiment on human cultures the way scientists experiment on cell cultures, but we can construct thought experiments that imagine the consequences of otherwise impossible changes. Successful thought experiments do not change the world, but change the way we *see* the world. This paper describes *Gedanken*-style reasoning in an AI system that allows a computer to understand, or at least speculate on, the surprising causal interactions between apparently unrelated concepts. This system ponders alternate worlds in which the amount of a conceptual ingredient *[X]* is increased or decreased, to see what unexpected and apparently incongruous effects might arise from this change. Our goal is to construct a creative generator of novel what-if scenarios that can be used in the generation of perspective-shaping stories, poems and jokes.

Changing the Present, Inventing the Future

Science-fiction writers have always been philosophers, of a sort. Like philosophers, such writers explore alternate worlds that, at their root, differ in some small but crucial detail from our own, and pursue the consequences of this change to the limits of logical reasoning. Metaphorically, speculative writing about other worlds and possible futures is akin to throwing a pebble into a lake, to see what ripples spread out from this simple perturbation of familiar reality. A special breed of philosophers – sophists and masters of persuasion – make selective use of the facts to challenge conventional wisdom, to change the way we see the world, and to show how a benign change can lead to an undesirable *reductio ad absurdum,* or how an apparently malign change may sometimes be surprisingly beneficial.

So science fiction stories often use tales about the future to engage with, and change our perspective on, key issues of the present. *Utopian* and *dystopian* tales thus highlight the role of certain core concepts in our present society. These utopias (idealistic futures) and dystopias (dysfunctional futures) follow certain recurring tropes which can be viewed as the seeds of novel sci-fi scenarios. Different scenarios may share the same seed (e.g. a world without love? a world without disease?) but may expand upon these seeds in novel ways. So the key to generating creative perspectives from familiar seeds is the way in which these seeds are elaborated. For example, is there an unexpectedly negative consequence of a world *without disease* (as in over-population and famine)? Perhaps there is an unexpectedly positive consequence of a world *without beauty* (with no ugliness, perhaps there will be less hate?). The classic TV series *Star Trek* found this a fertile furrow for thought-provoking stories, as when e.g. in Season 1, episode 23 (*A Taste of Armageddon*) our heroes encounter a world in which physically *violent* war has been replaced with a *peaceful* simulation of war. The episode concludes that efforts to make war clean and non-disruptive are actually counter-productive and cause even *more* suffering.

In this work we explore the causal consequences of such "*world with more/less* [X]" seeds, to produce thought-provoking perspectives on the world that deserve to be called truly creative. In particular, we will focus on pseudo-logical reasoning over implicit causal structures – *pseudo*-logical because these scenarios, like the literary speculations of science-fiction writers, cannot be "proven" in any strong sense – and call for a sophistry that is not afraid to be selective with the truth. The rest of the paper lays out our efforts to date using the following structure: we first use corpus-analysis techniques to identify the most common *Gedanken* seeds for our conceptual explorations; we then describe how corpus analysis is also used to create a knowledge-base of triples to serve as a representation of our quotidian world that is to be experimentally changed; we next consider how the implicit causality of these triples is projected across inferential pathways to derive surprising consequences of our most familiar beliefs.

Ponds and Ripples

Jared Diamond opens his book "*Why is Sex Fun?*" with a claim that is at once both obvious and thought-provoking:

> "The subject of sex preoccupies us. It's the source of our most intense pleasures. Often it's also the cause of misery, much of which arises from built-in conflicts between the evolved roles of women and men."

People's concepts are like people themselves: peel away the familiar veneer and one finds complexity, conflict and contradiction in abundance. We are preoccupied by a great many concepts like sex, whose outer appearance and affect offer a poor summary of their built-in causal complexity. As in Diamond's book, which examines just one of these grand themes of humanity, we can expose the conflict – to good humorous and philosophical effect – by considering how concepts interact with each other to generate causal ripples that achieve *emergent* effects. We consider the propagation of these causal ripples in the next section; but first, we must consider the structure of the pond itself.

We begin by defining a corpus-based criterion for what constitutes a cultural concept that preoccupies many of us. Ironically, our criterion speculates about the lack of such a concept: if a significant number of people on the Web speculate about the hypothetical world in which an aspect T is not present, then T is deemed to be an interesting and *changeable* topic of frequent preoccupation. As our Web corpus we use the Google n-grams (Brants & Franz, 2006), which is a large database of frequent Web snippets of between 1 and 5 words long. An n-gram such as "a world without sex" (here n=4) will be found in this database if its Web frequency (at the time the database was compiled) is 40 or larger. As it happens, the 3-gram "*world without sex*" has a frequency of 110 in the Google n-grams database.

To build our knowledge-representation of the quotidian world, we collect all the noun values of T for which the 3-gram "*world without [T]*" is found in the Google n-grams. To do this, we use the *Creative Information Retrieval* (CIR) model of Veale (2011), which allows us to find n-grams that match complex non-literal queries. Our sweep of the n-grams database reveals approx. 200 grand themes on which to anchor our representation of the world. The top 20 matching "without" n-grams (and, in parentheses, their frequencies) in the Google database are as follows:

war(7350), poverty(2551), hate(2386), love(1944), violence(1863), boundaries(1772), fear(1702), hunger(1561), information(1399), laws(1339), government(1254), religion(1063), lies(901), sin(830), hurt(821), music(816), money(788), conflict(690), pain(676), light(633), thieves(555), evil(548)

One could theoretically speculate about a world without any arbitrary concept or thing, such as a world without paper clips or toilet paper. One might even be able to construct an entertaining and well-argued narrative about how this simple lack ultimately causes a deep tear in the causal fabric of the world. But since the n-grams "world without paperclips" and "world without toilet paper" reside on the long-tail of the Web's content, and not in the rump captured by the Google n-grams, we do not consider these concepts to be core to our representation of the world.

With our stock of 200 or so core concepts, we now set about constructing a knowledge-base to connect them all. We use CIR to guide the way, by retrieving all Google 3-grams that match the pattern "{T}and "{T}", where {T} is our set of core concepts retrieved earlier. Each 3-gram of this kind represents a yoking of two topics in the popular imagination: these topics belong together, or at least deserve to be linked together in our world representation. Consider Sex: the Google 3-grams show that this topic is often coordinated with *drugs (97568), violence(65432), romance(50134), love(21334)* and *fun(16011)*. It is also coordinated with *crime(1318)*, *pain(1157)* and *sin(1034)*. If the set {T} provides the vertices of a knowledge-graph, these coordinations provide the unlabeled edges, and so we quickly move from a set to a graph representation.

Causal ripples will propagate across these edges as we speculate about hypothetical worlds with more or less amounts of a certain concept (less pain, no money, more love, no sin, less crime, etc.). However, as different edges will represent different kinds of relationships, and thus affect causality in different ways, we need to label these edges carefully. For the best quality knowledge, we do this labeling manually. For every edge A—B in the graph, we convert this edge into one or more labeled, directed arcs A→*rel*→B. As this is an impractical task to perform manually with a small number of annotators, we restrict ourselves to labeling the edges connecting one of the top 25 elements of {T} by n-gram frequency. As these topics will typically connect to less frequent elements of {T}, we achieve good coverage of the graph by focusing mainly on these *crossroad* concepts. In all, we label approx. 4000 of the edges in our knowledge-graph, to produce a sizable knowledge-base of over 6,000 semantic triples. Work is afoot to enlarge this knowledge-base considerably, but such a size is more than adequate for our pilot exploration.

These triples draw on an open inventory of semantic relationships (unlike a representation such as *ConceptNet*). For instance, the relationship between critics and artists is captured with the label *criticize*, while that between artists and their art is captured with the labels *produce* and *sell*. No attempt is made to capture the semantics of labels like *produce*, *sell* and *criticize* in any formal axiomatic sense, though as we shall see next, we broadly categorize labels according to a very simple model of causal reasoning. This allows a system to reason about the overall ramifications of a Gedanken change that produces a new speculative world.

Sophistry & Surprise on the path less traveled

When considering the broad implications of any triple of the form A→*rel*→B, we pose ourselves these questions: does this triple imply that a world with more A will likely have more B (*positive* causality), so that a world with less A will likely have less B? Or does this triple suggest that a world with more A will likely have less B (*negative* causality), so that a world with less A may have more B? Or is the nature of *rel* such that neither of these outcomes seems reasonable (*neutral* causality)? We then categorize each relationship *rel* that does not fall into the neutral class as either + (if positive causality) or - (if negative causality). Note the use of "suggest" and "likely" here: we care not for safe inference, but value instead the folk inferences that underpin gut feelings, intuitions and speculative fictions.

The triple format is not especially expressive, but triples can be chained together to form complex inferential paths. Thus, the triples A→*r1*→B, B→*r2*→C and C→*r3*→D can be chained to yield the path A→*r1*→B→*r2*→C→*r3*→D. Simple causal propagation rules can be defined to reason about the effect of the head of a pathway (e.g. A) on the end of a pathway (e.g. D). For instance, if *r1* and *r2* have positive causality and *r3* has negative causality, a system can reason that more A causes more B with causes more C which causes less D, so more A causes less D. Though our representation of the world does not directly link A to D, a system can broadly infer a causal consequence of A on D.

Causal pathways can be constructed from a knowledge-graph using simple processes of spreading activation and marker passing. Many of the pathways eked out in this way will be trite and uninteresting, but some will be surprising, and may even seem humorously incongruous to an average person when presented in the collapsed form *A→..→D*. These are the pathways that interest us here, the pathways that effectively create a jarring (but resolvable) *bisociation* – in the sense of Koestler (1964) – between two very different concepts A and D. These are the pathways beloved of science-fiction writers, jokers and sophists, the pathways that use what we already know to surprise us.

Sophistry is a natural by-product of this approach rather than an engineered goal, due in large part to our choice of representation. Notice how our concepts are denoted by linguistic labels such as *love*, *war* and *criticize*. As no attempt is made to sense-tag these vertices and arcs relative to a sense inventory like WordNet – since such an effort would be prohibitively expensive – labels are often used in multiple different senses simultaneously. Philosophers refer to this as the fallacy of *equivocation*, and consider equivocal arguments to be faulty arguments. For our part, we view equivocation as a locus of conceptual blending (Fauconnier & Turner, 2002; Veale & O'Donoghue, 2000), insofar as a label used with multiple meanings is deemed to denote a conceptual blend of all these meanings.

Consider this chained pathway of three triples: critics→*criticize*→artists→*produce*→art. The label *critics* is obviously (but implicitly) used here to denote art critics. Now consider this triple dictators→*suppress*→critics. The label *critics* is obviously (but again, implicitly) used here to denote political critics of a regime. Since the labels are undifferentiated by sense-tags, the system can combine both paths at the nexus *critics*, to yield: dictators→*suppress*→critics→*criticize*→artists→*produce*→art. That is, dictators suppress the critics that criticize the artists that produce art. Real dictators do not do this (if anything, they do the opposite, even in the case of dictators that were once artists themselves). What we see here is a blend of the two senses of *critic*, where art critics become political critics. We also see another implied blend: artists that suppress their critics, or try to, will effectively become dictators. Chaining triples into long inference paths will often create interesting blends at the points where paths link together.

Good pathways make for good stories, but what makes a good pathway, from a computational perspective at least? We employ a simple but effective criterion here, one that will be nuanced and elaborated in future work. An interesting pathway is one that links a concept A to another concept D by coherently chaining multiple triples together, where there is a bisociative tension between worlds with more A and worlds with more D. We expect a positive concept (such as *love*, *beauty*, *romance*, *art*, etc.) to have positive consequences on the world, by which we mean the proliferation of other positive concepts and the diminution of negative concepts. Likewise, we expect negative concepts (like *war*, *hate*, *jealousy*, *pain*, etc.) to have negative consequences on the word, and to diminish the effect of positive concepts. So a path *A→..→D* that shows how a positive concept A can have a positive causal effect on a negative concept D (so more A means more D), or that shows how a negative concept A can have a positive causal effect on a positive concept D (so less A means less D), will be considered interesting for our purposes.

For this criterion to work, we simply tag each element T in our inventory of core concepts {T} with a +/- sentiment label to indicate, on a gross level, whether it is affectively positive or negative. We then allow the causal path-finder to explore the space of possible triple chains to find these interesting pathways. Pathways that show a desirable, positive concept to have surprising negative consequences, or a negative concept to have surprising positive consequences, add nuances of emotion to these gross sentiment classes. In effect they show a topic T to reside in two mutually incongruous frames of reference at once – that which is desirable and that which is undesirable – and conform to how Koestler (1964) defined a bisociation: *"the perceiving of a situation or idea in two self-consistent but habitually incompatible frames of reference"*.

Reasoning about Inferential Surprise

We have described here a simple, triple-based knowledge-representation of those aspects of the world that preoccupy us the most, and shown how these triples can be chained together to form complex, and often surprising, inferential pathways. Additionally, we have shown how a system can reason about these pathways at a higher level by generalizing over their causal implications. By assigning coarse +/- sentiment classes to individual concepts, and coarse +/- causal classes to individual relation types, a system can infer the broad causal effect of the concept at the head of a path on the concept at the tail of a path. We have hypothesized that a path is more likely to be viewed as surprising if there is an apparent incongruity between the head and the tail of a path.

Suppose we have a path $S \rightarrow .. \rightarrow T$ that causally connects a start concept S to a target concept T via two or more chained triples. If S is a positive concept (sentiment-wise) we can denote it +S, and if it is a negative concept we denote it –S. Likewise we can denote T as +T or –T depending on its coarse-grained sentiment. Sentiment marking thus gives us four possibilities for our pathway:

i. $+S \rightarrow .. \rightarrow +T$
ii. $+S \rightarrow .. \rightarrow -T$
iii. $-S \rightarrow .. \rightarrow +T$
iv. $-S \rightarrow .. \rightarrow -T$

Cases (ii) and (iv) above correspond to inference paths that causally link two concepts of opposing +/- polarity. However, these paths need not be surprising or apparently incongruous, as an intervening negative relation is likely inverting the polarity. Thus it is hardly surprising that *policeman* (+) can be linked to *crime* (-) via the pathway policeman→*arrest*→criminals→*commit*→crimes. Any incongruity here, such as it is, resides entirely on the surface, between the superficial sentiment of the pathway's end-points rather than between its deeper cause and effects. Causally, this path can be summarized as follows: *more*(policeman)→*cause*→…→*cause*→*less*(crime). Since *crime* is a negative concept, *less(crime)* must be a positive concept, so when the causal implications of this path are considered, it actually links a positive to a positive.

Viewed this way, where *more*(+X) is a positive concept and *less*(+X) is a negative while *more*(-X) is a negative and *less*(-X) is a positive, the following generalizations apply:

v. *more(+S)→...cause...→more(+T)*
vi. *more(+S)→...cause...→more(-T)*
vii. *more(-S)→...cause...→more(+T)*
viii. *more(-S)→...cause...→more(-T)*
ix. *more(+S)→...cause...→less(+T)*
x. *more(+S)→...cause...→less(-T)*
xi. *more(-S)→...cause...→less(+T)*
xii. *more(-S)→...cause...→less(-T)*

We hypothesize that the generalizations that best capture the notion of a surprising inference pathway are (vi), (vii), (ix) and (xii). Pathways of type (vi) show how a world with even more of a positive quality must also have more of a negative quality, or less of another positive (ix). Pathways of type (vii) and (xii) conversely show that negative concepts can have positive effects, either directly (vii) or by diminishing the presence of another negative (xii). The pathway dictators→ *suppress*→critics→*criticize*→artists→ *produce*→art thus conforms to pattern (vii), as it implies that a world with more dictators may be one with more art.

Empirical Evaluation

We thus predict that, *ceteris paribus*, test subjects will find inference pathways that conform to the general patterns of (vi), (vii), (ix) and (xii) more surprising than those that conform to the patterns (ii) and (iv). For experimental purposes, we denote the former class of paths (patterns vi, vii, ix and xii) as *CausalSurprise* paths. We denote the latter class (conforming to ii and iv) as *SurfaceSurprise* paths. Pathways conforming to all the remaining patterns (i, iii, v, viii, x and xi) are denoted as *NoSurprise* paths.

SurfaceSurprise, *CausalSurprise* and *NoSurprise* paths provide the three test conditions under which we evaluate our hypothesis. A candidate pool of 80 inferential paths was randomly generated for each condition, 30 of which (per condition) were manually annotated as a gold standard to detect scammers, and 50 of which (per condition) were finally annotated by independent judges. The crowd-sourcing platform *CrowdFlower*[2] was used to recruit a panel of 70 human judges to estimate, for each of these 3x50 pathways, the degree of surprise exhibited by each. The full inferential pathway was presented in each case, so that judges could see not only its conceptual end-points, but the specific relationships and causal logic at work in each path. The gold standard paths were used to detect unengaged scammers, resulting in 2.5% of judgments overall being discarded. Ultimately, 50 pathways for each condition were judged by 15 or more judges, producing 765 judgments for the *NoSurprise* condition, 750 for the *CausalSurprise* condition, and 751 for *SurfaceSurprise*. Each judgment provided a measure of surprise on a scale from 0 (no surprise) to 3 (very surprising) for a given path.

The mean surprise value for the 765 judgments in the *NoSurprise* condition is **1.06**, that for the 751 judgments of the *SurfaceSurprise* condition is **0.96**, and that for the 750 judgments for the *CausalSurprise* condition is **1.44**. There is very little here to distinguish *Surface Surprise* from *No Surprise* at all, suggesting that the *Surface Surprise* paths are just not very interesting. Yet this is to be expected, for though these paths link concepts of conflicting sentiment

[2] *www.CrowdFlower.com*

(such as *justice* & *crime*), the causal reasoning that links the head and tail is explicitly articulated step-by-step.

However, as hypothesized earlier, there is a statistically significant difference in surprisingness between, on the one hand, the *SurfaceSurprise* and *NoSurprise* conditions, and on the other, the *CausalSurprise* condition. A one-sided Wilcoxon rank-sum test shows that the increase in mean surprisingness from the *SurfaceSurprise* and *NoSurprise* conditions to the *CausalSurprise* condition is significant at the $p < .001$ level. When seeking to eke out interesting inferences from a common-sense knowledge-base, patterns (vi), (vii), (ix) & (xii) yield the most surprising pathways by introducing a provocative causal *twist*.

Conclusions: Bots With Attitude

Surprising inferential pathways can be used to generate surprising linguistic artifacts, such as stories that take unexpected plot twists (an artist who becomes a dictator? a televangelist that becomes a jihadist?) or metaphors that choose provocative vehicles to describe a given target idea. Consider the latter application: pathways that end at the same destination point, and which involve much the same causal trajectory – as captured by patterns (vi), (vii), (ix) and (xii) – are causally similar, and may even be treated as morally equivalent. Thus, an artist who seeks to suppress his or her critics is no better than a dictator, while a devout televangelist who seeks to promote his faith by stirring up hatred for unbelievers is no better than a jihadi terrorist. Metaphors that make these comparisons and highlight these causal similarities may do more than surprise: they may actually change the way we view a target concept, or at least generate debate as to how it should be viewed.

The knowledge-based techniques described here allow a bot to be programmed to generate creative metaphors in this mold. The ***@MetaphorMagnet*** twitterbot is a fully automated system – described in Veale (2014a,b) – that generates such metaphors every hour, on the hour. Readers can judge for themselves just how thought-provoking its outputs are, while the pattern of re-tweeting will eventually provide us with further empirical evidence as to which causal patterns are most favored by social media mavens.

Acknowledgements

This research was supported by the EC project *WHIM: The What-If Machine*. *http://www.whim-project.eu/*

References

Brants, Y. and Franz, A. 2006. Web 1T 5-gram Version 1. *Linguistic Data Consortium*.

Diamond, J. 1997. *Why is Sex Fun? The Evolution of Human Sexuality*. Basic Books.

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

Koestler, A. 1964. *The Act of Creation*. Hutchinsons, London.

Veale, T. and D. O'Donoghue. 2000. Computation and Blending. *Cognitive Linguistics*, 11(3-4):253-281.

Veale, T. 2011. Creative Language Retrieval: A Robust Hybrid of Information Retrieval and Linguistic Creativity. *Proceedings of ACL'2011, the 49th Annual Meeting of the Association of Computational Linguistics, Portland, Oregon.*

Veale, T. 2014a. A Service-Oriented Architecture for Metaphor Processing. *Proceedings of the Second Workshop on Metaphor in NLP, at ACL 2014, the 52nd Annual Meeting of the Association for Computational Linguistics, Baltimore, June 2014.*

Veale, T. 2014b. Coming Good and Breaking Bad: Generating Transformative Character Arcs For Use in Compelling Stories. *Proceedings of ICCC-2014, the 5th International Conference on Computational Creativity, Ljubljana, June 2014*.