RESEARCH　　　　　　　　　　　　　　　　　　　　　　　　　　　**Open Access**

Language use in the Jihadist magazines inspire and Azan

David B Skillicorn[1*] and Edna F Reid[2]

Abstract

The language of influence or propaganda has been studied for a century but its predictions (simplification, deceptiveness, manipulation) can now be examined empirically using corpus analytics. Semantic models for intensity of belief and use of gamification as a strategy allow novel aspects of influence to be taken into account as well. We develop a semi-automated approach to assess the quality of the language of influence using semantic models, and singular value decomposition as a middle ground between high-level abstract analysis and simple word counting. We then apply this approach in a significant intelligence application: examining the use of the language of influence in the jihadist magazines *Inspire* and *Azan*. These magazines have attracted attention from intelligence organizations because of their avowed goal of motivating lone-wolf attacks in Western countries. Our approach enables us to address questions like: How good are the authors and editors of these magazines at producing influential language (and so how great is the impact of these magazines likely to be)? How does this change with time, and as a reaction to world events, and what does this tell us about competence and strategic goals? What is the impact of changes in authorship?

Keywords: Language model; Jihadist language; Integrative complexity; Gamification; Deception; Propaganda; Corpus analytics

Introduction

Inspire magazine is an online jihadist magazine, written in English. At the time of writing, twelve issues have appeared, the first in the middle of 2010 and the most recent in Spring 2014. The first eight issues were edited by Samir Khan, with substantial assistance from the charismatic preacher, Anwar al-Awlaki. Issue 8 did not appear until well after their deaths in 2011, but had clearly been written by them. Issues 9–12 are the product of different, so far unknown, editors.

The goal of *Inspire* appears to be to motivate lone-wolf attacks in Western countries by diaspora jihadists. Such potential attackers are assumed to have little Arabic, and so the magazine is written in English; to be familiar with high-quality publications, and so the magazine has high production values, including significant attention paid to layout and presentation; and to have had limited exposure to religion and particularly the religious language endemic in much of the Middle East. *Inspire* is thus much more "Westernized" than the typical discussions in, for example, Islamist forums. Intelligence agencies have taken *Inspire*'s role as a recruiting tool very seriously [1]. Indeed, it is against the law to possess copies in the United Kingdom, and several people have been convicted for this offence; an Australian man has also been charged with a similar offence.

Azan magazine resembles *Inspire* in "look and feel" and appears to share many of the design elements, so it seems reasonable to assume that it was modelled on *Inspire*. At the time of writing, five issues have appeared, the first in March 2013 and the fifth early in 2014. Its focus is more on a South Asian (English-speaking) readership, but its goal is apparently also to encourage self-initiated attacks in Western countries. The magazine's authorship is not known but the topics covered suggest concerns centered in Pakistan and Afghanistan, particularly a focus on the threat from drones. The textual content of issues of *Azan* is typically between three and four times longer than the textual content of issues of *Inspire*, and issues are appearing much more often so their author/editor is noticeably more productive.

*Correspondence: skill@cs.queensu.ca
[1]Queen's University, Kingston, Canada
Full list of author information is available at the end of the article

The effect of these magazines is difficult to judge. Several lone-wolf attackers have had copies in their possession; but it is also clear that the magazines are widely read in their target communities. There has not been a noticeable uptick in lone-wolf attacks since they began appearing, so it is difficult to conclude that they have had much effect. We investigate whether magazines such as these meet their goals or not. We focus on the language patterns used in these magazines and what they can tell us about the aims and perceptions of the authors, and the constraints on their success.

The contribution of this paper is threefold:

1. A mapping of high-level descriptions of the language of influence or propaganda into hypotheses about the ways in which particular, measurable language elements are expected to be used in documents intended to influence.
2. A computational methodology for instantiating models of particular kinds of language in terms of extracted word frequencies and variational analysis using singular value decomposition. This enables model intensity of different documents to be compared, especially for each of the magazines across time, and hypotheses about expected language use to be validated or falsified.
3. This, in turn, allows questions about the potential success of these jihadist magazines to be answered: how well they are able to deploy influence (how dangerous these magazines are), and how they react to world events and their own successes and failures (how professional or strategic they are). We are also able to determine authorship of individual issues.

We conclude that both magazines deploy the language of influence fairly well, although *Inspire* seems to have difficulty staying "on message". More worryingly, the intensity of jihadist language is rapidly increasing in *Inspire*.

We assume that intelligence analysts are assisted by automated or semi-automated analytic methodologies to carry out tasks such as extracting text and applying various language models to it, a task we carry out by hand. We make no special effort to handle issues that arise from the conversion process from Portable Document Format to text, or those that arise from automated part-of-speech labelling and word frequency extraction, on the assumption that a scalable version of this methodology should be robust in the presence of these issues.

Background

There is a long history of modelling the language of influence or, more pejoratively, propaganda, and these analyses provide indications of what to expect of the language patterns of documents whose primary purpose is influence. Martin [2] defines propaganda this way: "an organized deliberate attempt to influence many people, explicitly or implicitly". It is clear that *Inspire* and *Azan* meet these criteria. The essence of propaganda is that it is intended to be opinion forming, and the essence of its mechanism is that it is manipulative. Martin suggests that propaganda is characterized by: simplification, attractive visuals, deceptiveness of some kind, and psychological techniques that weaken or bypass rationality. Work in corpus analytics has made it possible to instantiate these characterizations in terms of particular sets of words and how they are used in documents and corpora.

We apply models with a semantic component: Rayson's models of informative language and imaginative language [3], Suedfeld's model of integrative complexity [4], Pennebaker's model of deception [5], Koppel's model of jihadist language [6], and a model for gamification developed by the authors. Gamification is being used as a strategy for motivation in business settings, but it has only just begun to be considered as a potential strategy for influence or propaganda.

Previous analysis of the approach and content of *Inspire* has focused on four questions: how does the content address the avowed purpose; what mechanisms are used to increase the "stickiness" of the message; what choice model for the readers is assumed; and are there contradictions inherent in the goal of the magazine? *Azan* has not yet received much academic attention, but the same four questions invite answers.

There are three audiences who appear to be addressed, at least implicitly, by these magazines. The most obvious and important are those sympathetic to jihadist ideology who are capable of being convinced to carry out attacks. Content for this audience consists of coercion towards carrying out attacks, and information about targeting and methodologies to support this. The second audience is the jihadist community itself who need to be shown that steps are being taken, who can point to the magazine as something to be proud of, and who can use it as a tool in their own recruiting. Content for this audience might be framed as morale building. The third audience is the population of Western countries at large. Content for this audience is intended to create terror by suggesting the presence of large numbers of motivated lone-wolf extremists in their midst, hard to identify and ready to strike anywhere without detectable precursors. A detailed analysis of the content of *Inspire* (up to Issue 7) was carried out by Ford [7]. She argues that the messaging in *Inspire* is consistent with themes in other Al Qaeda communication, emphasizing: a clash of civilizations, Muslim unity, the necessity of (violent) jihad, religious justification, Al Qaeda's

superiority, and glorification of martyrdom and hero worship. Interestingly, issues of *Azan* have also (somewhat plaintively) asked for help from a fourth audience, anyone with knowledge of drone operations to help them develop countermeasures.

The second aspect is how the authors attempt to make their messages "sticky", that is both standing out in the noisy media environment of the West, and compelling to the intended audiences. One major part of this is the use of narrative, both narratives of past successes, and biographies of previous heroes and martyrs of the movement. This is intended to create identification, to make abstract possibilities seem plausible and realistic, and perhaps even to create envy. Another part is the high production values of the magazine itself, making it more readable and serious-looking for Western eyes, used to the production quality of mass-market magazines. Considerable attention is paid to pictures and layout.

The contrast between *Inspire* and typical Islamist discussion forums, which also attempt to motivate attackers and attacks, is striking. There has also been research suggesting that crowdsourcing and gamification are being used in *Inspire* as a motivational strategy and to desensitize readers to the human consequences of violent attacks [8,9].

The third aspect is the choice model that is assumed by the writers and editors, that is what is supposed to go through the mind of a reader on the path towards willingness for violent action [10]. A rational choice model is implausible since the intention is to influence individuals towards actions that will result in their deaths or long terms of imprisonment. Two choice models have been suggested. The first is bounded rational choice, in which the available choices are made to appear constrained, either by controlling the available information, or the amount of resource available to make a decision. In settings where the consequences to the individual of taking action are severe, for example in the military, appeals to honor and community are often made, and these magazines use the same strategy. The second choice model is situational action theory, which models the process as the infection of vulnerable individuals. There is some empirical evidence to support this – it explains the *rates* of lone-wolf attacks well – but it does not have much explanatory power as a way of predicting *which* individuals may become infected [11]. Overall, those who are convinced to act in a violent way do not themselves enjoy the results of their actions. Accordingly, those who want to convince them must feel pressure to express the outcomes in abstract terms, and the goals in terms of what they are against, rather than what they are for.

The fourth aspect is the inherent contradiction in producing documents intended to convince *others* to carry out violent acts at a cost to themselves, an argument that has been made in detail by Ramsay [10]. If it is acceptable to contribute to jihad by writing for and producing magazines, then it must also acceptable for magazine readers to contribute to jihad in less-lethal ways, perhaps recruitment and fund raising. This may go some way to explaining why the effects of *Inspire* and *Azan* have apparently been so modest.

Lemieux *et al.* [12] have described *Inspire* from the perspective of the Information, Motivation, and Behavior (IMB) framework. This framework suggests that causing others to act requires providing them with information, providing them with motivation, and teaching them behavioral skills that will be required for the desired actions. They show that the issues match this framework quite well – for example, the most famous article in *Inspire*, "Build a bomb in the kitchen of your Mom" is a recipe book for building a certain kind of IED. The IMB framework may be used intuitively, but it is also possible that Khan and al-Awlaki had some informal psychological background and deployed this strategy consciously.

The theory of the language of influence or propaganda, and the specific theories that have been raised for magazines such as *Inspire* and *Azan*, suggest several hypotheses about the language patterns to be seen in these magazines.

- Informative language intensity will be high as the magazines try to train their readers in both targeting and techniques. (This is slightly unexpected: conventional theories of propaganda tend to emphasize transmission of ideas, rather than of actionable knowledge about, for example, bomb making).
- Imaginative language intensity will be high to increase emotional drivers of motivation, and perhaps as a way to bypass rationality.
- Integrative complexity will be consistent with other kinds of Western magazines – the goal is not to provide a textbook; but neither is it to become too overtly simplistic.
- Deceptive language intensity will be high because of the inherent difficulty of convincing others to carry out actions that are not in their individual interest, and because influence tends to become manipulative.
- Jihadist language intensity will be high, reflecting the mindset of the authors.
- Gamification language intensity will be non-negligible as as a way to drive action via rewards from the process itself, and to provide emotional distance between acts and consequences.

We investigate empirically whether these hypotheses are supported. The temporal variation in intensity of these

semantic models is of particular interest for two reasons: it provides us with insight about how author/editor strategy changes in relation to what happens in the outside world; and it allows us to gauge whether, for example, the danger posed by these magazines is increasing or decreasing.

Empirical evidence from Islamist forums has shown that high levels of jihadist language tends to be associated with low levels of deception [13,14]. The validation of the hypotheses relating to jihadist and deceptive language intensity will shed light on whether these magazines resemble forum language or not.

Since authorship is fairly reliably detectable from text, we can also investigate whether magazine issues are written by the same author(s). We expect the first 8 issues of *Inspire* to display consistent language; and we can address questions of authorship for the remaining issues of *Inspire* and all of the issues of *Azan*.

Approach

A language model measures some high-level property of the documents in a corpus by counting the occurrence frequencies of words that are associated with that property. At its simplest, the total number of occurrences of a given set of words can be computed for each document and treated as a score – for example, counting the number of adjectives in a document to measure its 'colorfulness'.

We compute a number of such measures for the magazine issues in what follows. Given a set of relevant words, either explicitly as a list, or by capturing all occurrences of a given part-of-speech, we compute a document-word matrix. Each row corresponds to a magazine issue, each column corresponds to a word of the particular language model, and the entries describe the frequency of each such word in each document.

The entries in each row of such a matrix could simply be summed to approximate the intensity of the language in each document. However, frequencies are necessarily larger in longer documents, so the rows must at least be normalized by dividing by the total number of word occurrences (that is, the length of the document), converting word frequencies into word rates.

In most language models, individual words are not equally significant markers of the property being considered, but summing across rows implicitly assumes that they are. Instead, we refine the analysis of a language model by transforming the space spanned by the words into a lower dimensional one, and embedding the (rows corresponding to the) documents as points in this space. If the language model describes only a single-factor property, this embeds the documents in a 1-dimensional space with documents that have the greatest intensity of the language model at one end, and those with the weakest

intensity at the other. In other words, the embedding creates a ranking of documents by the intensity of the property. Even when the language model consists of more than one underlying factor, a ranking can be constructed by projecting onto a a vector that captures variation along more than one axis.

A representation of the documents (and the words) in a 2- or 3-dimensional space allows the results to be visualized, and similarities and differences exposed. When there is a time sequence to the set of documents, as there is for the issues of these two magazines, we can also glean information from the trajectory over time of the intensity of each kind of language.

We use a singular value decomposition of the document-word matrix to carry out this embedding. Given an $n \times m$ document-word matrix, A, the singular value decomposition [15] expresses it as a product

$$A = USV'$$

where k is the minimum of n and m, U is $n \times k$, S is an $k \times k$ diagonal matrix whose non-increasing entries indicate the amount of variation in each dimension of the embedding, and V' is $k \times m$. The superscript dash indicates matrix transposition. U describes the coordinates of each document with respect to axes defined by V and, simultaneously and symmetrically, V describes the coordinates of each word with respect to axes defined by U. The decomposition can be truncated at values smaller than k, say 2 or 3, with minimal loss of structure, and then plotted so that structures can be visualized.

For the semantic language models, we create two artificial documents to represent both strong and weak language patterns. Before computing the SVD, word frequencies are converted to z-scores. Therefore we create an artificial document (matrix row) all of whose entries are $+1$ and another all of whose entries are -1. These represent documents that appear as if they use every word of the language model at one standard deviation above (resp. below) the mean usage rate in each corpus. These documents do not participate in the computation of the SVD, but are inserted into the resulting space by postmultiplying them (as row vectors) by $V * S^{-1/2}$. The resulting positions approximate where a document with intense use of the language model being considered would lie (indicated by an "H"), and where one with weak use of the language model would lie (indicated by an "L"). This is useful in two ways. First, the line joining "L" to "H" indicates the direction in which the language intensity is increasing, and so orients the roles of the other points (documents). Second, it allows us to claim that levels of particular language are high when many of the journal issues fall near or above the "H" level, since this means that many words of

the corresponding model must occur more than one standard deviation above the mean, *even within such a small corpus.*

Document-word matrices were extracted from versions of the magazines converted from Portable Document Format to plain text. This conversion, especially for such visually rich documents, introduces a number of artifacts. We remove only the most obvious of these, since a typical intelligence analysis pipeline would typically not have the resources to clean the text files extensively. Enough investigation was carried out to suggest that the remaining artifacts have negligible effects on the results.

Results

Three issues of *Inspire* are of particular interest in what follows:

- Issue 3 – bombs aboard cargo planes, claim of bringing down a UPS plane in Dubai that is widely believed to be untrue;
- Issue 5 – Arab spring;
- Issue 7 – 9/11 anniversary.

The results presented in this section for each semantic model are plots with one point corresponding to each document, placed at coordinates derived from the first three columns of the U matrix. These three columns define a three-dimensional space in which the cloud of points is centered around the origin (because of the normalization to z-scores). Its axes are linear combinations of all of the word frequencies; these are not directly interpretable and so we omit them from the images, and the space is isotropic so that distances in all directions are equally meaningful. This space captures the variation among documents derived from the variations in the way that they use the set of words of each model. Distance from the origin is a surrogate for significant variation, and points in the same direction from the center use the set of words with similar frequencies. Thus closeness corresponds to semantic similarity. For issues of *Azan* and for the first 8 issues of *Inspire* which were written by the same authors, we connect the points corresponding to issues in time order so that variation with time (trends) can be seen.

Informative language

Rayson's informative language model [3] represents how a document succeeds in conveying ideas in a rational fashion; in our context, how well a document represents a call to action based on convincing arguments. It is based on the rate of use of nouns, adjectives, prepositions, and conjunctions.

Figure 1 shows the informative language model applied to the issues of *Inspire*. Issues 1 to 12 are connected by a line in time order, while variation is shown in two dimensions. Thus we can see that most of the issues are quite similar to one another, but Issues 3 and 7 vary substantially

Figure 1 Patterns in the use of informative words for *Inspire*. In these, and subsequent figures, the points labelled "H" and "L" represent artificial documents with levels of the language pattern well above and well below the mean rate in the corpus.
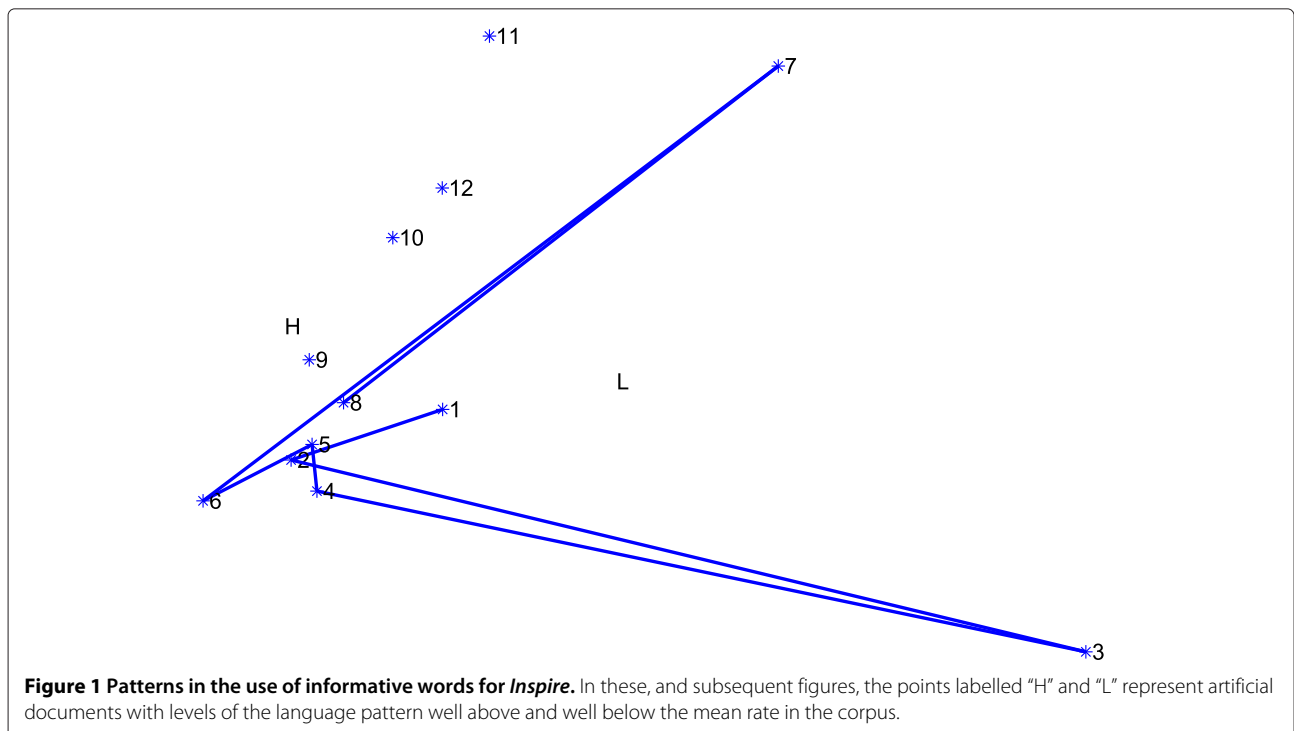
from these typical issues, and from one another. The issues by new authors also show a trend with time, and one that differs from the earlier trends.

The points marked as high ("H") and low ("L") intensity are the inserted artificial documents – documents that use every word of the model at one standard deviation above or below the mean rate. The line between these points defines the gradient from lower to higher intensity and so shows that informative language intensity increases from right to left. Issues 3 and 7, which we earlier commented were triumphalist, are also least informative, which seems consistent.

Figure 2 shows the issues of *Azan*, with most informative at the left of the figure and least informative at the right.

Figure 3 shows the combined plot, with informative intensity increasing from right to left. *Azan* clearly ranks higher overall on informativeness.

Imaginative language

Rayson's imaginative language model, in contrast, tries to convince by appealing to the imagination [3]. It is based on the use of verbs, adverbs, and pronouns. Figure 4 shows the structure of *Inspire* issues based on imaginative language, with intensity increasing from right to left. From this and Figure 1, we see that Issues 3 and 7 are both the least informative and the least imaginative. Note also that the recent issues exhibit a different, orthogonal pattern of informative language: similar in intensity to the other issues, but different in form.

Figure 5 shows the structure for *Azan* with imaginative intensity also increasing from right to left. Figure 6 shows the comparative imaginative intensity across both magazines. *Azan* is slightly more imaginative than *Inspire* across the board.

These results show that informative and imaginative language are not necessarily antithetical – it is possible for a document to be high in both, in just one, or in neither.

Integrative complexity

The integrative complexity model [4,16] measures the intellectual complexity of the content of a document. It can be further subdivided into two aspects: dialectic complexity, the awareness that concepts can be considered from more than one point of view, and elaboration, the awareness that concepts can contain more than one simple idea [17]. Integrative complexity is measured on a 7 point scale; its usefulness has been considerably advanced by the development of an automatic scoring algorithm which we use here [18].

The integrative complexities, shown in Figure 7 show a steady increase with time for *Inspire*, except for the anomalous Issues 10 and 12. *Azan*, in contrast, has an almost constant integrative complexity. Almost exactly the same patterns are visible for dialectic and elaborative complexities (Figures 8 and 9), indicating that the arguments being made include *both* an awareness of contrasting points of view *and* of depth within any single point of view.

It is interesting to consider how these values compare with values for other kinds of writing. Table 1 shows comparable values for randomly chosen articles in mainstream magazines. The values for *Inspire* and *Azan*'s integrative complexity and dialectical complexity are in the mid range compared to these magazines;
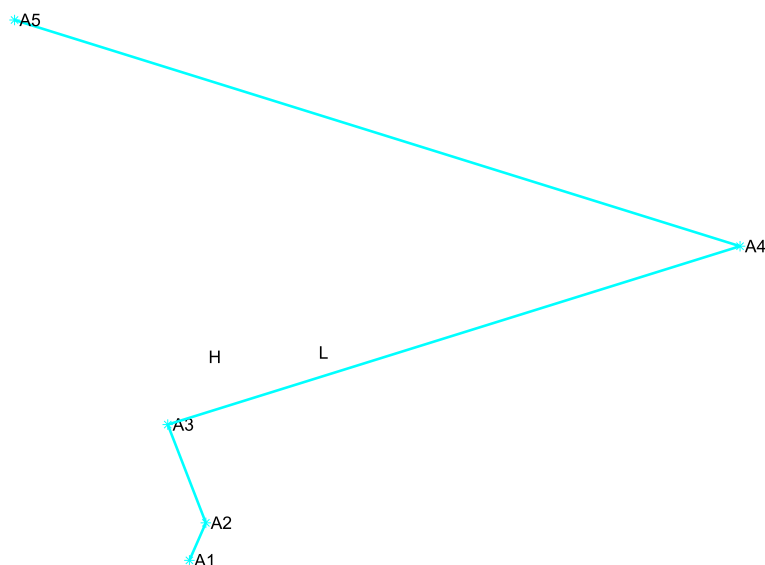

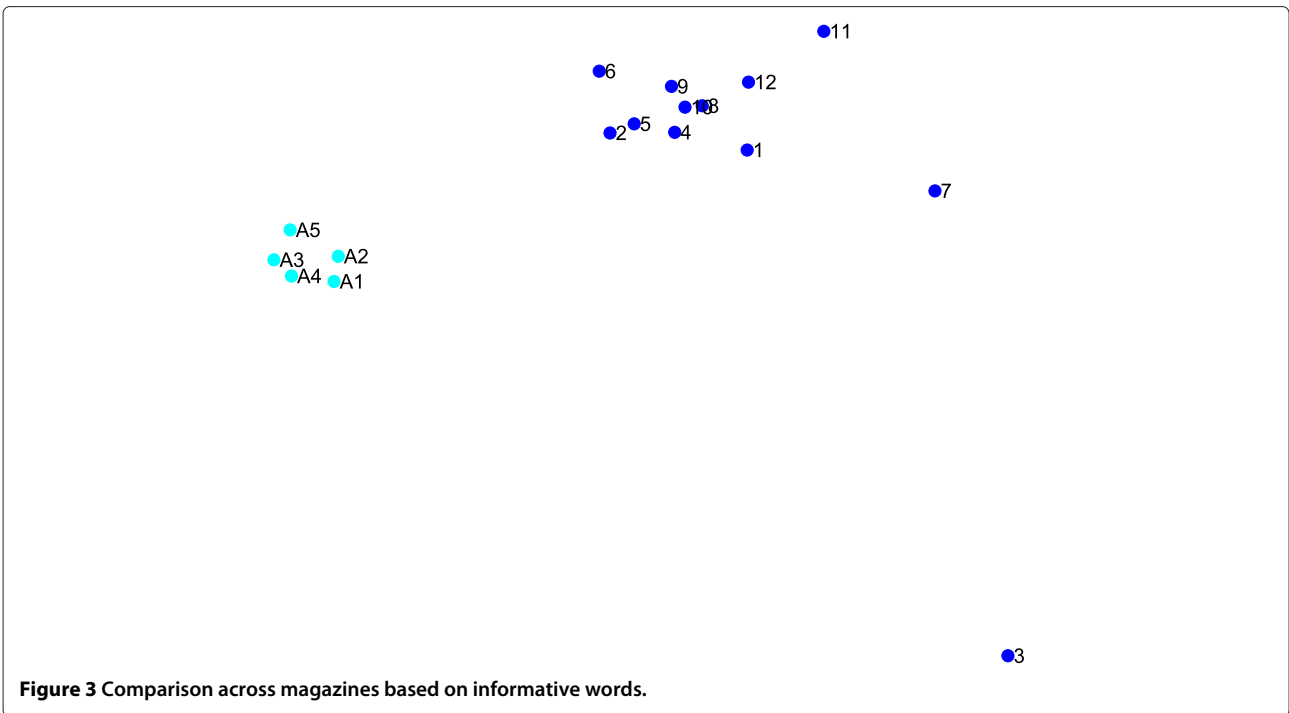**Figure 2 Patterns in the use of informative words for *Azan*.**

Figure 3 Comparison across magazines based on informative words.

their elaborative complexity scores are perhaps slightly lower.

Deceptive language

Deception is a broad spectrum property that includes both outright factual falsehoods and attempts to deceive by creating an impression that is knowingly not true

to the facts. Settings all the way from propaganda to advertising contain some measure of deception. We measure deception using characteristic changes in the frequencies of a set of words that were originally determined empirically but have since been verified in a large number of domains [5]. Deception is signalled by:
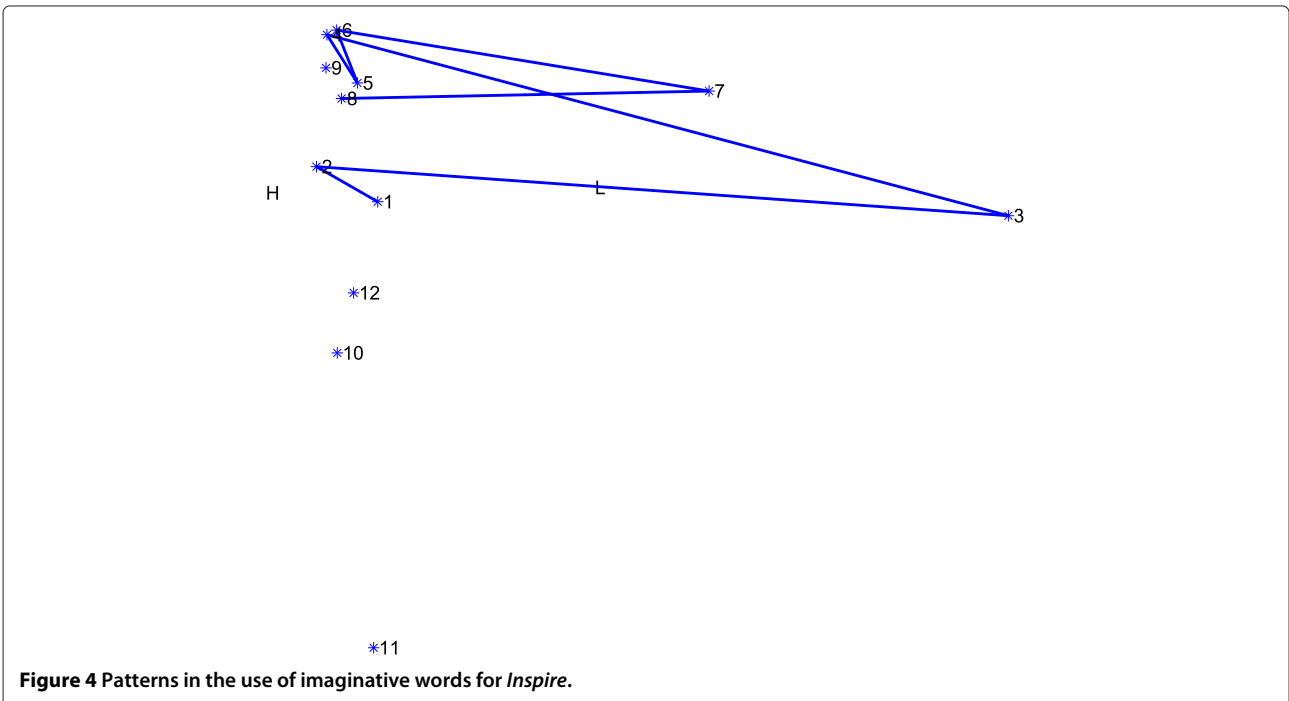
Figure 4 Patterns in the use of imaginative words for *Inspire*.

Figure 5 Patterns in the use of imaginative words for *Azan*.
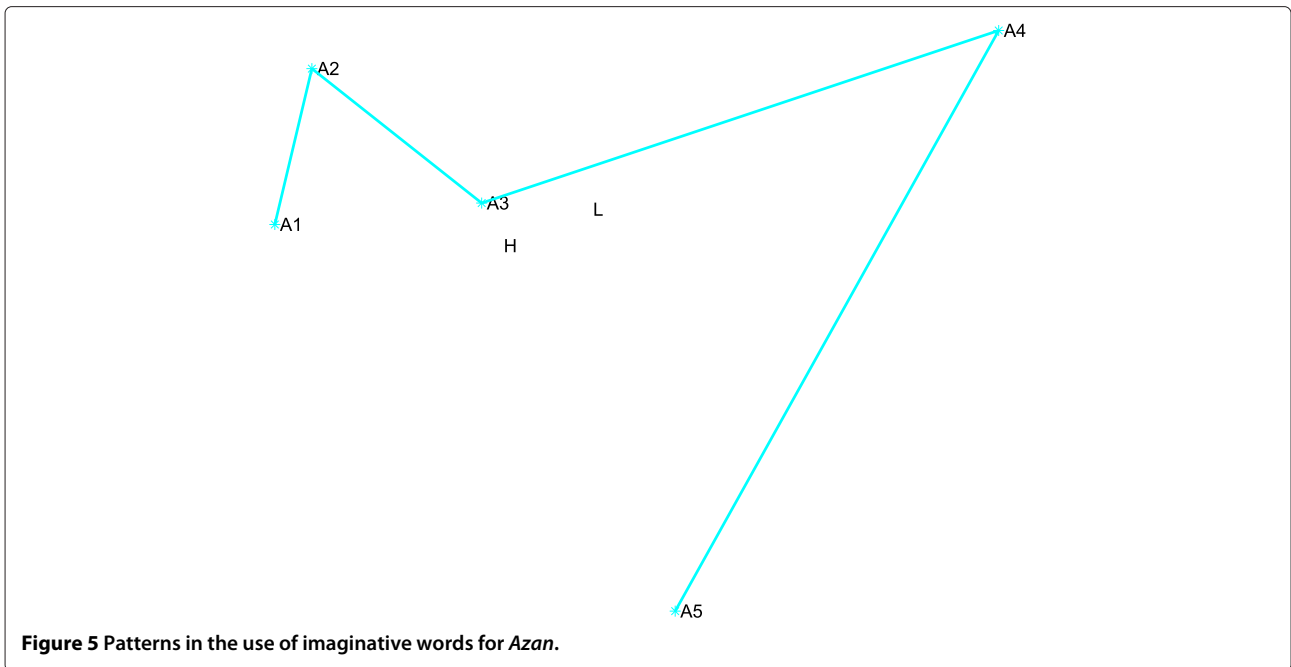
- Decreases in first-person singular pronouns ("I", "mine");
- Decreases in exclusive words, words that signal increasing refinement of the thought being expressed ("but" and "or");
- Increases in negative emotion words ("hate", "angry"); and
- Increases in action verbs ("go", "make").

Because intensity of deception is signalled by *decreases* in the frequencies of some words, word rates are mapped to z-scores, and those in the first two categories are then multiplied by −1 so that increasing magnitude is consistently associated with increasing deception. A singular value decomposition is then applied to the resulting matrix.

Figure 10 shows the structure of the *Inspire* issues based on the deception model. In this figure, deception increase from bottom to top, so that Issue 7 is by far the most deceptive among them.

Figure 6 Comparison across magazines based on imaginative words.
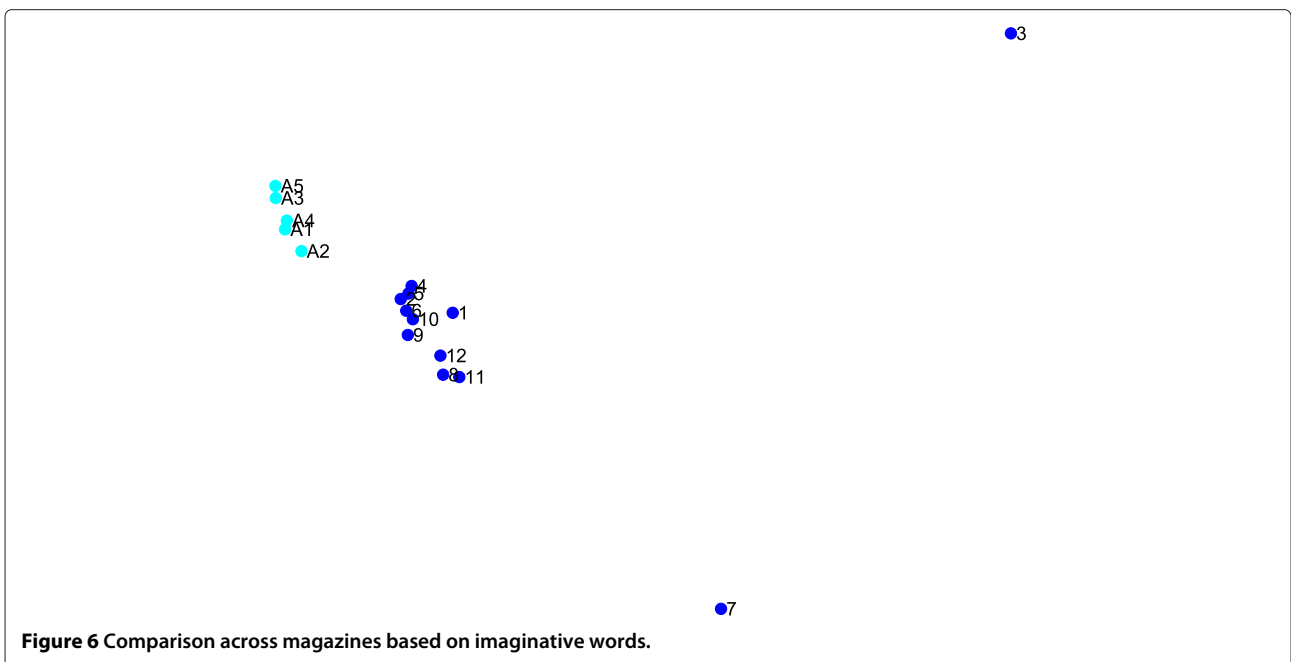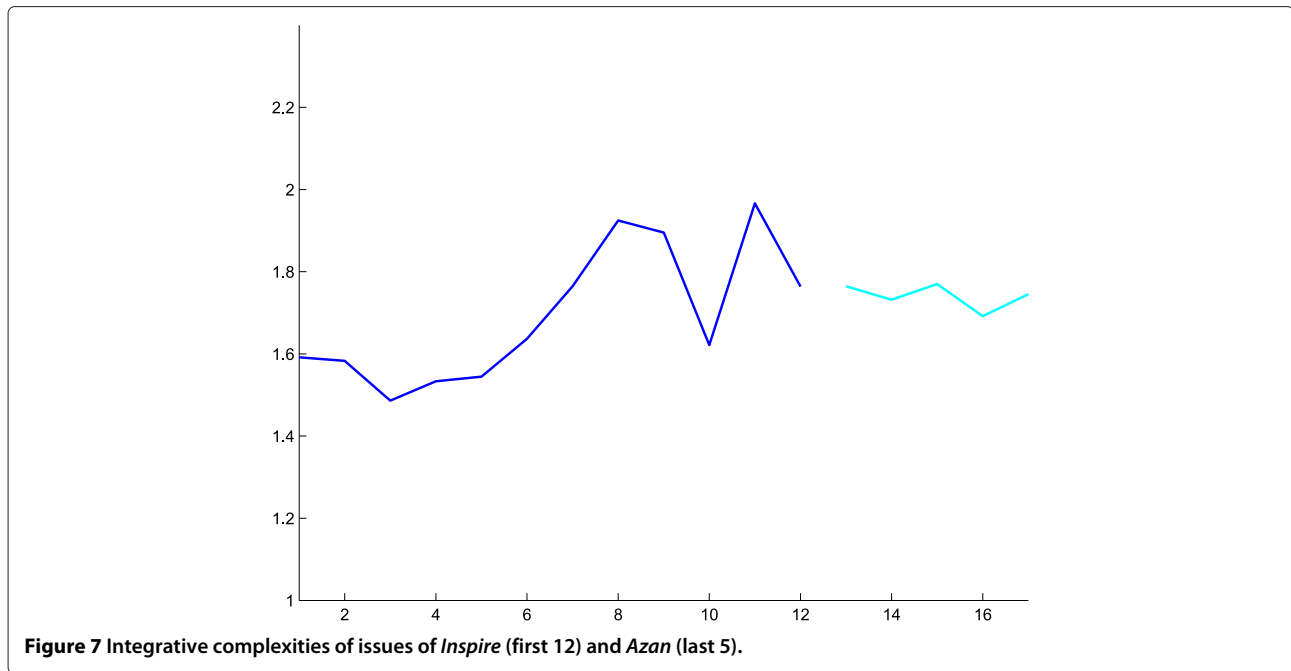
Figure 7 Integrative complexities of issues of *Inspire* (first 12) and *Azan* (last 5).

Figure 11 shows the ranking by deceptive language for the *Azan* issues. Deception increases from lower right to upper left in this figure, so Issue 4 is the least deceptive.

Figure 12 shows the deception ranking for both magazines. The issues of *Azan* are all less deceptive than the issues of *Inspire.* However, the separation between the two magazines is not as great as for some previous word categories.

Jihadist language

Koppel *et al.* [6] developed a language model that measures the intensity of jihadist language. Again, the model was developed empirically. The model words were originally in Arabic; we use English versions obtained using Google Translate. The resulting 85 English words have been shown in other contexts to produce a strong single-factor ranking of Islamic forum postings, suggesting that, although the construction

process is somewhat cavalier, the results are usable [13,14].

Figure 13 shows the ranking of *Inspire* issues based on this model, with intensity increasing from right to left. Thus the newer Issues 9–12 have much higher levels of jihadist language than the earlier issues. Some of the words associated with this increased intensity are "jihad", "mujahideen", and "killing". Figure 14 shows the ranking of *Azan* issues based on this model, with intensity also increasing from right to left.

Figure 15 shows the levels of jihadist language across the two magazines. The issues of *Azan* are at about the same level as the majority of issues of *Inspire.*

The results from the deception and jihadist language models are similar to previous analysis of Islamist forum posts [13,14]. There it was argued that a reasonable explanation was that high levels of jihadist language were plausibly associated with high sincerity, and so low levels of deception. The recent issues of *Inspire*, which have especially high levels of jihadist language, have levels of deception that are in the mid-range, or perhaps even a little lower than that, suggesting that the same relationship may be present.

Gamification language

Gamification has been used in many different settings, for example in business and learning, to increase the attraction of the desired process by adding elements of a game (goals, competition) to them. There are signs that a similar approach has begun to be used in jihadist propaganda as a way to change the calculus of carrying out an attack whose consequences are substantial, by focusing

Table 1 Integrative complexity scores for randomly chosen articles in mainstream magazines

Document	IC	DIAL	ELAB
Atlantic	2.44	1.97	1.63
Chatelaine	2.08	1.60	1.53
Harpers	1.63	1.50	1.28
National Geographic	1.70	1.50	1.26
New Yorker	1.68	1.44	1.28
Readers Digest	2.23	1.78	1.54
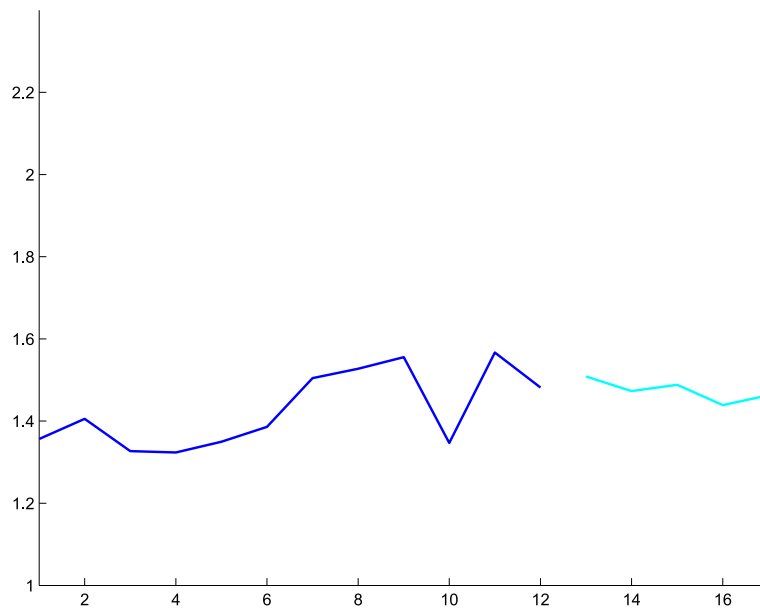
Figure 8 Dialectic complexities of issues of both magazines.

on the process rather than the consequences. In other words, gamification attempts to motivate individuals to take actions towards proximate goals, and perhaps in competition with others, providing notional real-world goals while at the same time blurring or down-playing the real-world consequences associated with achieving them.

We constructed a gamification lexicon using the commonest words from McGonigal's book [19] extended with words from some popular gamification web sites.

Figure 16 shows the intensity of gamification language for the issues of *Inspire*, with intensity increasing from lower right to upper left. As we have already seen, Issues 3 and 7 bask in the success of previous attacks; these results suggest that this comes at the expense of motivating new attacks, so it is not surprising that they show the lowest levels of gamification language. Figure 17 shows the intensity of gamification language for the issues of *Azan*, with the same orientation.
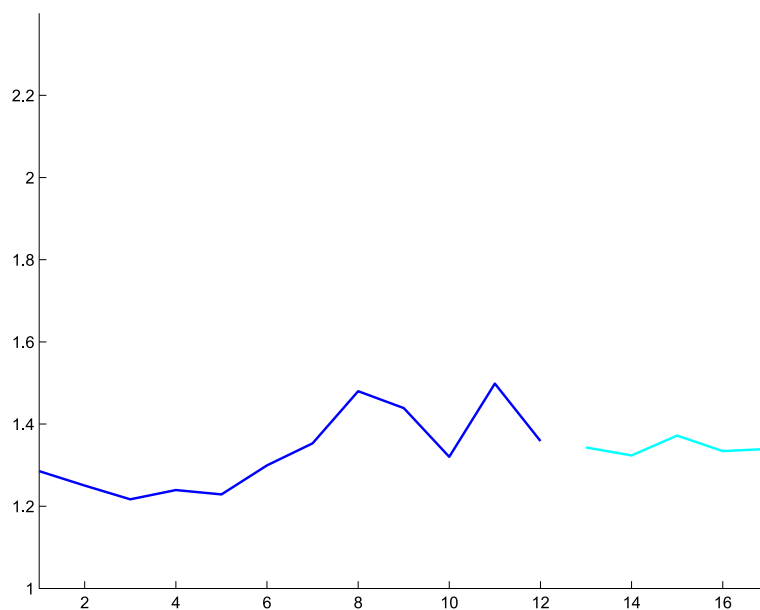
Figure 9 Elaborative complexities of issues of both magazines.

Figure 10 Changes with time in the use of deception markers for *Inspire*.
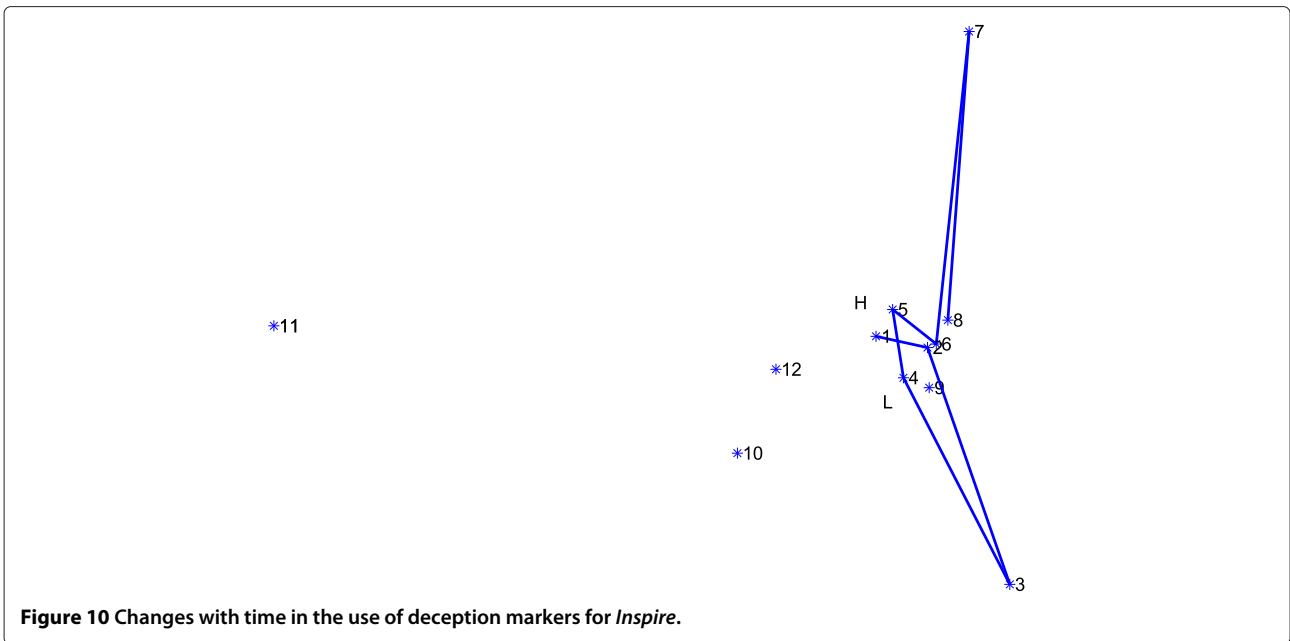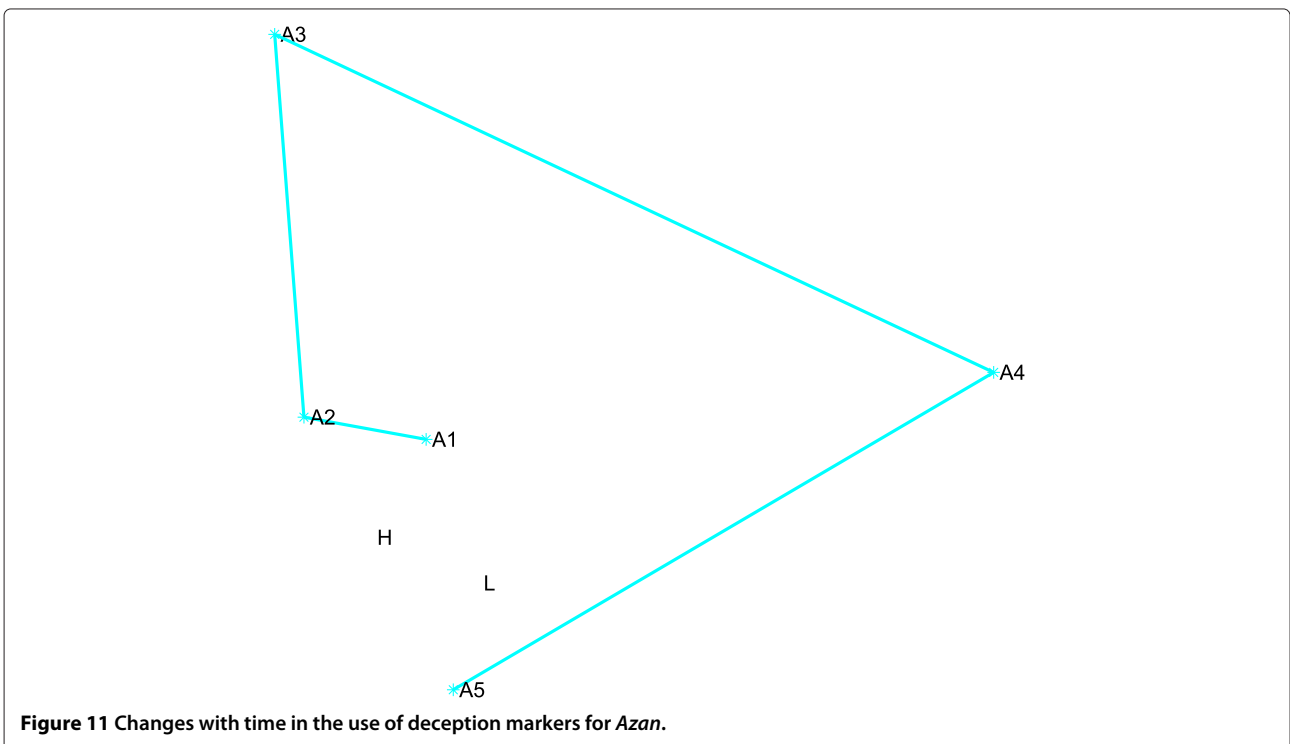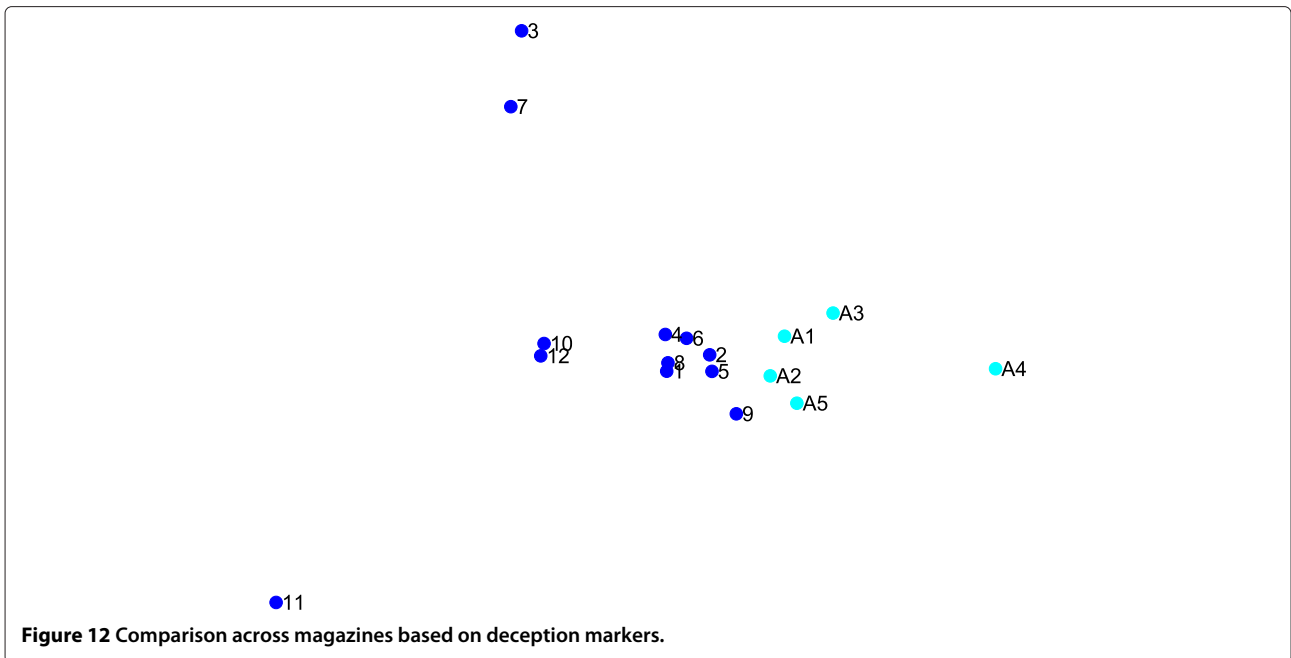
Figure 18 shows the ranking by gamification intensity across both magazines, with intensity increasing from lower right to upper left. There is considerable variation in the observed levels of gamification language, making it hard to judge whether there are attempts to use gamification, but in an uncertain way, or whether the ideas underlying gamification are being used unconsciously.

It is striking that the recent issues of *Inspire* differ so markedly from the others with respect to this model. *Azan* shows consistently higher levels of gamification language than *Inspire* does. So little is known of the authors and editors that it is hard to judge whether this represents intentional or accidental deployment of this style of language.
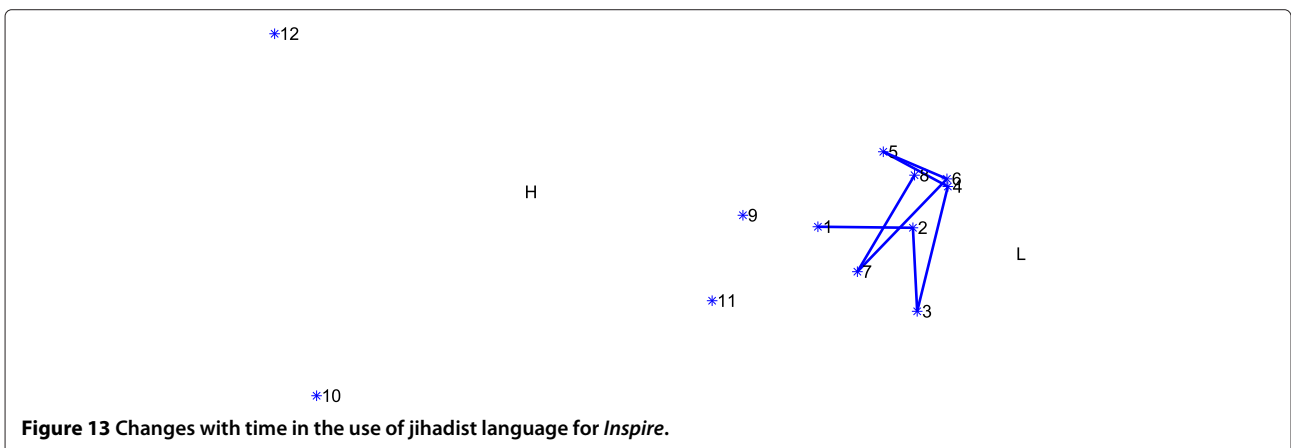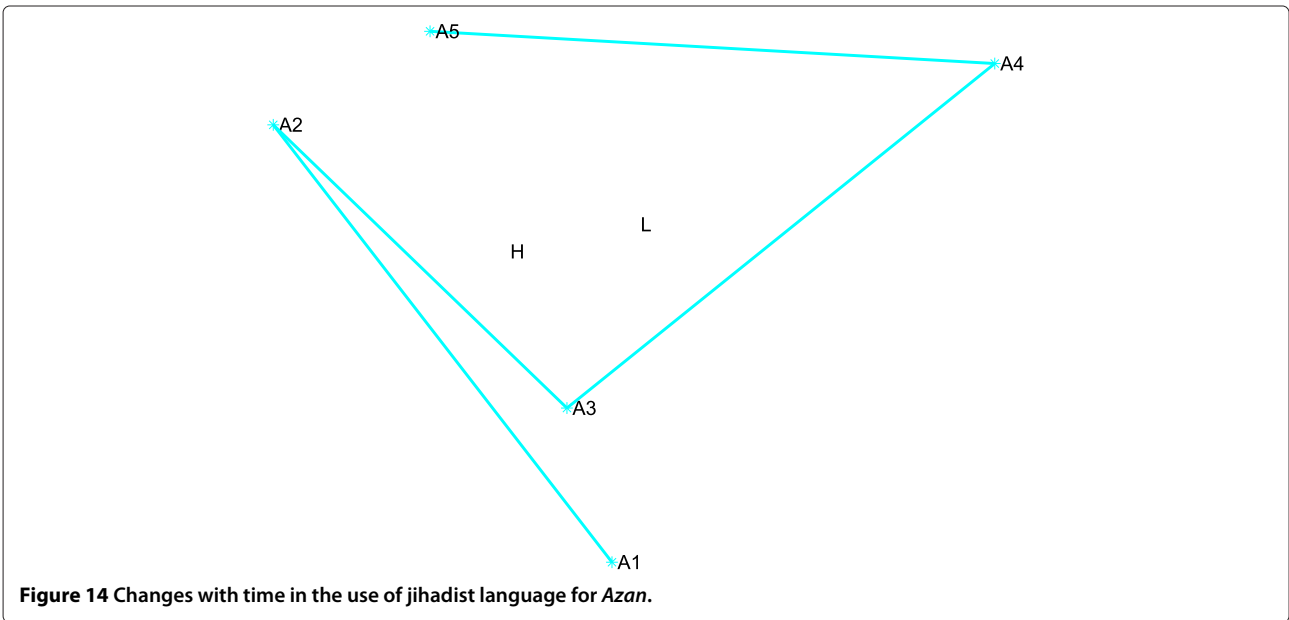

Figure 11 Changes with time in the use of deception markers for *Azan*.

Figure 12 Comparison across magazines based on deception markers.

Discussion

We now revisit our hypotheses, given the results of the empirical analysis. We justify claims of the form "intensity with respect to this particular language model is high" by appealing to the placement of the artificial documents in relation to the magazine issues. In particular, the document indicated by the point labelled "H" represents a document containing occurrences of every word of the model at rates one standard deviation above the mean. Such a document represents extremely high intensity, especially for the informative, imaginative, and gamification models where the 1000 most-frequent words are used. Thus a magazine issue that lies above the level of the "H" document must use the corresponding language at extremely high levels.

For the hypotheses we posited, we reach the following conclusions from the results:

- *Informative language intensity will be high.* This is supported; *Azan* is clearly better in this dimension, and *Inspire* wanders from informative to triumphalist several times.
- *Imaginative language intensity will be high.* This is partly supported; most issue lie just below the "H" marker. The issues of *Inspire* that were lowest on informative language also have low levels of imaginative language.
- *Integrative complexity will be consistent with other magazines.* This is supported; levels of complexity vary but remain in the general range of mass-market

Figure 13 Changes with time in the use of jihadist language for *Inspire*.

Figure 14 Changes with time in the use of jihadist language for *Azan*.
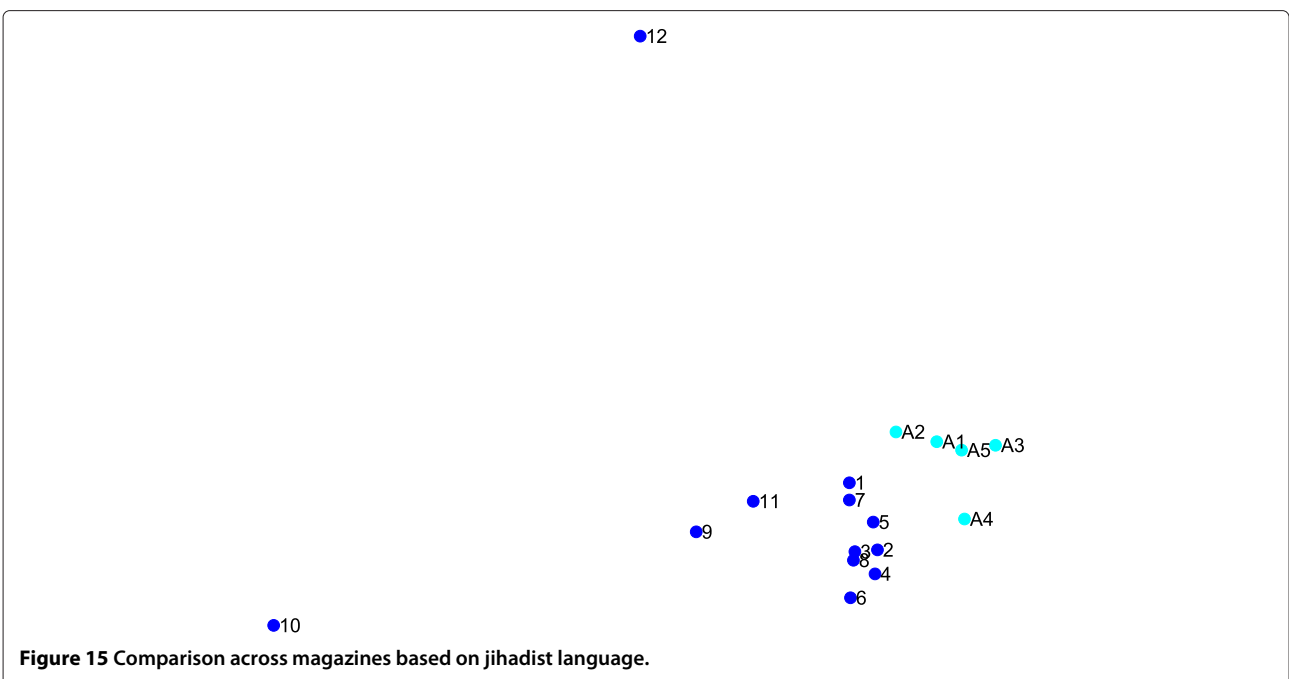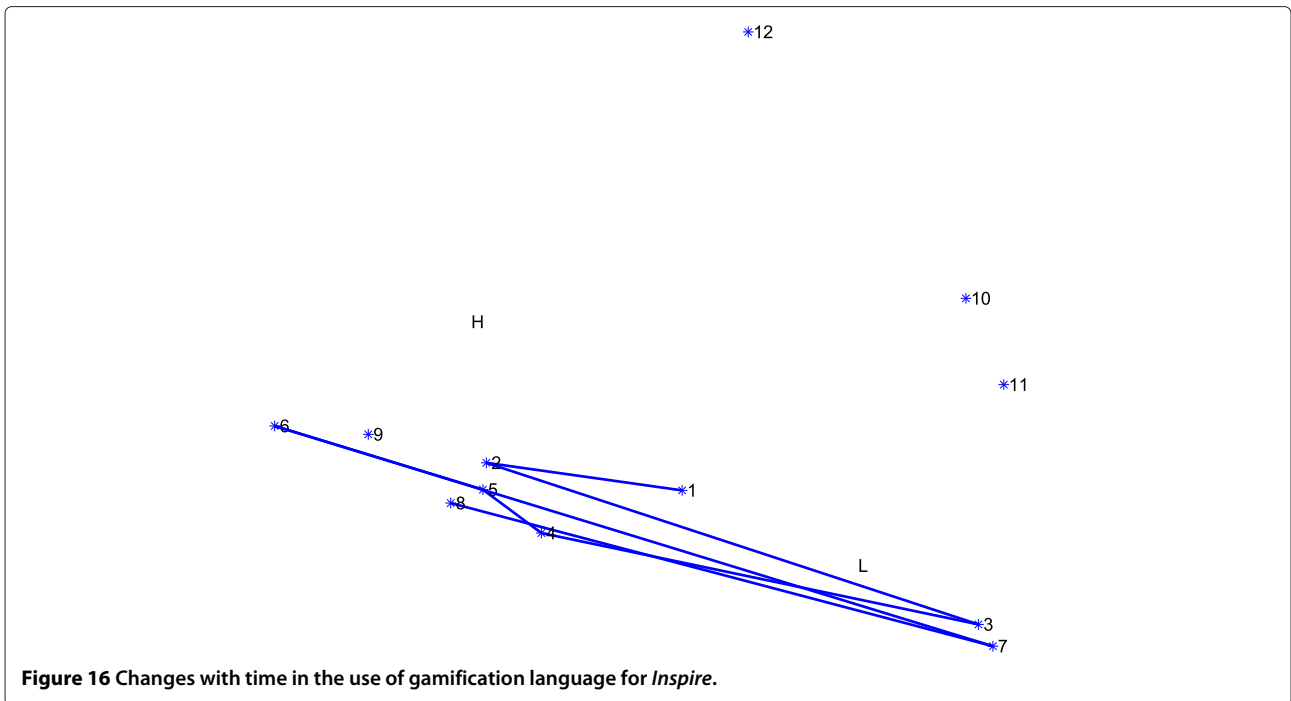
Western magazines. This is counter-evidence to the perception that Islamist propaganda is necessarily simplistic.

- *Deceptive language intensity will be high.* This is partially supported – most issues of *Inspire* are in the mid-range, with issues of *Azan* noticeably higher.
- *Jihadist language intensity will be high.* This is partially supported. Again, the issues of *Azan* have higher intensities than the older issues of *Inspire*, but the three most recent issues of *Inspire* show the highest levels overall.

- *Gamification language intensity will be non-negligible.* Some issues have very high levels of gamification, but others have very low levels, and there is no temporal pattern. The results suggest that gamification is being attempted, but with unsure technique leading to inconsistent levels.

The results based on these semantic language models are consistent with expectations of propaganda and the language of influence. Thus jihadists have adopted the

Figure 15 Comparison across magazines based on jihadist language.

**Figure 16 Changes with time in the use of gamification language for *Inspire*.**

playbook of governments, as they try to influence readers to adopt a jihadist ideology and, more importantly, to act on it.

Authorship is detectable. We can certainly discriminate the first 8 issues of *Inspire* from Issues 9–12, although Issues 3 and 7 are substantially different from the other earlier ones. The early issues may have been a collaborative effort between Khan and al-Awlaki, with the balance shifting depending on subject matter. All of the issues of *Azan* appear to be the work of a single author (despite the different author names associated with individual articles).

Azan issues are more consistent than those of *Inspire*, which shows a tendency to focus opportunistically on the issues of the moment, rather than on a strategic plan to convince and develop lone-wolf attackers. *Azan*'s admitted small number of issues have been focused and consistent, which speaks to tight editorial control, and intentionality. The early issues of *Inspire* were

Figure 17 Changes with time in the use of gamification language for *Azan*.
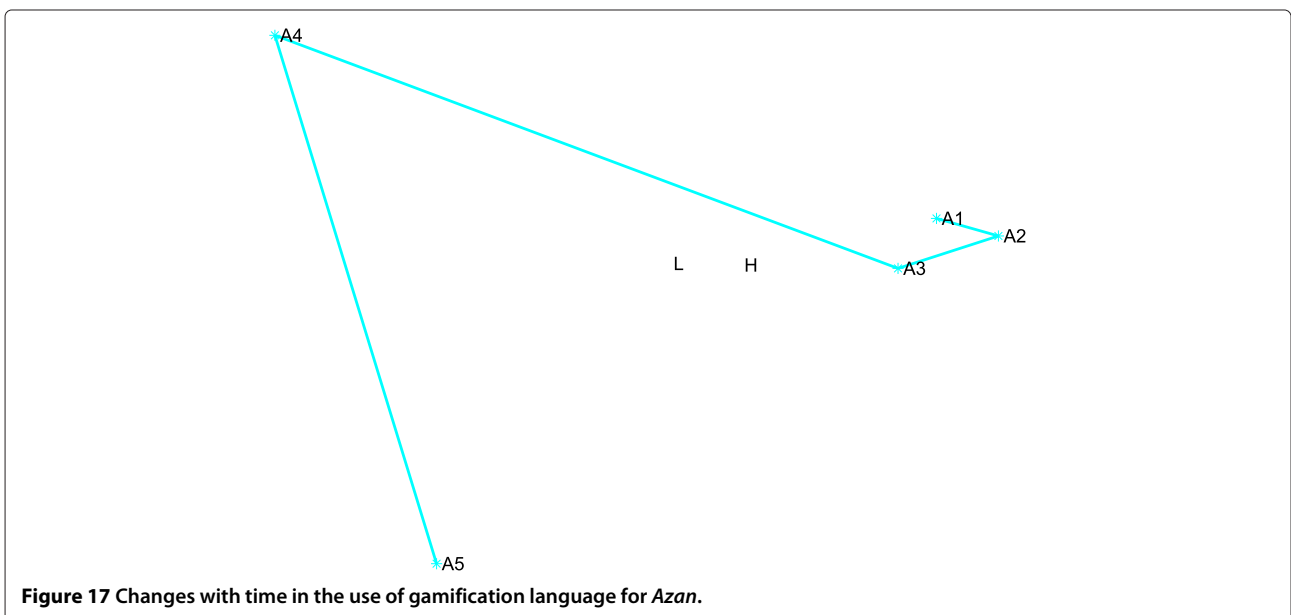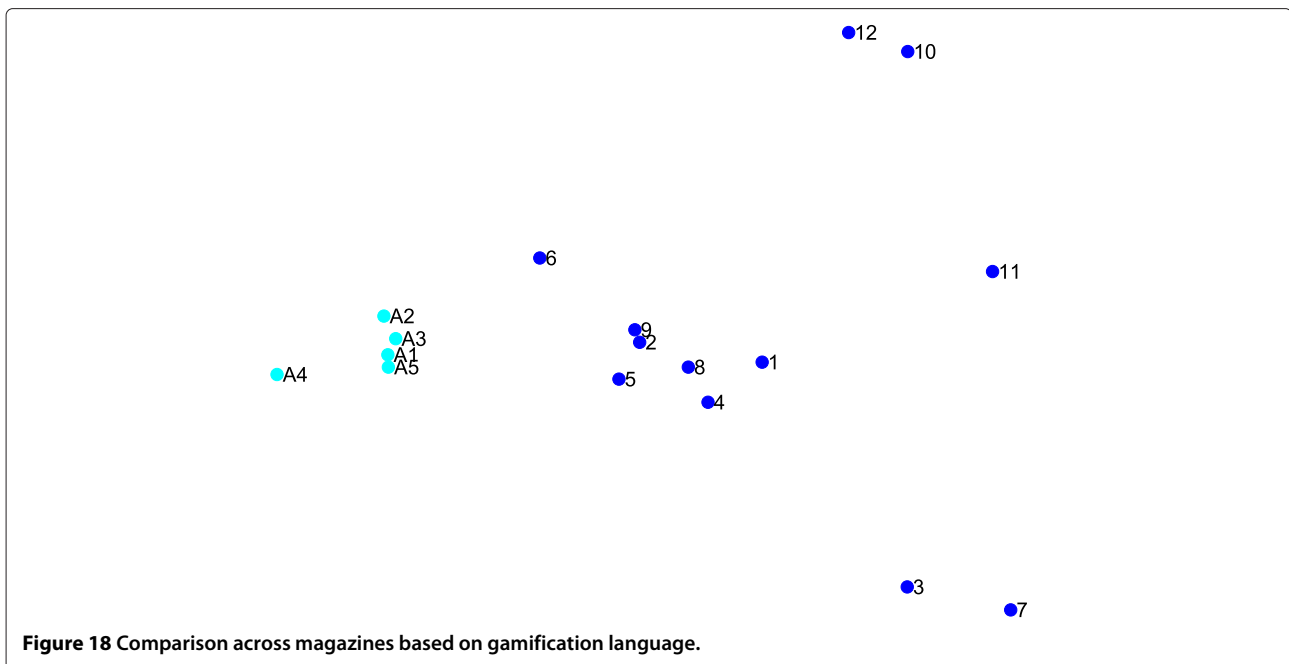
Figure 18 Comparison across magazines based on gamification language.

not nearly as focused or consistent; rather they tend to use current events as a jumping off point for each issue. This may be, to some extent, deliberate but it suggests the lack of a strategic and intentional view of the goals of the magazine. The more recent issues have been different from one another, but the common structures relating them in the various figures suggest that there is some deeper consistency that is now in play. This may some background figure who is exercising a kind of pseudo-editorial control, or it may reflect the homogeneity of the community from which it emanates.

It has been suggested that competition is one of the factors that might cause *Inspire* to become more strident [20]. The intensity of *Inspire* is visibly increasing, but there has only been minimal overlap in time of the two magazines, so it is too early to examine this conjecture.

We have demonstrated a methodology that can be applied to a corpus in at least a semi-automated way to extract data relevant to a set of language models; and to visualize the properties of the documents in the corpus with respect to these language models. Consideration of similarity/variation and temporal changes allow us to see aspects of the authors and of their mindsets.

From an intelligence perspective, this methodology allows us to draw conclusions about the changing editors at *Inspire* after the deaths of its founding editors; and the pool of authors and editors associated with *Azan*. We are also able to judge, to some extent, the sophistication of the authors and editors as they deploy persuasion techniques, visible via some of these language models. We are also

able to judge intensity of properties such as informativeness, imaginativeness, deception, and (most significantly perhaps) jihadist language.

Author details
[1] Queen's University, Kingston, Canada. [2] James Madison University, Harrisonburg, USA.

References
1. G Weimann, Lone wolves in cyberspace. J. Terrorism Res. **3**(2) (2012)
2. R Martin, *Propaganda and the Ethics of Persuasion*, 2nd edn. (Broadview Press, 2013)
3. P Rayson, A Wilson, G Leech, Grammatical word class variation with the British National Corpus sampler, in *Twenty First International Conference on English Language Research on Computerized Corpora*, (2000), pp. 295–306
4. P Suedfeld, P Tetlock, Integrative complexity of communications in international crises. J. Conflict Resolution. **21**, 169–184 (1977)
5. ML Newman, JW Pennebaker, DS Berry, JM Richards, Lying words: predicting deception from linguistic style. Pers. Soc. Psychol. Bull. **29**, 665–675 (2003)
6. M Koppel, N Akiva, E Alshech, K Bar, Automatically classifying documents by ideological and organizational affiliation, in *Proceedings of the IEEE International Conference on Intelligence and Security Informatics (ISI 2009)*, (2009), pp. 176–178
7. SA Ford, Inspiring a narrative: a content analysis of Al Qaeda's English-Language Inspire magazine. Master's Research Essay, Carleton University (2012)
8. EF Reid, AQAP *Inspire* magazine: implications for crowdsourcing and gamification. Technical report, Research Paper, National Intelligence University (NIU), Washington, D.C. (2012)
9. EF Reid, Crowdsourcing and gamification techniques in inspire (aqap online magazine), in *IEEE International Conference on Intelligence and Security Informatics (ISI)*, (2013), pp. 215–220
10. G Ramsay, Targeting, rhetoric and the failure of grassroots jihad. J. Terrorism Res. **3**(1) (2012)

11. J Woo, J Son, H Chen, An SIR model for violent topic diffusion in social media, in *Proceedings of 2011 IEEE International Conference on Intelligence and Security Informatics, ISI 2011*, (2011)
12. AF Lemieux, J Brachman, J Levitt, J Wood, *Inspire* magazine: a critical analysis of its significance and potential impact through the lens of the information, motivation, and behavioral skills model. Terrorism Pol. Violence. **1**(1), 1-18 (2014)
13. DB Skillicorn, Lessons from a jihadi corpus, in *Foundations of Open-Source Intelligence FOSINT 2012*, (2012)
14. DB Skillicorn, Applying interestingness measures to Ansar forum texts, in *Proceedings of KDD 2010, Workshop on Intelligence and Security Informatics*, (2010), pp. 1–9
15. GH Golub, CF van Loan, *Matrix Computations*, 3rd edn. (Johns Hopkins University Press, 1996)
16. P Suedfeld, PE Tetlock, S Streufert, Conceptual/integrative complexity, in *Motivation and Personality: Handbook of Thematic Content Analysis* (Cambridge University Press, 1992), pp. 393–400
17. LG Conway III, F Thoemmes, AM Allison, KH Towgood, MJ Wagner, K Davey, A Salcido, AN Stovall, DP Dodds, K Bongard, KR Conway, Two ways to be complex and why they matter: implications for attitude strength and lying. J. Pers. Soc. Psychol. **95**(5), 1029–1044 (2008)
18. L Conway III, KR Conway, LJ Gornick, SC Houck, Automated integrative complexity. Pol. Psychol. (2012). in press
19. J McGonigal, *Reality Is Broken*. (Penguin Books, 2011)
20. DB Skillicorn, C Leuprecht, Improving the language of influence, in *Proceedings of Foundations of Open Source Intelligence, at Advances in Social Network Analysis and Modelling (ASONAM)* (ACM & IEEE, 2013), pp. 1028–1033