A Social Influence Model for Exploring Double Subjectivity through News Frames in Online News

Loretta H. Cheeks
CIDSE, Computer Science
Arizona State University
Tempe, AZ U.S.A.
loretta.cheeks@asu.edu

Ashraf Gaffar
CIDSE, Software Engineering
Arizona State University
Mesa, AZ U.S.A.
ashraf.gaffar@asu.edu

***Abstract*—The Internet is the premier platform for the proliferation of news. Online news is unstructured narrative text that embeds facts, frames, and bias that can influence society about critical issues. Online news sources are carriers of news that operate within complex and dynamic networks. The informational flows, interactions, and structural variations of online news lead to asymmetries and influence societal attitudes and beliefs. Current efforts to express the internal embeddedness in online news text are limited by the use of existing computational tools. This research has the potential to inform advanced machine learning and to help researchers to understand implicit structural embeddedness to address real-world critical issues. This paper establishes the new concept of double subjectivity, proposes a formal definition of a news frame issues network, and introduces a social influence model for the discovery of distinct pathways. The paper also exposes opinion shifts in complex networks for future computational treatment of narrative text.**

Keywords—Artificial intelligence; bias; data mining; framing; machine learning; natural language processing; news source sentiment; sentiment analysis; social influence; subjectivity

I. INTRODUCTION

The Internet is the premier platform for the proliferation of online news and has been compared to a huge social and psychological laboratory [39]. News sources disseminate online news that operates within a complex and dynamic system. This system comprises informational flows, interactions, and structural variations, which lead to asymmetries and have a tendency to influence society. Online news is a form of narrative text that embeds facts, strategic communication frames, and bias for shaping and transitioning group standards, values, attitudes, and beliefs about critical issues. The embedded facts represent absolute, relevant truths characterized by their rigid nature. These absolute truths are shared among news sources. However, the exercise of news framing—and the bias that it conveys—is where multidimensional aspects of the content emerge and where expressivity happens for transformational and maximal influence. Framing is the use of strategic devices for presenting prominent aspects and perspectives about an issue using certain keywords as well as stereotyped images and sentences for the purpose of conveying latent meanings about an issue [30]. News sources play a critical role in the propagation of bias about issues, and bias in online news drives a wedge between evidence and beliefs [36]. In addition, the combination of facts, frames, and bias represents forces that offer a space of possibilities for exploring framing effects that could not be realized in the functioning of the parts alone.

When attempting to identify potential pathways to address the complexities and dynamics that exist in news text for understanding the production of language, a single computational approach is inadequate. Instead, this type of problem requires an integration of multiple models and perspectives that advance what is known about hidden meanings and latent structures in text. An acknowledgment of the multidimensional aspects—degrees of freedom—of the narrative text in news call for computational research to fill knowledge gaps and to transform the current understanding of alternative pathways to learning.

In exposing the discovery of new knowledge about text, transformations (or shifts) in which dynamical forces reach a critical point make clear the new structures, pathways, and influence patterns to be accepted or new questions or alternatives in the scope of an issue to be revised. The ability to seek out novel, uncharted relationships that are implicit from a body of unstructured narrative text—yet not explicitly stated within that text—has obvious scientific value. This process is called novel knowledge discovery. It enables the use of the current state of knowledge for an emergence of possible new relationships and pathways that have yet to be studied.

This manuscript includes an examination of online news framing within articles—a form of narrative text—to expose a deep level of expressivity—what is referred to in this paper as Double Subjectivity. The aim of the authors is to demonstrate news framing as a step toward deep learning about the production of language that may inform scholars about intelligent behavior in machines.

To the best of the authors' knowledge, in computer science scholarly research, examining double subjectivity using online news and framing effects has not been addressed previously for the class of problems discussed in this paper. Insight into the following questions is provided this research inquiry: What is the migration path of online news frames as it evolves over time? What is the effect of double subjectivity? What are the conditions observed for determining shifts in belief?

The following sections will describe the conceptual framework, related works, background on the issue of

978-1-5090-6435-9/17/$31.00 ©2017 IEEE

application, methods, possible space transitions, observations of the models in action, and the conclusion.

II. Conceptual Framework

The chosen method of communication has tremendous influence on the perceived and real value of knowledge. Through automation, to a substantial degree, computers have been used to leverage people's ability to communicate and acquire knowledge. However, although computers can be used to generate and disseminate data, they have less ability to analyze the hidden meanings of data. Humans still have the unique ability to get beyond the syntax of data in order to make meaning.

The automation of data processing can be generally categorized under two main areas: formal data and informal data. Formal data—represented by mathematical formulas and computer languages—have little room for subjective interpretation and are typically concise and predictable. For instance, a correctly developed source code can compile and run on a computing machine and will always produce the same output for a specific input. The combination of a formal set of instructions resulting from the successful compilation and run of source code can transition a computer through a set of predefined states, depending exclusively on the compiled code and input data. This leaves no room for interpretation or subjectivity. In other words, the meaning and interpretations must be fully represented in the source code. A good programming language must, therefore, have sufficient expressiveness to fully represent a feasible solution in a specific problem space. If a source code arrives at an undefined state, it will simply freeze or crash; it is unable to make an independent decision. Some areas of computer science such as fuzzy logic and AI attempt to address these inherent limitations. However, these approaches are not the focus of the work conveyed here.

Although the formal language of source code is a good medium for communicating with computers, natural languages are the medium of communication among humans, and natural languages are informal in comparison to the formal language of computer code. Such informal data can be further divided into two subcategories, depending on the source of origination: direct and indirect.

Direct informal data are the result of a person's encounter with an event and the perception, in part, of its details firsthand. Human perceptions of events lead to personal interpretations of events. Interpretation depends greatly on the viewer's perception, as different people can arrive at different interpretations of the same event. Perception is often associated with the viewer's state of mind, prejudice, background, and beliefs. Two persons witnessing the same event or being involved in the same situation could assign different interpretations and end with separate conclusions. Although the external situation is the same (the perception of which is uniquely defined in this paper as the absolute baseline), it could be internalized differently by different people. Such a case is a first-level subjectivity. Even when different people are asked to report an incident or a situation (the same absolute baseline), different accounts is often the result, indicating first-level subjectivity. Many factors further affect this subjectivity:

- Complex situations, such as movie showings, are harder to perceive and interpret compared to simple or atomic events like the occurrence of a rocket launch.
- Unlike a formal language, a natural language does not fully express the contents in its syntax. Natural languages report "about" something, rather than fully describing the event in a specific way.
- Humans have different expressive abilities and styles in reporting events using natural languages, leading to increase in subjectivity.
- The ambiguity of natural languages allows for different interpretations of spoken or written sentences.

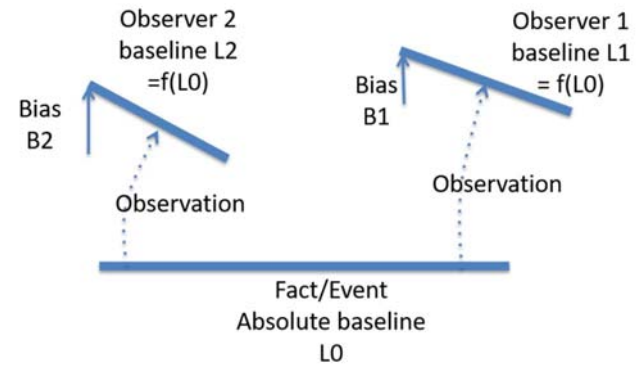
The aforementioned factors explain why—even for different people directly witnessing the same situation—substantial subjectivity can be present in reporting or discussing an event that was mutually witnessed.

People often try to reduce first-level subjectivity. For example, television reporters try to abide by certain rules to maximize neutrality and leave it up to the viewers to build their own opinions freely. The scientific community tries to show the stated facts through rigorous research and scientific experiments in an attempt to eliminate or reduce subjectivity. However, subjectivity is not so simple to discard. Consider newspaper and television reporters who are known for their opinions.

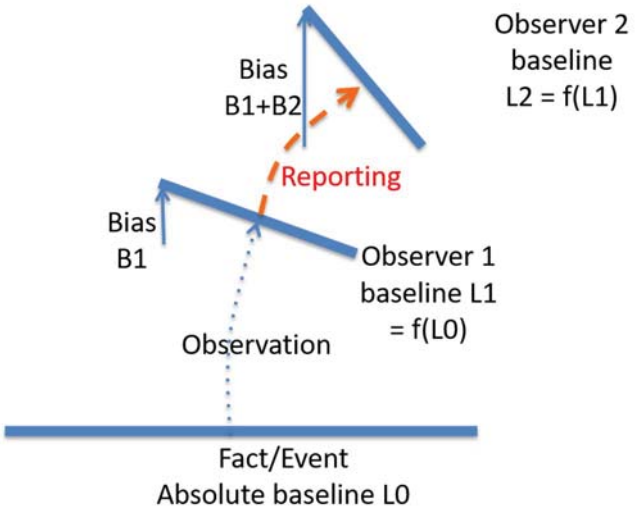
The Internet and the worldwide web have led to an explosion of indirect informal data although events can still be directly perceived and subjectively interpreted. Because of technology advancements, it is possible to receive a much larger number of events indirectly. Because these events are not directly witnessed, an absolute baseline is not available. Because knowledge of an event is often based on the written or spoken reports of others, the baseline will be relative to the reporter's subjectivity or simply "float."

With the acceptance of indirect reports and their interpretation, the receiver adds his or her own subjectivity. Again, no longer is the absolute baseline available. It is replaced with a reporter-relative baseline or a floating baseline. If the reporter's subjectivity is strongly biased towards certain beliefs, the receiver of the account already starts with a baseline that is completely different from the absolute baseline. This new baseline could strongly affect the receiver's beliefs. Double subjectivity is the term introduced in this paper to describe the secondary subjectivity that results after receiving indirect informal data. Fig. 1 illustrates the cumulative bias effects of double subjectivity.

978-1-5090-6435-9/17/$31.00 ©2017 IEEE



(a)



(b)

Fig. 1. Two representative illustrations of the double subjectivity framework: (a) There is an event (Observation) about an issue that two observers see: Observer 1 and Observer 2. Even though it is the same event, their perceptions are different. Yet, their bias may be different, but independent. (b) This graph illustrates the introduction of an indirect account of an event, whereby the reporter does not receive the information firsthand. Rather Observer 2 adopts the report of Observer 1, which sets into motion the cumulative effect of double subjectivity.

The illustrated formula depicts a generic function for double subjectivity in online news. Here, the assumption is that there is an observer-dependent function $f(x)$, such that for first-level subjectivity, it is expressed as:

$$L_1 = f_1(L_0), \text{ and} \quad (2)$$

Whereas, double subjectivity is expressed as:

$$L_1 = f_1(L_0), \text{ and} \quad (3)$$
$$L_2 = f_2(L_1)$$

Where, f is a bias intensity function that depends on the person observing, and then reporting a situation or event.

An analogy drawn from photocopying illustrates double subjectivity in that, ideally, a photograph copy should be a close replica of the copied photograph. Yet, in actuality, the resulting photograph is the copied photo with its bias plus the distortion bias gained from the copying process or machine.

Online news is exemplary of double subjectivity in that its assemblage consists of subjective properties such as context, interpretation, multimodal interaction, background, compensation, and assumptions. These properties are critical building blocks that characterize narrative text for its (1) signification (i.e. the semantic content signifying an association or concept given a word, sentence, or phrase) and (2) significance (i.e. the relevance, rank, importance, or capacity to make a difference). Unfortunately, there remain substantial gaps in advancing what is known about the production of language in narrative text where double subjectivity exists. Therefore, exploring deeper levels of expressivity with an emphasis on double subjectivity may inform a new direction in research.

Discovering double subjectivity in narrative text poses a difficult problem, as it seeks to give insight and meaning to open-ended statements and indirect data. More complex structures within the narrative text provide an expansive landscape for advancing effective tools and crafting applications to automate many language-related tasks, for example, document summarization, automated text generation, and many others. To show the gains in logic and reasoning, Fig. 2 depicts advances made using current tools and new frontiers that double subjectivity may shed light on, such as poems, metaphors, idioms, jokes, art, and storytelling. This is to be achieved by exploring levels of narrative text.

III. Related Works

In the literature reviewed, no computational treatments were identified to address (1) the concept of double subjectivity for automatic emergence of knowledge of narrative text or (2) a model for exploring the interactions and implications of news frame choices. However, several works were identified that have some relevance for the research topic.

978-1-5090-6435-9/17/$31.00 ©2017 IEEE

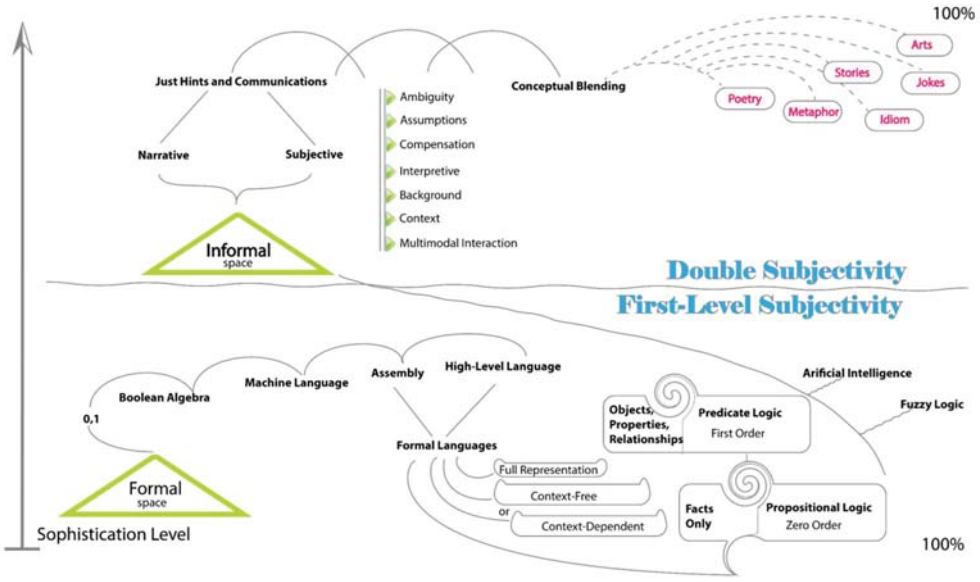


Fig. 2. A depiction of the evolution of first-level subjectivity and double subjective potentials.

A. *Text Summarization, Text Data Mining, and Information Retrieval*

In computer science, scholarly research on latent meanings in the association between terms and documents to reveal relationships is found in literature related to text summarization, text data mining, and information retrieval. The earliest paper on text summarization identified for the current report is a description of work at IBM in the 1950s [15]. In his work, the author proposed that the frequency of a particular word provides a useful measure of its significance, and he identified the concept "term frequency" (TF), which states that it is possible to identify the significant terms based solely on the term's calculated frequency within that document [15]. Term frequency relates to average information, or entropy, of a term or group of terms in the ranking of their relationship to each other.

Outstanding contributions to text data mining in the 1990s and beyond, with the advent of machine learning, include text representation and models construction [16] [17] [18] [19] [20]; data dimensions reduction research in feature extraction [21] [22]; and deep semantic mining [23].

The research on online news frames reported in this paper is motivated by framing theory, which focuses on understanding the latent meanings of observable messages in their contexts [24] and can provide important insight into how the presentation or framing of an issue affects the choices people make. Other disciplines have focused on framing. In linguistics research, similar approaches are also described as "latent semantic analysis" (LSA) [25]. Furthermore, social network analysis (SNA) focuses on the importance of relationships among interacting units [26].

B. *Opinions and Social Influence*

Whereas text data mining techniques and methods have given insight into first-level subjectivity, dynamic systems modeling provides insight into the double subjectivity that exists in narrative text. Models of social influence, cultural dynamics, and information diffusion (i.e. topic modeling, sentiment analysis, and opinion mining) are active areas of research [7] [1] [27] [28].

The opinion models, such as the voter model, give insight into the spread and distribution of opinions [1]. It has been posited that social impact theory explains the impact of a social group on an agent as being dependent on the prominence of the social sources, their proximity, and source group mass (i.e. the number in the group) [6]. Political opinions and Axelrod's cultural dynamics model behave similarly to these models that capture the interplay between selection and influence [2] [3] [4] [5].

IV. BACKGROUND ON THE ISSUE

The framing of an issue has an impact on the ultimate course of the issue. In order to observe the potential pathway of informal, natural language, the authors of this paper took as a subject water policy. The ways that water policy and decisions are framed affect water rights allocations, policy decisions, human consumption, emerging technologies, farming techniques, and agricultural outcomes [29]. Water is a fundamental resource affecting all aspects of life on earth. Water is used for human consumption, industrial processes, production of food, sanitation, as well as other usages. The issue of water insecurity in the U.S. Southwest region is particularly important in that seven states (Arizona, California,

978-1-5090-6435-9/17/$31.00 ©2017 IEEE

Colorado, Nevada, New Mexico, Utah, and Wyoming) rely primarily on fresh groundwater flows originating from the Colorado River. The Colorado River was a sufficient source of water for decades; however, due to a decade of drought, this once plentiful source of water cannot meet the demand to sustain life as it exists in this region. This situation contributes to water insecurity throughout a considerable region of the United States.

Water insecurity in states that source their water from the Colorado River exemplifies an issue that is appropriate for study through the lens of double subjectivity and that can be observed via the framing strategies used by online news sources. The topic of water insecurity is ideal because this issue is context specific, complex, and characterized by multiple interacting forces. Newspapers in the Southwest regularly produce in-depth articles about drought, water, and climate change that are published online. However, the general public outside of the U.S. Southwest has very little experience with water insecurity, even though the consequences of water insecurity affect other parts of U.S. in significant ways. Biased framing is more likely to influence uninformed respondents [31] or respondents with reduced exposure to or interest in an issue [23]. Therefore, citizens' attitudes and beliefs about water insecurity are likely influenced by the way reporters frame this issue. However, coverage of elites in articles is of particular interest in this study as well, as this factor has an effect on the bias intensity that news sources adopt over time. According to Druckman (2001), the way elites frame an issue is a driving force for shaping public opinion.

The source of unstructured text analyzed in this research offers researchers a body of content in which to explore the production of language. The issue of water insecurity is a case that can be studied to explore the patterns of interaction between online news sources and the effects of their frame choices. This research can generally apply to any large, unstructured dataset that is affected (or biased) by framing. It can, therefore, be extended to other applications such as economics or political discourse.

The aims for constructing a news frame issues network using water insecurity as a contextual case are (1) to gain a deep understanding of how framing advances novel knowledge discovery with sensitivity to double subjectivity as a factor in the production of language in text narratives and (2) to explicitly examine the temporal dynamics of structural changes that may exist within the network. The overarching research problem is the following: "How is the issue of water insecurity in the Southwest region of the United States being framed in online news media over time?"

V. METHODS

A. Dataset Description

The news data for the study were collected from Google News,[1] a newsfeed aggregator. Each article was selected based on four characteristics: (1) its publication by an online news source, (2) the prominence of the news source that generated the articles, (3) the frequency of news publications by sources over time, and (4) the salient terms contained in the article. All articles were published on the Internet between January 2006 and March 2015.

Articles were gathered by limiting the search for keywords associated with water or subtopics of water consequences within Arizona, Colorado, California, and Nevada. These states were chosen because they represent the fastest growing and most populated states that rely on Colorado fresh water. After the collection of the articles, the articles were stored in a database for undergoing data pre-processing that included removing errors and inconsistencies to improve data quality.

B. Construction of Issues Networks for Online News

The first step in the construction of the issues network for news frames is to define the nodes. The issues network is a collection of news sources and article nodes of interest joined together in pairs by edge connections. In this research, the news articles being published by online news sources and the news sources that generate the articles served as the nodes. The boundary of the network in the study was determined to be water or subtopics of water consequences within Arizona, Colorado, California, and Nevada and the online news sources who produce articles on this issue. These states are included in the map in Fig. 3, which shows states that are in drought. It is the reasoning of the authors that news sources play a critical role in the propagation of bias about issues, and bias in online news drives a wedge between evidence and beliefs [36].

Since water is a fundamental resource affecting all aspects of life, one can expect that the number of articles being produced on the issue is an indicator of its prominence and the framing choices of the news sources. The framing of water issues is key for driving policy and decision-making, in addition to shaping attitudes and values. Such framing also indicates areas of bias. Connections are made between news sources through the capture of framing adoption over time as it happens during migration. Migration pathways can be understood as carriers and adopters of online news frames. Undoubtedly, reciprocal relationships (i.e. news sources using the same frame choice) can be valuable, but such relationships are a challenge to capture using the article content. This is largely due to news sources not crediting other news sources

[1] http://news.google.com/

978-1-5090-6435-9/17/$31.00 ©2017 IEEE

in reports. The resulting network aim is to capture structures that were created and maintained through repeated patterns of framing and to capture the migration path of those structures over time.

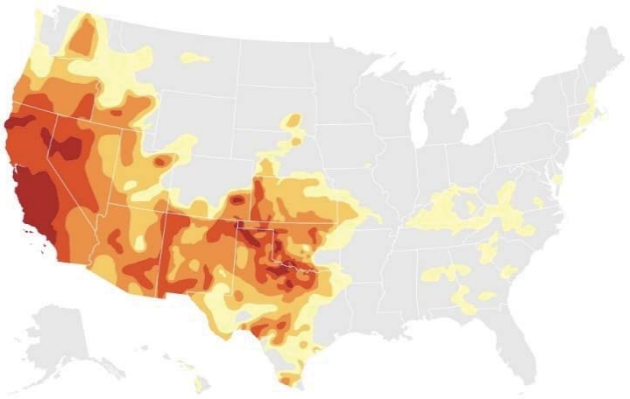


Fig. 3. Drought severity impacting the U.S. Southern region. Source: National Climatic Data Center, National Drought Mitigation Center, U.S. Department of Agriculture, National Oceanic and Atmospheric Administration.

C. Graphs and Problem Formalization

Framing of critical socio-environmental issues in online news is a complex and dynamic social network, which can be studied as a graph, as it is here. Graph analysis has become important in understanding the dynamic process in the production of news frames. Graphs allow for the visualization of the social network, showing interdependency of actors (or nodes) in terms of the social relationships such as friendship, kinship, or financial exchange [33] [34]. Graphs represent objects where order and disorder coexist. Graphs serve well for showing social interactions, influence, migration paths, and framing effects. For instance, a graph may expose which news sources hold central positions that function as points of prominence, control, and stability and may show edges that act as highways for lead relationships and exchanges of news frames. These intertwined dynamics, coupled with the vast amounts of online news being produced daily, make graphs an important tool for visualization and analysis of information flows pertaining to critical issues.

The overarching research problem is expressed through the construction of an issues network, and the model of a news frame issues network is presented in this paper as a graph $G = ({S_g}^t, {A_h}^t, {E_r}^t)$ consisting of (1) S_g that denotes a set of source nodes containing $g = \{s_1, s_2, \cdots, s_g\}$ elements and (2) A_h as a set of article nodes containing $h = \{a_1, a_2, \cdots, a_h\}$ elements that operate within the time window$t$. In general, t indicates one discrete time step of a reporting period. Let $S_g = \{C_i, F_i\}$ be the set of source properties, where (1) C_i is the set of terms (i.e., sub-issue keywords or content descriptors) and (2) F_i is the dominant frame choice (identity) that the news source uses. Each relation has a corresponding set of edge connections, ${E_r}^t$, directed/undirected as edge elements. The issues network subscript $r = 3$, corresponds to the total number of relations. The edge set represents the communication channels between node pairs. A_h are children of S_g; therefore, they may inherit the properties of the parent news source, as shown in Fig. 4.

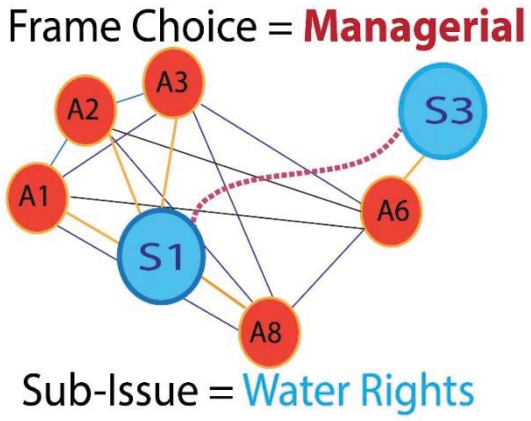


Fig. 4. A subset of the news frame issues network with two news source nodes with connecting articles using the same sub-issue and the same frame choice. The blue circles around notation S_g show online news sources node that produced the article and propagates a central organizing idea (i.e., framing), orange circles around A_h show an online news articles node, black lines show edge connections among neighboring articles on the similar issue, golden lines show edge connections of news sources who produce articles covering similar issue, and the red line shows a managerial news frame.

The relations (or rules) for E_r edge connections are as follows:

$\boldsymbol{R_1}$. A & A (articles to articles). A non-directional edge connection is constructed for E_1 when $A_h(a) = \{a: (a_1, a_2, \cdots, a_h) \in E_1\}$ is the set of neighbors of $\boldsymbol{a_h}$ such that $\boldsymbol{a_h}$ represents articles with a similar sub-issue.

$\boldsymbol{R_2}$. A & S (articles produced by the same source). A directional edge connection is constructed for E_2 when $S_g \in E_2 \rightarrow A_h(a) = \{a: (a_1, a_2, \cdots, a_h) \in E_2\}$ is the source (or producer) of the set of neighbors of $\boldsymbol{a_h}$ articles, where $\boldsymbol{a_h}$ represents articles with similar sub-issues.

$\boldsymbol{R_3}$. S & F (sources with sub-issue to sub-issue with the same frame choice). A non-directional edge connection is constructed for E_3 when $S_g(s) = \{s: (s_1, s_2, \cdots, s_g) \in E_3\}$ is the set of neighbors of $\boldsymbol{s_g}$ such that $\boldsymbol{s_g}$ shares the same dominant frame choice.

Problem: Given $G = ({S_g}^t, {A_h}^t, {E_r}^t)$ about I issue, construct a graph for determining how frames are produced in online news media over time.

D. Model Setup

To model the news frame issues network, the authors of this paper propose formulating a new model for information feedback based on a social system. The new framework is within the tradition of previous frameworks. These are the previous models [1] [7] [28] [38] for studying the evolution of personal position, social influence multi-dimensions, and forces that cause cultural shifts. Axelrod's model [7] is built on two simple assumptions: (1) selection, a phenomenon in which people are more likely to interact with those who are

978-1-5090-6435-9/17/$31.00 ©2017 IEEE

more similar to them, and/or to be more receptive to influence from those who are similar and (2) influence, a term that describes the case in which interactions tend to cause similarity among interacting actors. The model proposed in the present study promotes the additional emphasis of bias that emerges as double subjectivity in narrative text. Although the behavior of the proposed model is similar to that in Axelrod's [7], when deciding how to interact in the network, the proposed model differs in that insight is given into the endogenous and exogenous forces that may be at play in dynamic systems. Therefore, the double subjectivity in the proposed social influence model integrates cultural dynamics extensions [28] with the cusp catastrophe model [38]. The cusp catastrophe model is useful for describing non-linear relationships such as those found in narrative text. In the proposed model, the particular focus is on the different forces at work in the bias slant employed in news frames as it allows for the degrees of freedom whereby alternative pathways may be realized. The cusp catastrophe model allows for understanding the different forces at work when quantifying bias and the migration path of bias formation.

Conceptually, the switch from linear to non-linear relationships involves taking into account not only a news source's capacity to select (or affect), but also another news source's capacity to be influenced (or affected). Thus, an important aspect of selection and influence is the characterization of dynamical systems not just by their properties, but also by their capacities.

Consider this contextual example. When a news event occur that impacts society, this thereby becomes an issue. It is partly defined by its properties, such as the absolute baseline facts, as well as being in a certain state, like the event having taken place on a certain day. This same news framing, however, has the capacity to intensify its bias that may have the causal effect of a revolt or eruption by the people, an exacerbation of sentiment about a group of people, or the viral sensation of the news frame. This can happen through interactions with news sources that have the capacity to exercise strong bias and to simultaneously weaken the agency of the people, group, or opposing view.

Properties are always reality based, since at any given point in time the facts—absolute baseline—are either true or false. Hence, facts are logical, formal, and rigid, but the causal capacity of news when exercising double subjectivity has a cumulative effect, which acts as a fabrication of what is real without being actual (or absolute). Unlike the limitations that facts in online news offer, bias intensity within the construct of double subjectivity allows for the freedom to change dimensions through the parameterization of different forces—frame identity, justice, and bias intensity—and distributions of sudden changes. The end effect is the perception of reality about an issue in online news that may offer new structure formations (or pathways) that lead to alternative views for shaping attitudes and beliefs. Stated differently, that which was only potential becomes a reality when changes in the different forces reach a certain critical threshold and a sudden jump happens. At this point, changes in beliefs can be tracked with the proposed model, even if the news framing bias mechanism morphs the facts into something far from the truth.

The cusp catastrophe model calls this possibility space "the response surface space and control space," which represents all possible states in the system. In this research, the relationship among frame identity, justice (or a sense of fairness), and bias intensity are acknowledged through the cusp catastrophe differential calculus formula, which allows the measurement of the rapidity or slowness with which forces can change. In the geometric approach to the calculus, each degree of freedom becomes one dimension of a possibility space. The space of possible states in the system allows for the differential relations between them to determine a certain distribution of stable states around points or loops of attractions. They have a fractal dimension (intermediate between one and two) and are referred to as "chaotic attractors" [37]. Thom's cusp catastrophe model [38] shows specific transitions of the forces, which have the tendency to cycle through the same set of states over and over.

The integration of cultural dynamics and the cusp catastrophe model for defining the new double subjectivity social influence model may shed insight into deep levels of expressivity in narrative text where bias offers alternative pathways and possibilities. This remains an emerging research topic for understanding the production of language that may help in synthesizing vast amounts of unstructured text and learning online social behaviors.

The authors of this study explore the process that news sources use when producing online news articles. The main factors considered in this production are the facts, bias intensity, and signals that derive from neighboring news sources. The signals that derive from news sources are likened to the reporter-relative or a floating baseline as mentioned previously. In the process of time, the news sources' absolute baselines will undergo adjustments as the bias intensity changes.

E. Double Subjectivity Social Influence Model

The proposed double subjectivity social influence model integrates cultural dynamics extensions [7] [28] with the cusp catastrophe model [38] [40] for modeling double subjectivity in narrative text. The cusp catastrophe model is useful for describing non-linear relationships such as those found in narrative text. The cusp catastrophe model allows for understanding the forces at work when quantifying bias and the migration path of bias formation.

Online news sources tend to produce articles about an issue I of interest to society, such as water insecurity, the state of the economy, or affordability of health care. Additionally, news sources may receive signals β from other news sources on the importance of the issue, based on observed increase to the number of articles other neighboring news sources produce. They receive facts f about an issue. Each news source, then, selects a bias strategy α using as the basis of its calculation the bias intensity. More specifically, each news source S_g with interest in I issue at time t generates news articles.

$$I(\mathrm{g}) = \alpha_g + \beta_g(f) \quad (1)$$

978-1-5090-6435-9/17/$31.00 ©2017 IEEE

Where, α_g and β_g are the coefficients of the news source while f are the facts received. Furthermore, it is supposed that α_g corresponds to the endogenous propensity of the news source nodes to express its own bias (or double subjectivity), while β_g represents the exogenous force in the network. In the first interaction, set $\alpha_g = 0$, denotes bias intensity; this value will be calculated upon interacting with neighboring nodes and should change to give shape to the news source's absolute baseline. Time evolves in discrete steps t = 0, 1, 2, …, and $S_g(t)$ denotes the mass count on S_g at time t. The maximization of $\{S_g\}$ gives the sub-issue most important, M_i— a count of the largest cluster (the mass) of nodes reporting on a water sub-issue. This initialization of α_g and β_g provides access to new nodes to enter the overall conversation about water consequences by linking to the most prominent mass, thereby, leveraging facts and the neighboring nodes of the relative baseline. This is similar to the operation observed in the hidden Markov model (HMM), whereby the news sources' absolute baseline is hidden with adjustments made over time. In contrast, the relative baseline is likened to the observations one can make in the HMM for getting a sense of the state [13].

Problem: Given $\{S_g\}$ interactions within graph G_t over time, determine the migration path and the non-linear, causal linkages about $I(\mathrm{g})$.

At the start of the process, each news source $g \in S$ has a nonnegative node mass associated with it, corresponding to the fraction of the news sources that initially reports (i.e. through signaling) on similar sub-issues about water insecurity. The dynamical system allows for each news source to switch frame choice, $F(e)$, and sub-issue interest, thereby enabling random selection of news source interaction. Also, news sources are susceptible to being influenced when they interact with neighboring news sources that share similar bias intensity. The full state space may be calculated by counting the news sources as expressed as the mass vector $S_g(t)$.

F. Defining Bias Intensity for Expressing Double Subjectivity

The quantification of bias intensity is situated in terms of the mathematical catastrophe theory [38]. Catastrophe theory is a branch of non-linear dynamic systems theory that originated with the work of the mathematician Rene´ Thom [38] to help explain biological morphogenesis as one of the great mysteries confronting mathematical biology. A key property in catastrophe theory is that the system under study is driven toward an equilibrium through its use of gradient descent or potential function for seemingly automatic guidance (i.e. through the law of attraction) occurring in the system, which is important in research on social influence, particularly when considering the property of convergence and stable states. The cusp model is the most well-known and the simplest model of catastrophe theory, positing that non-linear transition within a system from one state to another is guided by two controlling variables: asymmetry and the bifurcation factor.

Consider, for example, the social and economic inequality movement Occupy Wall Street. What began as small grassroots pressure points to address pay inequality among fast food workers and Walmart employees has transitioned to giving voice to 99% of the U.S. population and moving the political conversation in the U.S. election. This transition, from a grassroots protest in Manhattan's Zuccotti Park to making inequality and the wealth gap the core of the political race, is a catastrophe in accordance with catastrophe theory.

Let $M = \{Managerial, Economic, Conflict\}$ and $S = \{Human\ Interest, Science\}$ be the aggregate count of the sets of frame choices obtained from the edge connection. S_g is assigned the frame identity corresponding to its F_i property at time t.

Definition: Bias is represented by, a projection on the behavior surface for predicting the migration path that shows the gradual shift in bias intensity of the online news source when shaping the frame narrative about the issue of water insecurity. This factor is a measure of the strength of bias, that is, "weak bias" is considered safe and is represented by *S(ecure)*, and "strong bias" is considered insecure or harmful and is represented by *I(nsecure)*.

Definition: Asymmetry control (or normal) factor is represented by $\boldsymbol{x}$. This factor receives the label *Justice*, denoted on a scale that ranges from $J(ust)\ to\ U(njust)$. Justice is a measure of the news source's perception of fairness of water consequences.

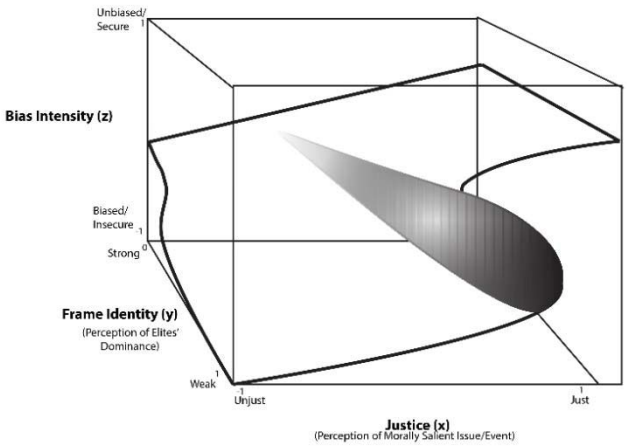


Fig. 5. The image depicts the cusp catastrophe model for measuring the shift in news sources bias intensity.

Definition: Splitting factor or bifurcation factor is represented by $\boldsymbol{y}$. This factor receives the label *Frame Identity*, denoted as $S(trong)$ or $W(eak)$. This is the perceived dominance by elites when reporting the news.

The authors, here, argue that *Frame Identity* and the perception of *Justice* are key forces for predicting the bias intensity of news sources when producing articles. *Frame Identity* is one form of bias; it is shaped by strategic devices for presenting prominent aspects and perspectives about an issue using a strong slant for the purpose of conveying latent meanings about an issue [30]. The position of the authors is that *Justice* is of equal importance, as it captures the perceived

978-1-5090-6435-9/17/$31.00 ©2017 IEEE

sense of fairness about an issue. Fig. 5 provides a depiction of the news source bias intensity in accordance with the cusp catastrophe model for predicting gradual shifts of views and showing the possibility space. This allows one to visualize discrete points in time over the possibility space. A discretized approach to opinion research has been conducted in scholarly research [40] and applications of cusp catastrophe [41] [42].

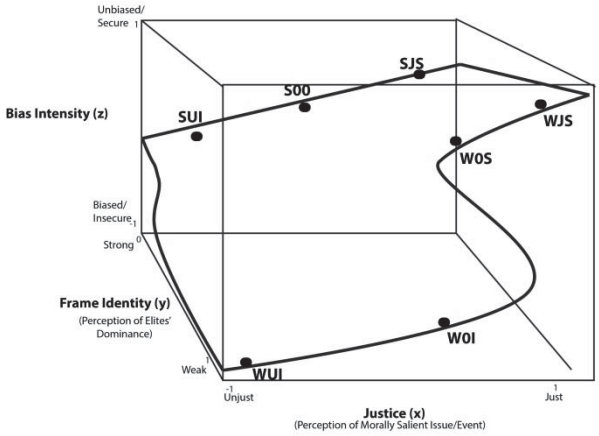


Fig. 6. This is a depiction of the state in the possibility space (surface space).

VI. Possibility Space Transitions

Non-linear relationships represent a variety of possibilities and pathways of which the linear case is a limiting one. The authors propose to express the possibility space for the study using three variables as shown in Table 1.

Index-1, Index-2, and Index-3 are combined to denote state transitions for updating shrinkages and growth. For instance, SJS denotes strong frame identity (S), a sense of justness (J), and a sense of security or being safe (S). Fig. 6 provides an illustration of the states on the cusp catastrophe surface space.

News sources' selection and influence occur through signaling news message bias intensity. Table 2 indicates where news sources change bias intensity based on interactions. Each column shows location of growth in the number of recipients switching from one news source to another, as a result of observed bias intensity. The shrinkage and growth at a single site are proportional to the number of recipients and the number of senders signaling. Table 2 illustrates the shrinkages and growth that may occur over time.

TABLE I. Possibility Space for Migrating the Network

FRAME IDENTITY (SPLITTING FACTOR)		
INDEX-1	**DESCRIPTION**	**VALUES**
S	S(trong)	0
W	W(eak)	1
JUSTICE (NORMAL FACTOR)		
INDEX-2	**DESCRIPTION**	**VALUE**
J	J(ust)	1
U	U(njust)	-1
0	N(eutral)	0
BIAS INTENSITY		
INDEX-3	**DESCRIPTION**	**VALUE**
S	S(ecure)	1
I	I(nsecure)	-1
0	N(eutral)	0

Note. Explanation of Index meanings: Index 1 represents the control variable frame identity, which is the perceived dominance by elites when reporting the news. Index 2 represents the sense of justice in the handling of the issue. Index 3 represents the dependent variable bias intensity, which has the effect of signaling a secure (or safe) or insecure (or harmful) issue.

TABLE II. Possibility Space for Migrating the Network

	Signaling content (state of news sender)						
Recipient	**SUI**	**S00**	**SJS**	**WUI**	**W0I**	**W0S**	**WJS**
SUI			S00			WUI	W0I
S00	SUI		SJS				
SJS	S00			WOS	WJS		
WUI	SUI		W0I				W0I
W0I	S00						
W0S			S00				
WJS	W0S		W0S	WOS			

VII. Observations of the Models

From the search of articles for keywords, 55,000 news articles were collected and stored in a database for pre-processing. After pre-processing, 30,000 articles were deemed relevant for the dataset; 280 news sources were identified using strategic framing, to potentially influence public perceptions of the issues of water insecurity.

A. Influence Dynamics and Migration Path

In this paper, models of cultural dynamics are integrated with the cusp catastrophe model to explore double subjectivity

978-1-5090-6435-9/17/$31.00 ©2017 IEEE

through quantifying the bias intensity. The analysis shows that this measure has potential to change the network structure, as prominent mass count is not the only law of attraction at work.

Although the proposed model is presented to scale to full dataset capacity, a small subset of the network is used for this preliminary evaluation. The model allows for news sources that generate articles about water consequences to interact freely with other news sources. However, because news sources receive signals that provide a relative baseline—leveraging other news sources' opinions, interpretations, and perceptions—it is most likely that a news source will choose interactions with similar sources. A chance exists for a news source to interact with another news source based on an increase in its bias intensity although no connection exists. However, the chance is higher that news sources will interact only with connected sources; thus, breaking out of the cluster is caused by catastrophic shifts in bias.

With the bias intensity function, a variant of the cusp catastrophe model shows possibility space in behavior when making decisions about trust and distrust. Any space on the surface response represents the state of the observed news source. As the news source perceives that the issue will not lead to fair water consequences, the cusp triggers them to start attributing blame toward the population causing harm. The contextualization of this case with water insecurity shows that a news source with ties to nodes that employ the frame choice of perceived unfairness (unjust) will notice that the sources more aligned with elites are becoming increasingly bullish on the water issue. These news sources who initially framed the issue as unjust attract other news sources (i.e. homophily) because their narrative is spreading. When feelings of unfairness intersect with the dominance of the elites, the news sources start scanning for those to whom to attribute blame, thereby entering a space of distrust and strong bias. This kind of model is associated with hysteresis, making it hard to shift between surface response planes, as the migration path will be different.

VIII. Conclusion and Open Questions

In this paper, a new concept was presented that allows for an adjustment of the absolute baseline of a news source (or agent) that takes into account one's bias for expressing double subjectivity through news frames. Further, offered in this paper is (1) the first known formal news frame issues network in computer science, (2) a model for learning the migration paths and patterns about issues, and (3) the first known formulation of a bias intensity function using the cusp catastrophe model for showing the possibility space and gradual shifts in views that are sensitive to news framing bias. Preliminary experiments suggest that the integration of cultural dynamic models with a cusp catastrophe model is promising for exploring double subjectivity for revealing latent relationships found in online news articles. Future work will involve testing other hypotheses and unfolding the potential function to explore conditions of convergence. Also, future work will give treatment to other hypotheses associated with bias that leads to trust or distrust, as a result of perceived positive water consequences.

References

[1] P. Clifford and A. Sudbury. A model for spatial conflict. Biometrika, 60(3):581{588, 1973.

[2] Eli Ben-Naim, Paul Krapivsky, and Sidney Redner. Bifurcations and patterns in compromise processes. Physica D, 183, 2003.

[3] Claudio Castellano, Santo Fortunato, and Vittorio Loreto. Statistical physics of social dynamics. Reviews of Modern Physics, 81:591–646, 2009.

[4] Guillaume Deffuant, David Neau, Frederic Amblard, and G´erard Weisbuch. Mixing beliefs among interacting agents. Advances in Complex Systems, 3:87–98, 2000.

[5] Rainer Hegselmann and Ulrich Krause. Opinion dynamics and bounded confidence: Models, analysis and simulation. Journal of Artificial Societies and Social Simulation, 5(3), 2002.

[6] B. Latane′. The psychology of social impact. American psychologist, 36(4):343, 1981.

[7] R Axelrod. The dissemination of culture - A model with local convergence and global polarization. Journal of Confict Resolution, 41(2):203{226, APR 1997. ISSN 0022-0027. doi: f10.1177/0022002797041002001g.

[8] K. C. Chen, M. Chiang, H. V. Poor, "From technological networks to social networks," published in IEEE J. Sel. Areas Commun.., 2013.

[9] Y. F. Chou, R. G. Cheng, and H. H. Huang, "Performance evaluation of referral selection algorithms in human centric communication networks,"in Proc. 2012 International Symposium on Wireless Personal Multimedia Communications, pp. 638–642.

[10] Miller McPherson, Lynn Smith-Lovin, and James M. Cook. Birds of a feather: Homophily in social networks. Annual Review of Sociology, 27:415–444, 2001.

[11] Mullainathan, Sendhil, and Andrei Shleifer. 2005. "The Market for News." American Economic Review 95 (1): 1031-1053.

[12] Thomas A. DiPrete and Gregory M. Eirich, Cumulative Advantage as a Mechanism for Inequality: A Review of Theoretical and Empirical Developments, Annual Review of Sociology, Vol. 32: 271-297 (Volume publication date August 2006)

[13] L. R. Rabinerand B. H. Juang,"An introduction to hidden Markovmodels," IEEE ASSP Mag., vol. 3, no. l, pp. 4-16, 1986.

[14] P. Raghavan, S. Amer-Yahia and L. Gravano eds., "Structure in Text: Extraction and Exploitation", In. Proceeding of the 7th international Workshop on the Web and Databases(WebDB), ACM SIGMOD/PODS 2004, ACM Press, Vol 67, 2004.

[15] H. P. Luhn. "The Automatic Creation of Literature Abstracts," IBM Journal of Research and Development, 2, No. 2, 159 (April 1958).

[16] Salton, G., A. Wong, and C. S. Yang. "A vector space model for automatic indexing." In Communications of the ACM 18 (11), 1975: 613-620.

[17] Steinheiser, R., and C. Clifton. "Data Mining on Text." 2012 IEEE 36th Annual Computer Software and Applications Conference IEEE Computer Society, 1998:630.

[18] A. Tan, N. Zhong and L. Zhou, "Text Mining: The state of the art and the challenges." Proceedings of the PAKDD Workshop on Knowledge Discovery and Data Mining, 65—70, Beijing, China 2000.

[19] S. Tan, X. Cheng, B. Wang, H. Xu, M. M. Ghanem, Y. Guo, "Using dragpushing to refine centroid text classifiers", In Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval (SIGIR '05). ACM, New York, NY, USA, 653-654. 2005.

[20] M. Tsytsarau & T. Palpanas, Survey on mining subjective data on the web, Data Mining and Knowledge Discovery, 24(3), 478-514, 2012.

[21] H. Karanikas, C. Tjortjis, and B. Theodoulidis. "An Approach to Text Mining using Information Extraction", Proc. Workshop Knowledge Management Theory Applications (KMTA 00), 2000.

[22] Q. Hu, D. Yu, Y. Duan, & W. Bao, A novel weighting formula and feature selection for text classification based on rough set theory, in Proceedings of the 2003 International Conference on Natural Language Processing and Knowledge En C. Jiang, F. Coenen, R. Sanderson, M. Zito, "Text Classification Using Graph Mining-Based Feature

978-1-5090-6435-9/17/$31.00 ©2017 IEEE

Extraction", Knowledge-Based Systems, vol. 23, no. 4, pp. 302-308, 2010.gineering (pp. 638-645), 2003. IEEE.

[23] B. Liu, Sentiment analysis and opinion mining, Synthesis lectures on human language technologies, 5(1), 1-167, 2012.

[24] D. A. Scheufele, "Framing as a theory of media effects", The Journal of Communication, 49(1), 103-122, 1999.

[25] S. Dumais, G.W. Furnas, T.K. Landauer, S. Deerwester, Using latent semantic analysis to improve information retrieval. In Proceedings of CHI'88: Conference on Human Factors in Computing, New York: ACM, 281-285, 1988.

[26] C. Fellbaum, "A Semantic Network of English: The Mother of all WordNets," in: Computers and the Humanities 32: 209-220. 1998.

[27] D. Bindela, J. Kleinberg, & S. Orenb, How bad is forming your own opinion?, Games and Economic Behavior, 92, 248–265, 2015.

[28] D. Kempe, J. Kleinberg, and ´ E. Tardos, "Maximizing the spread of influence through a social network," in Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, pp. 137–146, 2003.

[29] M. Lombard, J. Snyder-Duch, C. C. Bracken, Content analysis in mass communication: Assessment and reporting of intercoder reliability. Human Communication Research, 28, 587–604, 2002.

[30] R. M. Entman, "Framing: Towards clarification of a fractured paradigm", Journal of Communication, 43(4), 51-58, 1993.

[31] G.F. Bishop, R.W. Oldendick, A.J.Tuchfarber,. Effects of presenting one versus two sides of an issue in survey questions. The Public Opinion Quarterly 46, 69–85, 1982.

[32] J.A., Krosnick, Survey research. Annu. Rev. Psychol. 50, 537–567, 1999

[33] K. C. Chen, M. Chiang, H. V. Poor, "From technological networks to social networks," accepted and to be published in IEEE J. Sel. Areas Commun.., 2013

[34] Y. F. Chou, R. G. Cheng, and H. H. Huang, "Performance evaluation of referral selection algorithms in human centric communication networks,"in Proc. 2012 International Symposium on Wireless Personal Multimedia Communications, pp. 638–642.

[35] D. P. Cartwright, "Analysis of qualitative material", In L. Festinger & D. Katz (Eds.), Research Methods in the Behavioral Sciences (pp. 421-470). New York: Dryden, 1953.

[36] R. K.Garrett, B. E. Weeks, R. L. Neo, Driving a Wedge Between Evidence and Beliefs: How Online Ideological News Exposure Promotes Political Misperceptions, Journal of Computer-Mediated Communication, 4 August 2016

[37] I. Stewart, Does God Play Dice?: The Mathematics of Chaos, Basil Blackwell (Oxford), 1989, pp 84–94

[38] Thom, Rene´. Structural Stability and Morphogenesis, 1972. D. H. Fowler (trans.). Reading, MA: W.A. Benjamin.

[39] Skitka L, Sargis E, The internet as psychological laboratory. Annu Rev Psychol 57: 529–555, 2006.

[40] Sobkowicz, P. (2012). Discrete model of opinion changes using knowledge and emotions as control variables. PloS one, 7(9), e44489.

[41] Homer-Dixon, T. (1999). Environment, Scarcity, and Violence. Princeton University Press.

[42] Homer-Dixon, T. F. (2006). The upside of down: Catastrophe, creativity, and the renewal of civilization. Washington: Island Press.

978-1-5090-6435-9/17/$31.00 ©2017 IEEE