



Save to disk



Hide/Show menus

⌘ { Measuring the Physical Properties of Memes } ⌘

A PHYSICS OF IDEAS

“One of the main obstacles to productivity today is the growing problem of information overload. Information overload results because we lack effective tools for automatically organizing information collections into meaningful and relevant chunks.” *continued* ▶



by Nova Spivack

This article is reprinted
with permission from
the author from:
<http://mindingtheplanet.net>

INTRODUCTION

This paper provides an overview of a new approach to measuring the physical properties of ideas as they move in real-time through information spaces and populations such as the Internet. It has applications to search, information filtering, personalization, ad targeting, knowledge discovery and text-mining, market research, trend analysis, intelligence gathering, organizational behavior and social/cultural studies.

One of the main obstacles to productivity today is the growing problem of information overload. Information overload results because we lack effective tools for automatically organizing information collections into meaningful and relevant chunks. For many years I have been thinking about a new way to approach this problem that is based on some ideas in classical physics. For quite some time I didn't say anything about it because it seemed like good material for a patent, but now I've decided it would be better to just put this in the public domain since I would rather have it be prior-art than patented by anyone. This is fundamental and useful and everyone should benefit from it.

In this article I propose the beginning of what might be called "a physics of ideas." My approach is based on mapping classical physics to memes that move through information spaces over time. The key to this is to measure the momentum of ideas as they move through space and time, and thus the momentums of documents that contain them. This provides a means to quantify the strength and trajectory of ideas as they move through a given corpus (and by inference, through the populations that create and consume the documents in that corpus) — and this enables us to start applying classical physics to empirically measure and understand the dynamics of

ideas that are shaping our world. In other words, we can start to objectively analyze interactions between ideas as well as the impact that various ideas have on people, organizations and events in the "real world" and in turn the impact that those things have back on ideas.

Ideas are perhaps the single most powerful force shaping our world today, other than the climate. Humanity's behaviors are nothing but the results of various ideas — the phenotype of the ideas that are actually at work in the population at a given time. This is true for any organization, and even for individuals — ultimately much if not all of their behavior is conditioned by their ideas. So if we can form a science of ideas that enables us to begin to quantify and analyze their dynamics, we can start to gain new insights into human behavior and the "hidden" forces shaping our world today. My approach is to find a way to map what is going on in the realm of ideas to existing methods in classical physics — I want to make it possible to treat ideas as ideal particles in a Newtonian universe. It may then be possible to use the wealth of techniques that physicists have developed for analyzing the dynamics of particle systems to understand the dynamics of ideas within and between individuals and groups.

BACKGROUND

But first some background about how I came up with this idea...In 1993 I worked as an analyst at Individual, Inc., back in the pre-Web days. In that job I was part of a sophisticated information filter. Individual published filtered personalized newsfeeds. They aggregated content from thousands of sources and then filtered it into strategic newsfeeds tailored to the interests of their customers. You may have used Newpage

or Heads Up, in the past. Chances are, if you did, I was your analyst. The way that the Individual system worked was that first a set of AI agents did a first pass on the incoming content to sort it into buckets. These buckets were routed to a team of human analysts with expertise in the relevant fields. The analysts would then go through the articles in the buckets to prioritize them, remove duplicates or items that had come through in previous articles as well as items that did not belong, and add in any items that should be included.

I want to make it possible to **treat ideas** as **ideal particles** in a **Newtonian universe**.

What this meant in practical terms for me as an analyst was that every night from about 8 PM until 1 AM I had to personally read through around 1600 news articles. My beat was emerging technology, software, broadband, online-services, multimedia and satellite applications, so I enjoyed it (yes, I am a serious news hound!). But still it was a challenge to keep on top of such a fire hose. Not only did I have to figure out what was important and how to prioritize it, but I also had to remember if I had ever seen and published anything about a given subject before in the previous year. By trial and error I evolved a solution to this problem. In summary, what I realized was that whether or not something is relevant is much more subtle than merely keyword matching!

A good example can be found in nature — specifically frogs. Frogs have interesting visual systems. They are tuned to focus on things that move. They are most sensitive to size and velocity, but they also notice changes in velocity. Things that are small and that

don't move are not of particular interest to them. Things that move in erratic ways are most interesting. Humans are slightly more sophisticated — we notice momentum, a measure of the "mass" or "size" of things and the way they change over time.

Every night from about 8PM until 1AM, I had to personally
read through around **1600 news articles.**

What I realized at Individual back in 1993 was that the way I figured out what articles to prioritize was not so different from how a frog finds flies to eat — but more sophisticated. I realized that I filter information according to the momentums of ideas (or "memes") — how the various memes in the articles I was reading were moving through space and time in the culture I lived in and the communities I was interested in.

Here is a formula that provides an overview of my heuristic for filtering news articles:

Rank of item (such as an article) in a list = function of (keyword relevancy of item to query, reputation of publisher or source, date of publication, relevancy of source to query, 'document momentum' of item with respect to memes in corpus).

The last item, "document momentum" is the key to it all. The document momentum is a measure of the cumulative real-time momentum of the memes that occur in the document. This requires some explanation. The rest of this article describes how to measure the momentum of a meme and compute the document momentum for a document, and how this can be useful for a number of information tasks.

MEMES

First of all what is a meme (pronounced “meem”)? — basically it's any idea, any concept.

More formally, a decent definition of a meme is:

"/meem/ [coined on analogy with `gene' by Richard Dawkins] n. An idea considered as a {replicator}, esp. with the connotation that memes parasitize people into propagating them much as viruses do. Used esp. in the phrase `meme complex' denoting a group of mutually supporting memes that form an organized belief system, such as a religion. This lexicon is an (epidemiological) vector of the `hacker subculture' meme complex; each entry might be considered a meme. However, `meme' is often misused to mean `meme complex'. Use of the term connotes acceptance of the idea that in humans (and presumably other tool- and language-using sophonts) cultural evolution by selection of adaptive ideas has superseded biological evolution by selection of hereditary traits. Hackers find this idea congenial for tolerably obvious reasons." (Definition from: The Hacker's Dictionary)

Memes are essential to the way the human brain processes ideas and how it decides what is important. We are basically "meme processors" — we are "life-support systems for memes" to put it another way. To use a computer analogy, our physical bodies are like the hardware and operating system, and our minds — the dynamical activity and state of this hardware — are like the software applications and content running on the hardware. Our minds could be viewed as systems of interacting memes — complex systems of ideas that interact within us, and across our relationships.

Memes are capable of spreading across human social relationships, via human interactions, and via human interactions with static storage vehicles such as printed media, audio or video, and digital storage media — they are highly "communicable." (And

soon, as I have proposed in other articles, with the coming Semantic Web memes will be able to spread and interact without needing humans at all — machines will be able to process them on their own.)

Because memes spread and even appear to evolve over time there has been some discussion about using biological metaphors and ideas from population biology and evolutionary theory to analyze memes. I think these techniques are definitely useful for doing macro-memetics. What I am proposing here is a way to start doing micro-memetics — to understand what is happening precisely for each meme at any level of granularity.

WHAT MAKES A MEME "HOT"?

One of the many interesting skills that humans have but that computers are so far not able to replicate very well is the ability to "intuitively" figure out what is important in a complex set of information — we humans are excellent at detecting patterns in complex information. We are good at tracking and measuring memes. But more significantly we are excellent at picking out the most important memes, even in highly cluttered, chaotic and noisy environments. We do this naturally all the time. But how?

Our minds experience thousands of memes every day — But how exactly do we decide which memes are more important at a given time, relative to all the other memes? For example, how do we figure out that something is a "hot" trend? We may hear it mentioned casually in conversation, read it in an article, see it on the Web, etc. But what determines whether we decide it is important and "hot?" It is not merely the

total number of times that memes related to it have been mentioned (what we might call the "mass" of the meme) — that just tells you how "big" the memes are but not how "hot" or "forceful" they are relative to other memes. Furthermore "hotness" is not merely the velocity or frequency of new mentions per unit time — while this is useful, velocity alone doesn't say much about relative importance of memes. Acceleration of memes — the change in velocity — is also useful but not enough.

What we really need is a method that relates the "size" of a meme to the way it moves in space and time. In other words, in order to determine whether a meme is "hot" we need a way to measure its 'meme momentum' — we need a physics of ideas.

THE MEDIA IS THE MIRROR: IDENTIFYING THE REFLECTIONS OF MEMES

Before we can measure the physical properties of memes, we need a way to identify the memes we are interested in analyzing. We can identify memes by analyzing textual media such as document collections, wire services, and the Web.

THE MEMES WITHIN TEXT APPEAR TO BE DORMANT — they are frozen digital representations of memes. They do not move or reproduce on their own — they need help from humans (for the moment). But by inference, static textual representations of memes provide a mirror of the actual "live memes" that are taking place in the minds of the people who author and consume that media. By analyzing textual media we are not merely looking at the memetic properties of text, we are looking at the memetic properties of people's minds and of organizations, societies and cultures.

THE MEDIA IS THE MIRROR — the Media is a mirror of the mind and its culture. By analyzing suitably selected information sources (for example, "all news articles from USA newspapers") we can effectively focus on a reflection of the memes that are actually present within the minds of humans in a particular place, time, industry, community, demographic, etc. The more we know about the information sources, the more we can infer about the memes we find, and thus the memes taking place within the minds of the people who interact with those information sources.

We really need **a method** that **relates** the *size* of a meme to the **way it moves** in **space** and **time**.

The simplest approach to identifying memes in textual media is to simply pre-specify a list of memes we are interested in and to then search for any matching strings. For example we might be interested in measuring memes related to a particular trend, such as "Java technology," so we could compile a list of terms related to Java and then use search techniques to locate all instances of those terms. We can then measure their properties.

A more sophisticated approach than specifying interesting memes in advance is to discover them empirically by analyzing text to see what's there. To do this we might automatically identify nouns or noun-phrases and then measure their dynamics to see whether they are interesting enough to warrant further analysis. There are many existing computational linguistics techniques for isolating parts of speech and linguistic expressions.

Each of these nouns or phrases is a potential meme (we may consider them to all be actual memes or we may filter for only those memes that exhibit dynamics in space and time that meet our threshold for what constitutes "interesting" or "memelike" behavior. Another, more brute-force approach, would be to simply analyze every noun and phrase in a document or corpus for any that exhibit "memelike" dynamics in order to discover memes empirically instead of specifying them and then gathering their stats.

We can use various **standard methods** from text-mining and natural language processing to do a smarter job of **identifying memes**.

We can use various standard methods from text-mining and natural language processing to do a smarter job of identifying memes (for example, we can use stemming to consolidate various forms of the same word, we can use translation to consolidate expressions of the same meme in different languages, and we can use conceptual clustering and even ontologies to consolidate different memes that are equivalent to the same underlying meme). But for now, we can start by identifying memes in a simple way — the same way we might identify "topics" or "keywords" in a document. Once we can do this we can then measure the physical properties of those memes as they move through time and various spaces of interest.

(Note: We don't necessarily have to analyze every document in a corpus to gather valid statistics for memes within it. We can use random sampling techniques for arbitrary

degrees of accuracy if we wish to optimize for faster results and less computation. Instead of analyzing every occurrence of each meme, we can analyze a statistically valid sample of the corpus.)

A PHYSICS OF IDEAS

I suggest that the physics of ideas will be quite similar, if not equivalent to, the physics of the natural world. Everything in the universe emerges from the same underlying laws, even memes. The intellectual processes taking place within our own minds, as well as across our relationships and social organizations are similar to the dynamics of particle systems, fluid flows, gasses, and galaxies. We should therefore be able to map existing physical knowledge to the memescape, the dimension of memes.

Here are a set of basic measurements of the physical properties of memes and documents:

- » Absolute meme mass = how "large" the meme is = the total number of times it is mentioned in the corpus since the beginning of time to the present. (Note: When computing meme mass, we can choose to count repeat mentions or ignore them — doing so has slightly different effects on the algorithm. We can also, if we wish, get more fancy and look at clusters of memes that relate to the same concepts in order to compute "concept-cluster momentums" but that is not required.)
- » Absolute meme velocity = how fast the meme is moving in the corpus in the present time interval = The rate of occurrences (or "mentions") of the meme per unit time (minutes, hours, days, etc.) in the present time interval.

- » Absolute meme momentum = the force or importance of the meme in the corpus = the meme's absolute mass x the meme's absolute velocity.
- » Relative meme mass = the mass of a meme within a subset of documents or data in the corpus representing some set of interests. (Note: we call a subset of mutually co-relevant documents a "reference frame" or a "context.") such as a set of interests, a particular period in time, etc. (rather than in the entire corpus).
- » Relative meme velocity = the velocity of a meme within a reference frame.
- » Relative meme momentum = the relative meme mass X the relative meme velocity.

On the basis of these we can then compute derivatives such as:

- » Absolute meme acceleration = how the absolute meme velocity is changing in the entire corpus = The change in absolute velocity per unit time of the meme in the corpus.
- » Relative meme acceleration = the change in relative velocity of a meme.
- » Absolute meme impulse = the change in importance per unit time = the change in a meme's absolute momentum.
- » Relative meme impulse = the change of a meme's relative momentum.

Next, we use the above concepts to look at sets of memes, for example documents:

- » Absolute document momentum = the force or importance of a document in the entire corpus = the sum of the absolute momentums of each meme that occurs in the document. (Note: we may choose to count or ignore repeat occurrences of an article in different locations or at different times — this has different effects).
- » Relative document momentum = the force or importance of a document within a reference frame = the sum of the relative meme momentums in the document. This is a more contextually sensitive measure of document momentum — it couples momentum more tightly with a context, such as a particular query or time interval, or demographic segment. (Note: we may choose to count or ignore repeat occurrences of an article in different locations or at different times — this has different effects).
- » Hybrid document momentum = a measure of momentum that combines both relative and absolute measurements = either relative mass X absolute velocity or absolute mass X relative velocity.

HOW TO ANALYZE A CORPUS USING THESE METHODS

We can then apply the above measurements to entire corpora (collections of documents). This enables us to empirically rank the ideas occurring in the corpus in any interval of time. Furthermore it enables us to rank and prioritize documents in the corpus according to their momentums within any time interval — in other words, how representative they are of "important" or "timely" ideas within any time interval.

To do this, first we must create an index of stats for all memes we are interested in. We can use the above mentioned techniques for identifying memes to do this. For each meme we identify, we create a record in our index that lists the stats we find for it by source location and time. We then analyze our text sources and update the records in this table (for a historical analysis we do this all at once; for a real-time analysis we do it continuously on an ongoing basis or in batches). As new instances of memes are found we append the corresponding records in the index.

Looking at **the impulse** (the change in momentum)
we can also **chart the direction** of these trends
(increasing or decreasing).

We can now use these statistics to plot memes and documents according to our measurements of meme and document mass and velocity. This enables us to segment the memes or documents according to the various possible configurations of these dimensions. Each of these configurations has a useful meaning, for example a document with low absolute mass, moderate or high relative mass, high absolute velocity and high relative velocity contains "newly emerging trends of interest to the current context" whereas a document with high absolute mass, low relative mass, high absolute velocity and low relative velocity contains "established large trends that are not very relevant to the current context."

By looking at the impulse (the change in momentum) we can also chart the direction of these trends (increasing or decreasing). Memes that have high positive impulse are becoming more "important" than those with lower impulses. This enables us to determine whether memes are "heating up" or "cooling off" — a meme is heating up if it is important and timely and has positive impulse.

These **techniques** provide a way to **rank documents** that is in some respects like **Google's algorithm**, except that it **works for all types** of information.

Thus documents that have high document momentums contain memes that have high meme momentums — in other words they are representative of whatever ideas happen to be most important now. Tomorrow, when the momentums of various memes may have changed, the same documents might have different document momentums.

These techniques provide a way to rank documents that is in some respects like Google's algorithm, except that it works for all types of information — not just information that is highly interlinked with hotlinks or citations but even for flat text — and it is capable of arbitrary resolution in time and space. For example, Google is basically estimating document popularity — or effectively, endorsements implied by citations — for each query. Google determines the rank of a page in a set of results by estimating the community endorsement of that page as implied by the number of relevant pages that

link to it. Using the proposed physics of ideas however we can accomplish the same thing in a different and possibly better way — we can now compute the 'potential community value' of a document — without actually requiring links in order to figure that out. Instead, we can determine the relative strength of the ideas (the memes) that are present in the document and compare them to the memes that are present in the community of documents that are relevant to the keywords in our query.

For example, we do a query for "space tourism" and get back 6,830,000 documents in Google. Next we compute the above stats for each of those documents. We then rank the documents returned by this query according to their relative document momentums. This has the effect of ranking the documents according to the strengths of memes that are particularly of interest to the community represented by the query results. Thus it enables us to rank the resulting documents for our "space tourism" query to favor those documents that contain the highest momentum memes relative to set of memes that matter to the community — in other words the documents that contain ideas that are most "timely for the community" would appear higher. So this is a way to figure out not just what is relevant but what is important or in other words timely at a given point in time to people with a given set of interests.

EXAMPLE APPLICATIONS

Using the above techniques we can use momentum to provide a more sensitive way to filter any collection of information objects for which we can gather stats representing mass and velocity. There are numerous useful applications of doing this. Below I describe some of them.

Filtering E-Mail

For example, one might filter their e-mail using meme and document momentums in order to automatically view messages, people and topics with high momentum, low momentum, growing or declining momentum, etc. One could also use these techniques to data-mine the articles in a news feed or corpus for those that contain the "hottest trends." It could be used to automatically detect "emerging hot topics," "people to watch," "companies to watch," "products or brands to watch," etc. When ever you send a message the system measures the memes in that message and updates a special meme-stats index called "my interests" which just has the meme-stats for memes in messages you send. All incoming e-mail messages you receive can then be ranked according to their document momentums with respect to the meme momentums in the "my interests" index. This e-mail filter is automatically adaptive — as you send messages it learns what your current interest priorities are and this is reflected in changing meme momentums, even as your interests shift over time. These updated momentums are then used to filter incoming mail. So your mail filter learns what is important to you as you work and adapts to focus on your current priorities and interests, without you having to teach it. It just learns and adapts to model your current interests as you work.

Media Analysis

Beyond just that, these techniques can be used to perform more precise media analysis — for example they can be applied to measure the success of an advertising or marketing campaign by correlating the campaign placements with changes in momentum of the memes for the brand or product in the media.

Predicting Changes to a Stock Price

We can also use these techniques to make predictions — for example, we can correlate meme momentums for memes related to a company with technical properties of its financials and stock price and then make predictions about price changes by analyzing news articles to detect changing meme momentums related to the company. We can also do pure statistical correlations between meme momentums and stock momentums for example. The financial news media is like a mirror reflecting what is taking place in the markets — but investors also use this mirror to decide what to do in the markets. So by measuring what appears in this mirror we can predict what investors are likely to do next.

Prioritizing Search Results and Implicit Query Expansion

We can also use these techniques to prioritize Internet search results — or any search results for that matter. For example, a set of Web documents can be prioritized by their document momentums, such that those that represent the memes that are currently the hottest can score higher — in other words, documents that are currently more timely can score higher than those that are less timely, and documents that are more timely yet less relevant (on a keyword level) can be ranked higher than those that are less timely but more keyword-relevant.

For example, suppose you search for "Asian restaurant." If the meme "Vietnamese food" is currently in vogue in the media, meaning that it has higher momentum currently, then documents about restaurants that contain "Asian" or "restaurant" and that contain "Vietnamese food" will score higher than those that only mention "Asian" or

"restaurant" and "Chinese food" (assuming that Chinese food currently has a lower momentum). But this could change later as trends change. In other words, although we searched for "Asian food" we ended up getting documents ranked not merely by the keywords "Asian food" but by what topics related to Asian food have highest momentum today. This is a form of "implicit query expansion" and "implicit filtering." In other words the system can prioritize search results for you according to the present momentums or in other words, the timeliness, of memes that occur in them. So it can show you the documents that are likely to be most important to you NOW in light of current trends and events, versus just the documents that have the best keyword relevancy.

Market Research

To make things even more interesting, we can add additional arguments to our "Rank of item" function and our meme-stats table — for example, not just a measure of mentions but also a measure of "hits" — hits on a meme increase whenever a document containing the meme is viewed. We can also add another dimension to represent the spatial distribution of memes. This will enable us to track the vectors of memes through time and space. We can do this by associating each source (each publisher) with a geographic location. We then segment our meme-stats table by geography to break out the momentum of each meme in each geographic region. This enables us to do things like filter documents by "how important they are to people in New York."

By adding further dimensions — such as demographic profiles gleaned for example from the reader-surveys of publishers we can also segment by demographics, so we can even filter documents by "how important they are in the last month to

professional, Democratic party affiliated, college educated, women in New York City who earn a median household income of \$100,000."

The system can **prioritize search results** for you according to the **present momentums**.

By adding still one more dimension to measure "sentiment" for each mention of a meme (as a function of the positive or negative language occurring near it or better yet, about it), we can even start to rank memes according to the percent of members of a given population that support or oppose them. In other words, this system can be used to empirically measure what polls and focus groups do informally. The notion here is that by selecting media sources that are representative of the community you are interested in understanding; you can then view memes and meme data relative to that group. You can also do this in the other direction, simply look to discover what memes have interesting stats for the group your are interested in. Another use of this technology might be to analyze intellectual history by computing meme-stats from historical documents or past news articles.

We can also leverage the fact that meme dynamics can be correlated with those of other memes to determine dynamical dependencies amongst them. This enables us to determine that some memes positively or negatively reinforce others. It also enables us to discover sets of related memes — such that we can learn that stats on a given meme should be inherited by related "child memes" in an automatically or manually generated taxonomy of memes.

Measuring and Mapping Ideas in the Semantic Web

We could also reference metadata about the semantics of various memes. We can even filter for various types of memes — such as "memes related to vehicles" or "memes representing people" or "memes representing products ," etc. This enables us to start measuring ideas as they occur and interact on the emerging Semantic Web — but not just particular memes, even conceptual systems of memes that are interacting or somehow ontologically related. By linking with an ontology, for example, we can track the momentum of all memes related to "American cars" versus those for "German cars." The ontology enables inferences that help us find all memes that represent types of cars and classify them by nationality of manufacture.

Detecting Terrorist Activity

These techniques might even be used to detect signs of potential terrorism, and to "get inside the minds" of various people or groups of interest — simply analyze the meme-stats for memes in documents they create or view to automatically generate a profile of the main ideas currently occupying their minds. Next by tracking this over time you can start to plot trajectories and make predictions. Intelligent agents can then be trained to notice "interesting" patterns in these trajectories and alert analysts as needed.

Targeting Ads

The same methods could be used to better target advertisements or recommendations to users. Knowing what memes are currently most important to a party enables

better personalization and targeting. In this case a Web site could track what memes are hottest for a given user account — derived from what pages they view and what messages they write or respond to. This data could then be used to augment the users' interest profile with more dimensions of detail about each interest — such as how timely it is to the user, what particular nuances are specifically interesting, what their sentiment is. This could result in less irrelevance and spam for users and better results for marketers.

Knowledge Discovery

Now what gets interesting is the above methods can be used on both directions. We can use them to ask questions about memes we are interested in and we can also use to empirically discover memes we should be interested in within any corpus. So for example we can just empirically compute meme momentums and document momentums in any collection of information and then filter for whatever dynamics we are interested in, for example, "hot new emerging trends to watch."

A New Kind of Portal

Using these methods it is possible to build a new kind of portal that provides a window into the collective mind of the planet (or any community of interest). It would show what people within the desired segment think is important over time. We could watch an animation on it of how memes for "jihad" have spread, or for how those for a technology like "Java" have spread versus those for "Microsoft .Net," or how a particular war is currently viewed by the public in different states or different demographic

segments. A user could "drill down" into any meme to see its stats, all articles where it was mentioned, and related items on the Web, and maybe even products etc.

Open Questions & Directions for Further Research

It is important to note that these simple physical concepts could be taken much further. For example, using the above approach we should be able to determine the "gravity of a meme" or of a document or any set of memes or documents. We can then start to model the shape of memetic manifolds — the shape of space-time for ideas. We can also start to look at systems of memes as fields. Perhaps there may even be applications of fluid dynamics, relativity theory, or even quantum mechanics to what is taking place in the memescape — but today we are just taking baby-steps, just as Newton and the early natural philosophers did long before us. We need to begin to simply have the ability to measure memes and their basic interactions before we can go on to higher levels of analysis. I leave it to the physicists among us to take this to the next level of formalism — would anyone like to try their hand at formalizing the basic equations this physics of ideas?

There are a number of open questions I am still thinking about that suggest opportunities to refine these techniques. In particular, should we normalize documents somehow so that large documents don't have an unfair advantage over small documents (because large documents have more terms in them and thus have higher document momentums)?

Another question is whether or not we should rank documents first by relevance to query, and then within each "relevancy band" further rank by document momentum

within that band? This has the effect of limiting the impact of momentum versus relevancy — which may be useful if relevancy is considered to be more important. For example the top 100 most relevant documents are ranked by relevancy and then within that set they are ranked by document momentum and displayed, next the second 100 most relevant documents are ranked by relevancy and then within that set they are ranked by document momentum and displayed, etc.

[Is] there is an **ideal set** of priorities for the various **measurement dimensions** above with which **to rank** documents for **general searches**?

Another question is whether there is an ideal set of priorities for the various measurement dimensions above with which to rank documents for general searches. We can let users choose their own priorities of course, for example, by letting users set their priorities for various memetic dimensions, we can then tailor our ranking for their needs. Are they just looking for all documents that are relevant to a query, or are they really trying to find documents that are representative of the most timely issues relevant to a query? We might enable users to set their weights for the absolute and relative measurements of documents in order to view different rankings of search results. Better yet, we could simply provide them with natural language filters to apply, such as "Filter for documents that contain currently hot topics related to this query." In other words they can set priorities for the above dimensions in order to favor one

dimension over another — so they might decide that query relevance is most important, document mass is second and velocity is least important. This would translate to a constraint such that it would be more difficult for documents with low relevance to be ranked higher than documents with high relevancy just because they have higher momentums. On the other hand, they might want to favor momentums — for example if they really want to find documents that mention the latest trends related to a query — in which case we would favor document mass and/or velocity above document relevancy in our ranking. I am still thinking about the best way to handle these tradeoffs. Letting the user set their priorities is one way — but it may be possible to do a good job of satisfying most people with a particular set of default priorities. What is the best set of default priorities for general use?

Can we **determine the vector** of each meme at **each sector** in a **geographic map**?

There is also the question of how to best represent the "footprint of a meme" in geographic space. We can detect mentions of memes and using the above methods we may be able to associate each mention with a particular geography (the geographic region of the publisher and/or the intended audience — if the source has an audited audience demographic survey — as most publications that sell advertising do — then it is easy to associate any memes that occur within its content with particular geography and demography). Now the question is suppose we are tracking a particular meme — can we determine its geographic trajectory over time? Can we

determine the vector of each meme at each sector in a geographic map? And can we represent that in an animated map for example, perhaps with something like a fluid flow animation?

Really **interesting memes** — the memes that represent important emerging trends or current hot issues — will exhibit **high volatility**.

Another open area to study is to analyze the higher order distributions of memes in order to automatically detect memes that are "interesting" (i.e. not "noise" according to our priorities). One easy way to do this is to automatically ignore any memes that have a random distribution. We may also want to de-emphasize memes that have regular distributions — such as memes for which the dynamics have been the same for a reasonable period of time. In other words, we want to filter for memes that have dynamics that deviate from being predictable or stable (randomness and regularity are both predictable). My hypothesis is that the really interesting memes — the memes that represent important emerging trends or current hot issues — will exhibit high volatility. For example, imagine for a moment that we are tracking memes related to "digital music" — if we look back in time there will be a point where the word "Napster" suddenly appears — at first it is a relatively "small" meme but gradually it spreads and gains momentum. Then there is a critical point where it begins to grow exponentially. Then it probably levels off for a while or even inflects after the initial hype phase

ends. Next another dramatic increase in momentum should be seen around the time of the music industry's lawsuits against Napster. Then following the resolution of these we should see Napster fall off dramatically. Later we see momentum increase again as the new commercial version of Napster is announced. This type of pattern is what we are looking for. Can we characterize these patterns well enough that we can detect them automatically?

This approach could be used to **train intelligent agents** that specialize in **detecting particular types of trends**.

Perhaps one way to do this is by training a neural network to recognize the types of patterns that interest us — we could do this for example by taking historical content (such as the last 10 years of the Associate Press) and then telling a neural net what memes are most important to us. The neural net can then learn from this training data. We can then run the neural net on current or more recent news and let it guess what is important to us based on the patterns of past important trends. We can rate these guesses to provide further feedback to improve learning. This approach could be used to train intelligent agents that specialize in detecting particular types of trends — for example, we could train agents to alert us when a major new technology trend is about to erupt, or when we should invest in a technology stock, or when a company we track is experiencing a major change of some sort, or to tells us when

a new competing product emerges or when an existing competing product overtakes our own product, etc. We could also potentially train agents to recognize the early signs of important cultural or political issues, significant changes in sentiment or focus for a given community we are interested in, or even signs of emerging threats.

ARE THERE IDEAL MEME DISTRIBUTIONS?

Perhaps one of the most interesting questions I have thought about in relation to the physics of ideas is whether or not there are perhaps "ideal distributions" of memes that get the best response from humans? In other words, do the higher order distributions of memes that become major trends, or that get the most attention in noisy environments, have similar characteristics? If it turns out that this is the case then it could provide a powerful new technique for advertising, information filtering, and even for user-interface design. I believe we can analyze memes to answer this question. Here's how we might do it:

APPROACH 1: We choose a representative set of memes for major trends. We analyze their higher order distributions in the media. We then attempt to figure out whether these distributions have anything in common that we can isolate. We then search the media for other memes that have distributions with similar properties and test whether they are in fact major trends. We can provide feedback by scoring the output of these trials and using an evolutionary algorithm to evolve successively better filters. Eventually through such a process we can evolve an agent that is good at discovering major trends in the media.

APPROACH 2: We can do a perceptual psychology experiment to discover and evolve memes that get the most attention. Create a noisy environment in any sensory modality — let's use visual information for the moment. Put 100 human subjects in a room and show them a computer generated slideshow. Our slideshow consists of 100 images. We change slides rapidly. Each slide is shown many times in the course of the slideshow, with a frequency according to one of many different distributions we wish to test. For example, one slide is shown such that it has low mass, low velocity — a low momentum. Another is shown to have high momentum. Others are shown to vary

By doing this experiment many times with many distributions we can **experimentally determine** which **memetic distributions** have the strongest **cognitive impact**.

such that their momentum inflects and is volatile. We can test a number of different momentum curves in this manner — such as linear or nonlinear momentum growth, etc. At the end of the slideshow we give each subject all the slides and ask them to prioritize them in order of most important to least important — we ask them to tell us what they think the most important slides in the slideshow were. This effectively tests the various distributions we ran in the experiment to see which ones had the strongest cognitive effect on the subjects. Two weeks or a month later we repeat this rating

test to see which distributions have the strongest long-term effect as well. By doing this experiment many times with many distributions we can experimentally determine which memetic distributions have the strongest cognitive impact. The next step would be to test whether the distributions we discover are applicable across sensory modalities — for example, do the distributions we found for vision also work for the auditory system. My hypothesis is that they do hold across modalities. If this is the case then we have discovered a key underlying meta-pattern in the human perceptual system — the pattern by which humans recognize what to tune their attention to.

An advertiser or marketer could use this knowledge to **design campaigns** that get the most attention and that are not “**tuned out**” by people as quickly.

There is another interesting and related question to the above experiments: Do certain distributions retain attention better than others? The human perceptual system attenuates to signals very quickly — we tune out anything regular or predictable and focus on identifying novelty. But what is "novelty"? Any new meme that occurs is novel at first, but whether or not it remains novel or gets tuned out is another question. Which meme distributions do NOT get tuned out as quickly, or ever? Is there an optimal way to vary the distribution of a meme such that it continues to remain novel? In thinking about this, are there any meta-patterns to the memes that have gotten your attention in the past? For example, is there something about the way that particular

technology trends or celebrities have moved through the media that made them appear to be hotter and more important to you? Having high momentum at a given time is part of this, but it may in fact be the change in momentum over time — the "meme impulse" — that really makes the difference. For example in my own experience I notice that trends that exhibit exponential growth in momentum quickly get my attention — but as soon as the growth becomes predictable I lose interest. So it seems that the trends that retain my interest the best are the ones that have more variable graphs — graphs that are neither random nor regular. Is there an ideal balance between randomness and order? What patterns have this balance — can we quantify this and define it more concretely?

Mememes are now among the **most powerful**
“hidden” **forces shaping** our individual minds,
our relationships, organizations and **our world**.

A better understanding of the cognitive effects of various higher order distributions of memes in various human sensory modalities could be particularly useful for advertisers, marketers, and user-interface designers. An advertiser or marketer could use this knowledge to design campaigns that get the most attention and that are not "tuned out" by people as quickly. A user-interface designer could use this information to design interfaces for managing changing information in which the signal-to-noise ratio is optimized so that users can quickly focus on just the most important changing

information — for example the information display of a stock-trading terminal, executive information system, military situation room, or fighter jet cockpit user-interface could perhaps be improved using these principles.

CONCLUDING REMARKS

Given that memes are now among the most powerful "hidden" forces shaping our individual minds, our relationships, organizations and our world, wouldn't it be great if we could really measure them and analyze them empirically?

That is what I hope the basic techniques provided above will help to catalyze. By making this hidden layer visible we can gain a much better understanding of our world. Let me know if you end up using these techniques for anything interesting (and hopefully you will make your ideas open-source too so everyone can benefit).

What these basic techniques provide is a way to measure the movement of ideas in time and space. For example, we can track the trajectories of ideas in our workspaces, our teams, enterprises, cities, nations or interest-communities. We can also track them across geography or any other set of dimensions.

Because we can compute basic physical properties of memes we can start to apply Newtonian physics to analyze them. Perhaps by doing so we can really develop a "Physics of Memetics" with which we may begin to predict the outcomes of interactions among memes, the future trajectories of memes, and the influence changes to memes have on events in the so-called "real world" and vice versa. With this in hand we could potentially teach systems to learn to detect memetic patterns of interest

to us — for example the early "fingerprints" in the media that indicate the outcome of a proposed act of legislation or a vote, or a stock price, or a political change. We could also use it to detect emerging cultural trends, and to measure and compare the dynamics of brands or competing technologies in various markets in order to predict winners.

By putting this information into the public domain I hope to see these techniques in use as widely as possible. They will provide dramatic benefits in managing large volumes of information, improving knowledge worker and team productivity, and in discovering and measuring trends in communities.

Ultimately, I would like to see this embodied in a "grand cultural project" — a real-time map of the memetic dynamics taking place around the globe. This map would be filterable in order to show relative memetic dynamics in different places, communities, etc., and to show how various memes are spreading and interacting over time around the world. The data would be open and accessible via an open API so that all services that manage information could provide information to it and query it for stats when needed.

info

ABOUT THE AUTHOR

Nova Spivack is President and CEO of Radar Networks, a stealth-mode software company that is developing a platform for the Metaweb — a new technology for working with microcontent such as RSS, weblogs, Atom, and the coming Semantic Web. Nova is the grandson of Peter F. Drucker and shares strong interests in improving group cognition, innovation, and knowledge work. Nova studied philosophy with a focus on artificial intelligence and cognitive science, at Oberlin College, did cellular-automata and complex systems research at MIT, and did graduate study at The International Space University, which led to his flight to the edge of space in 1999.

DOWNLOAD THIS

This manifesto is available from <http://changethis.com/7.PhysicsOfIdeas>

SEND THIS

Click here to pass along a copy of this manifesto to others.

<http://changethis.com/7.PhysicsOfIdeas/email>

SUBSCRIBE


Learn about our latest manifestos as soon as they are available. Sign up for our free newsletter and be notified by email. <http://changethis.com/subscribe>

info

WHAT YOU CAN DO

You are given the unlimited right to print this manifesto and to distribute it electronically (via email, your website, or any other means). You can print out pages and put them in your favorite coffee shop's windows or your doctor's waiting room. You can transcribe the author's words onto the sidewalk, or you can hand out copies to everyone you meet. You may not alter this manifesto in any way, though, and you may not charge for it.

NAVIGATION & USER TIPS

Move around this manifesto by using your keyboard arrow keys or click on the right arrow (→) for the next page and the left arrow (←). To send this by email, just click on .

KEYBOARD SHORTCUTS

	PC	MAC
Zoom in (Larger view)	[CTL] [+]	[⌘] [+]
Zoom out	[CTL] [-]	[⌘] [-]
Full screen/Normal screen view	[CTL] [L]	[⌘] [L]

BORN ON DATE

This document was created on 1 November 2004 and is based on the best information available at that time. To check for updates, please click here to visit <http://changethis.com/7.PhysicsOfIdeas>

info



COPYRIGHT INFO

The copyright in this work belongs to the author, who is solely responsible for the content. Please direct content feedback or permissions questions to the author: <http://www.mindingtheplanet.net>

This work is licensed under the Creative Commons Attribution–NonCommercial–NoDerivs License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/2.0> or send a letter to Creative Commons, 559 Nathan Abbott Way, Stanford, California 94305, USA.

ABOUT CHANGETHIS

ChangeThis is a vehicle, not a publisher. We make it easy for big ideas to spread. While the authors we work with are responsible for their own work, they don't necessarily agree with everything available in ChangeThis format. But you knew that already.