Statistical and Physical paradigms in social sciences

(Sociophysics, Econophysics, Mediaphysics)

Igor Mandel and Dmitri V. Kuznetsov

Media Planning Group (MPG)
New-York-Boston
May 2006

Sociophysics and Statistics,
or
Two prisoners dilemma

"I did that not in the interest of being right, but in the interest of being correct" – said he enigmatically."

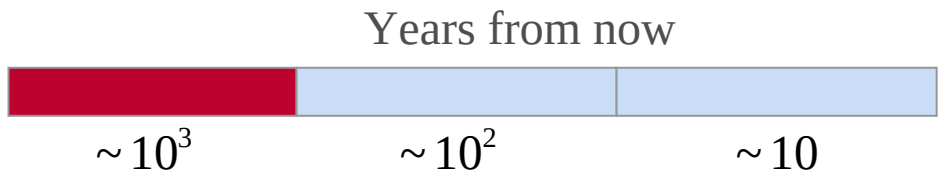
Ilf and Petrov, *Twelve chairs*

Invasion of physics into social sphere, where statistics plays a key role as a methodological tool, created already and will create a lot of problems, different from those what science faced earlier when different …-physics were established.

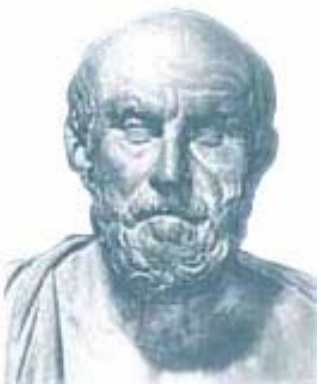
Some of them are the topic of this presentation.

1. Sociophysics – historical highlights

Sociophysics: **View from today.**



Sociophysics. ***Analogy:***



Empedocles

495-435AC

"Some people are like water and wine, mixing easily, while others

are like water and oil, not mixing" *(cited by Stauffer, 2004).*

•First (?) of many other analogies between human and physical worlds, spread over centuries, including philosophical ones;

•V.Pareto (1897) compared discovered by him law of wealth distribution with Kepler's laws of planets motion, what become quite prophetic and somehow unique;

•In modern time, the most general work in that direction is, maybe, "**Sociophysics**" by P. Arnopolous, 2005

Sociophysics. *Naming and Revelation*

Adolphe Quetelet
1796-1874

*"**Sur l'homme et le developpement de ses facultés, essai d'une physique sociale**", 1835*

"...the greater the number of individuals observed, the more do individual peculiarities, whether physical or moral, become effaced, and leave in a prominent point of view the general facts, by virtue of which society exists and is preserved"

- Direct physical analogies from physics and astronomy to social life;
- Promotion of importance of normal distribution for social phenomena;
- Concept of an ideal "average man" with individual random fluctuations, overly ridiculed later;
- Direct influence to J.Maxwell (kinetic theory of gases), F.Galton (genetics, eugenics, regression analysis), K. Marx (theory of price), E. Durkheim (sociology), K. Pearson (statistics) and many others;
- Weight/Height (body mass) index (*index Quetelet*), used as obesity measure until now – a (very physical) concept of constants in human life

Sociophysics (Econophysics). ***Strong implementation of physics***

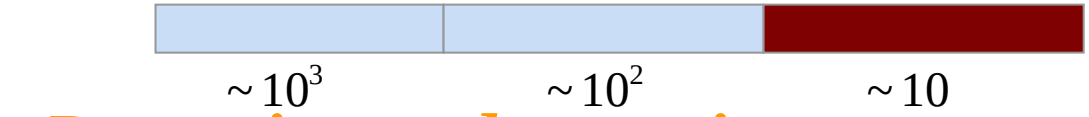
Louis Bachelier
1870-1946

Theorie de la Speculation, 1900

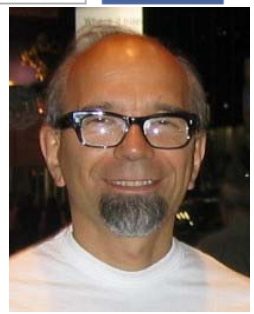
"The manner in which M.Bachelier deduces Gauss's law is very original....He derives it in ***a chapter whose title may at first seem strange, for he calls it "Radiation of Probability." In fact, the author makes a comparison with the analytic theory of heat flow.*** *A bit of thought shows that the* ***analogy is real and the comparison is legitimate****.The reasoning of Fourier, almost without change, is applicable to this problem so different from the one for which it was originally created."*

(Appell, Poincare, Boussinesq, from the letter recommending Bachelier's thesis to publication, translated from French, 1900, our bold – I.M., D.K.)

- Revolutionary work, advancing many directions for decades, admired by some very early (A.Poincare, A.Kolmogorov) and abandoned by many others;
- Concept of random walk, before Browning motion model by A. Einstein's theory (1905);
- Concept of stochastic processes, before A.Markov (1906) and N.Wiener (1940s);
- First work in financial mathematics;
- The Bachelier Finance Society was founded in 1997



Sociophysics. *Renaming and emerging*

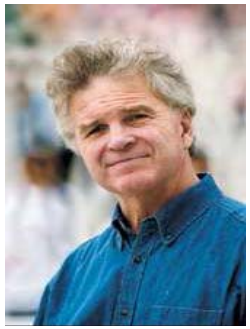


Serge Galam

Galam.S., Y. Gefen, and Y. Shapir , 1982 **Sociophysics**: **A new approach of social collective behaviour**. 1. Mean behaviour description of a strike. J.of. Mathematical Sociology, 1982, v.9

- Gave a name of a new field of research with reference to previously published physical implementations which didn't use this name;
- Now – hundreds of articles, 18500 references in Google;

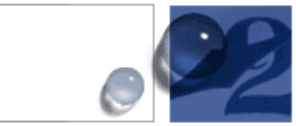
Econophysics. *Naming and emerging*

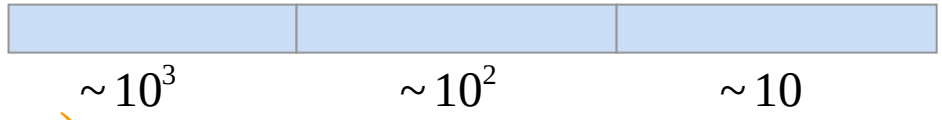


Eugene Stanley

H.E. Stanley coined a term in 1994 (http://www.unifr.ch/econophysics/PHP/principal/redirect_news.php?id=62) and then together with R. Mantegna published a most cited until now book on subject **An Introduction to Econophysics**, 2000.

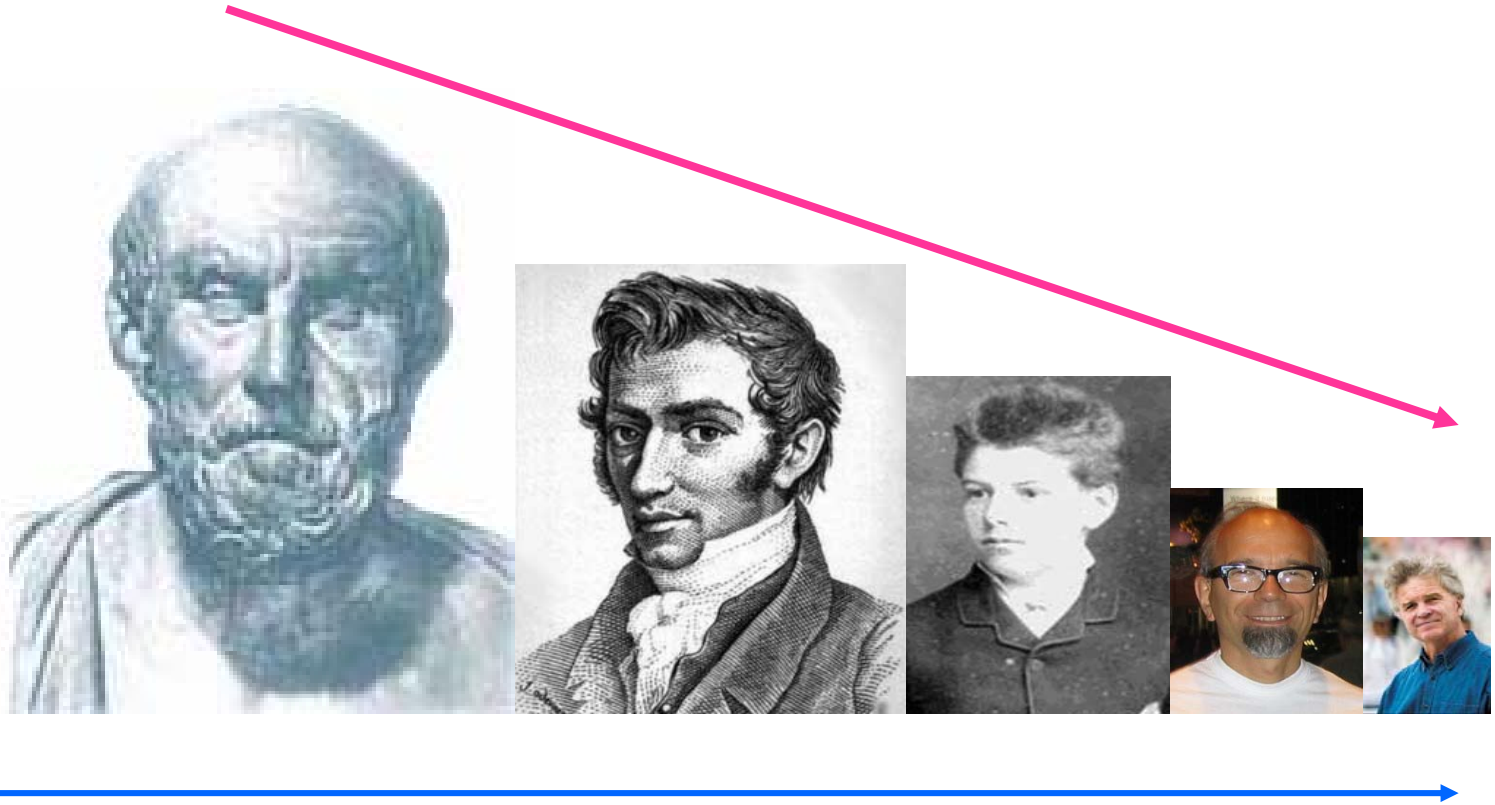
- Econophysics originally was considered as a replacement for "phynance", i.e. "physics+finance", and, indeed, was and is understood by many as a science about financial markets. However, later many other traditional economical areas have been also covered;
- As economics is included in social sciences, **econophysics**, in broad sense, **should be considered as part of sociophysics**;
- Now – several books, thousands articles, regular conferences, 208,000 references in Google





PHYSICS (**side question**): Should we measure importance of one's contribution in this (classical) way, because the earlier the wider influence is spread?

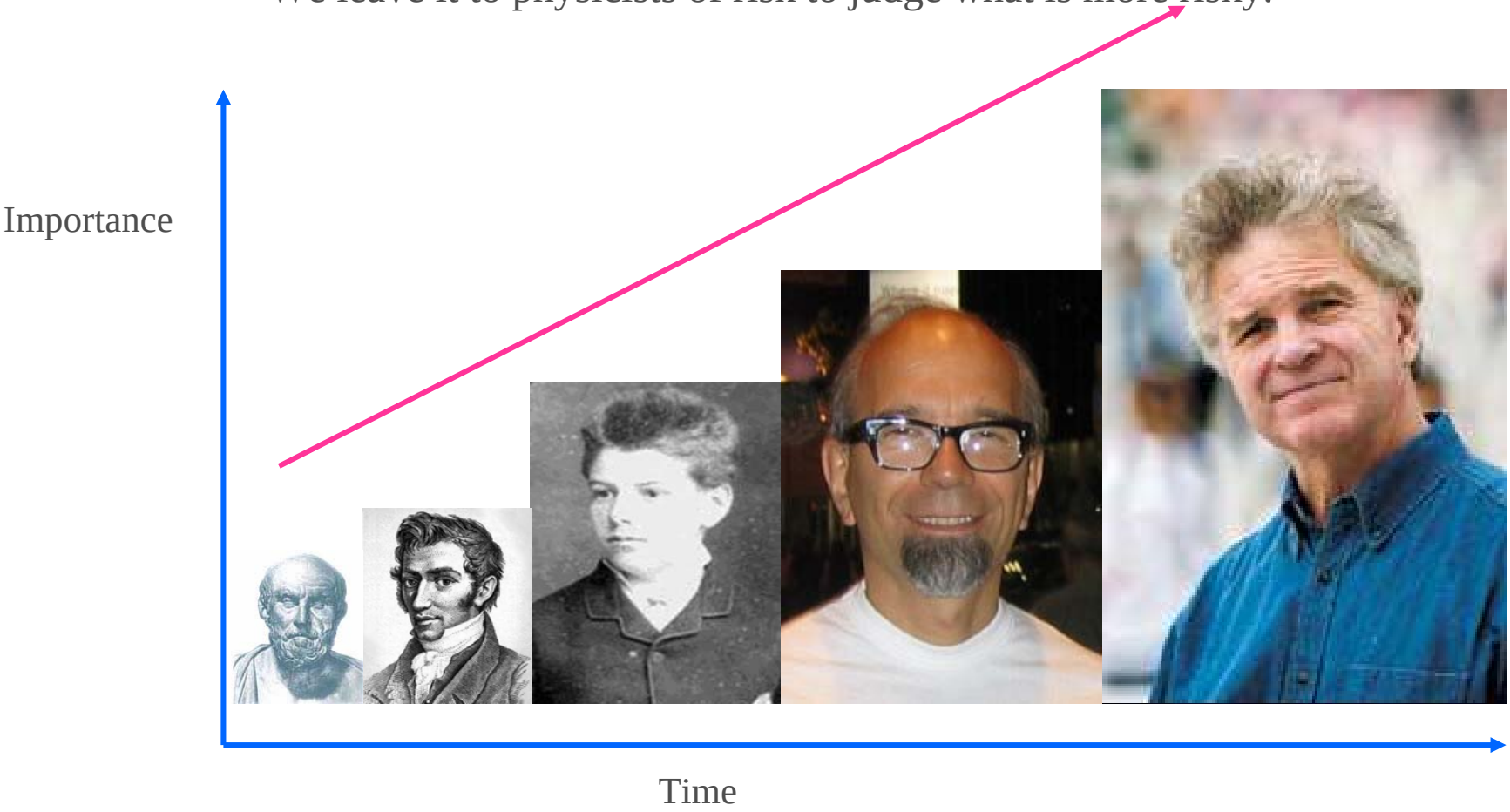
Importance



Time

PHYSICS (**side question**): Or in this way, because the later the "more scientific" it is?

We leave it to physicists of risk to judge what is more risky.



2. Sociophysics and Statistics

Sociophysics: *seven branches*

•**Empirical** – demonstration of existence of some, supposedly physical-like, phenomena in social life (mainly distributions of different types);

•**Modeling** – building physical models, possibly explaining empirical facts (mechanisms of generation of observed distributions);

•**Analytical** – known models of different processes from physics receive social interpretation and modified to get new analytical results (Ising model getting opinion-exchange interpretation), without direct link to observed data;

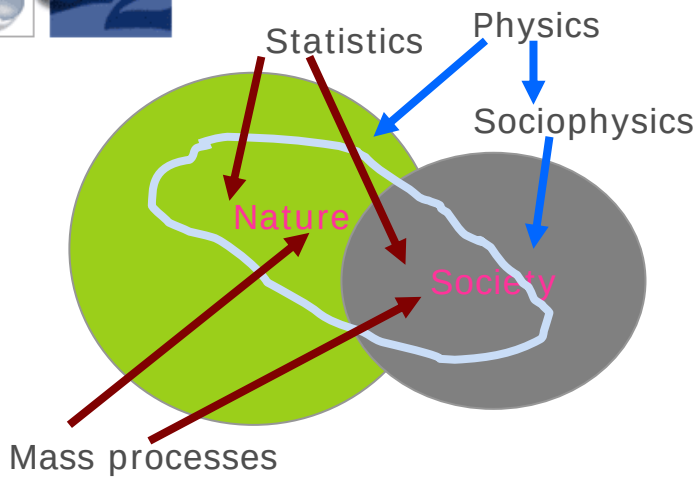
•**Simulation** – some physical concepts are used as basics for building social analogy and then are simulated in order to understand the convergence of the process (assuming that interaction between people go in certain way, what does it yield when some time passed);

•**Aggregative** – using theoretical physics (like in modeling and analytical branches) and simulation, merged with idea of solving particular problems and approximating concrete data, like in statistics (mediaphysics);

•**Conceptual** - no exact models are proposed, but similarities between two worlds are traced, like expansion of general laws of nature to social realm;

•**Nonphysical –** pure mathematical or econometrical models or simulations of possible social situations without link to physics, published under the “sociophysical roof” (often some publications are undistinguishable from publications on artificial societies and so on);

Sociophysics and Statistics. ***Interactions***



1. Subject

- Statistics studies only mass phenomena in nature and society, while physics not only those;
- Physics has born many branches (astrophysics, biophysics, sociophysics and so on), exactly as statistics (social, economical, biological, etc.);

- In social studies, statistics is the only way to measure things, while physics may work on phenomenological level as well. It provides a ground for possible conflicts;
- So far statistics was substantiated by respective subject sciences (economics, biology, etc.). Should physics replace it? If not, what is statistical and what is physical?
- For the first time, physics learns principally not physical phenomena, what makes her status **analogical.** Even if analogy is a basis for **any** modeling, in this case the distance between object to be modeled and means of modeling is especially big.

Sociophysics and Statistics. *Interactions*

2. Methodology of analysis

•Not only methods between statistics and physics where inter-exchangeable (statistical physics), but from social life to physics (Quetelet, Bachelier) and backward (typically in modern time). It tells about similarity in some extent. **But differences are no less important.**

•Statistics has two main paradigms: **analytical and probabilistic**. Within the first, any data could be analyzed somehow and some conclusion derived. Second presumes existence of "general population" with some assumed features (distributions, etc.), from which data was drawn, and all (still analytical) conclusions should be considered only in respect to this population and associated assumptions.

•Physics does not work in analytical mode at all; it does with probability, though in a principally different manner. It creates a model of **individual behavior** of constituents, and just after that statistical distributions become derived from the model.

•Statistics:

mass assumptions → concrete data → mass conclusions;

• Physics:

individual assumptions (model) → abstract data (or no data) → mass conclusions.

•As a result, they have different perception of **causality** – very clear in physics and very fuzzy in statistics. Physical formulas have nothing to do with statistical "regression equations with many factors"

Sociophysics and Statistics. ***Interactions***

3. Methodology of forecasting

•A forecasting is an ultimate goal or at least very important ingredient of any science;

In physics:

- conditions under which model is derived are usually very transparent and repetitive;
- if a model is correct, than forecasting is “automatically” correct within given conditions;
- if model is incorrect, forecasting still some times might be not bad (a geocentric model in past; McCutcheon’s “final theory” today (*is it really wrong?*).

In statistics:

- conditions under which a model is derived, usually are just declared, but not checked (because it’s almost impossible) and not repetitive over time and space;
- model is very rarely correct in causal, not in “approximational” sense;
- respectively, forecast could be bad or good, what often not really depends on the model’s quality, but rather on stability of a life and luckiness of a researcher (look at thousands of “financial gurus” forecasts).

Sociophysics and Statistics. ***Convergence***

4. Convergence?

As Babel's personage once desperately cried, "Where Benya Krick *(a local mafia godfather-I.M.,D.K.)* is ending and where police is beginning?" Should one even try to make a demarcation line between two sciences discussed (as we did) or it will be no need in it?

Do one likes it or not, it should be done. The fact that statistics anyway cannot be replaced from social studies for it is the only tool there, sociophysics has to find some **complimentary** place in there.

As it goes from all considered , **a deep contradiction between solidness of physical laws and transient nature of statistical regularities in social life might be a main battlefield for sociophysics.** And this problem is much more serious that mutual relations between different departments in universities or different directions in journals (as similar problems are often understood).

That's why so lovely topic in sociophysics now is **power low**: it is, maybe, the only example, where some strong empirical statistical evidence could be happily married with different, some times very exotic physical theories. And it's not random, that power low is a one-dimensional distribution; if one goes to any two-dimensional problem – this beauty will be immediately distorted.

However, the importance of physical paradigm as opposite to statistical one in social science is obvious. The new ways of co-habitat should be found. **Mediaphysics** is intended to work in that direction.

3. Mediaphysics:

Introduction and motivating example

http://arxiv.org/abs/physics/0506217

Mediaphysics: *Declaration of Dependence*

Definition: **Mediaphysics** describes processes of mass communications in social life by techniques of statistical physics (**media** is understood like medium, where individuals act);

Second meaning: way of analyzing **mass media** from statistical physical prospective;

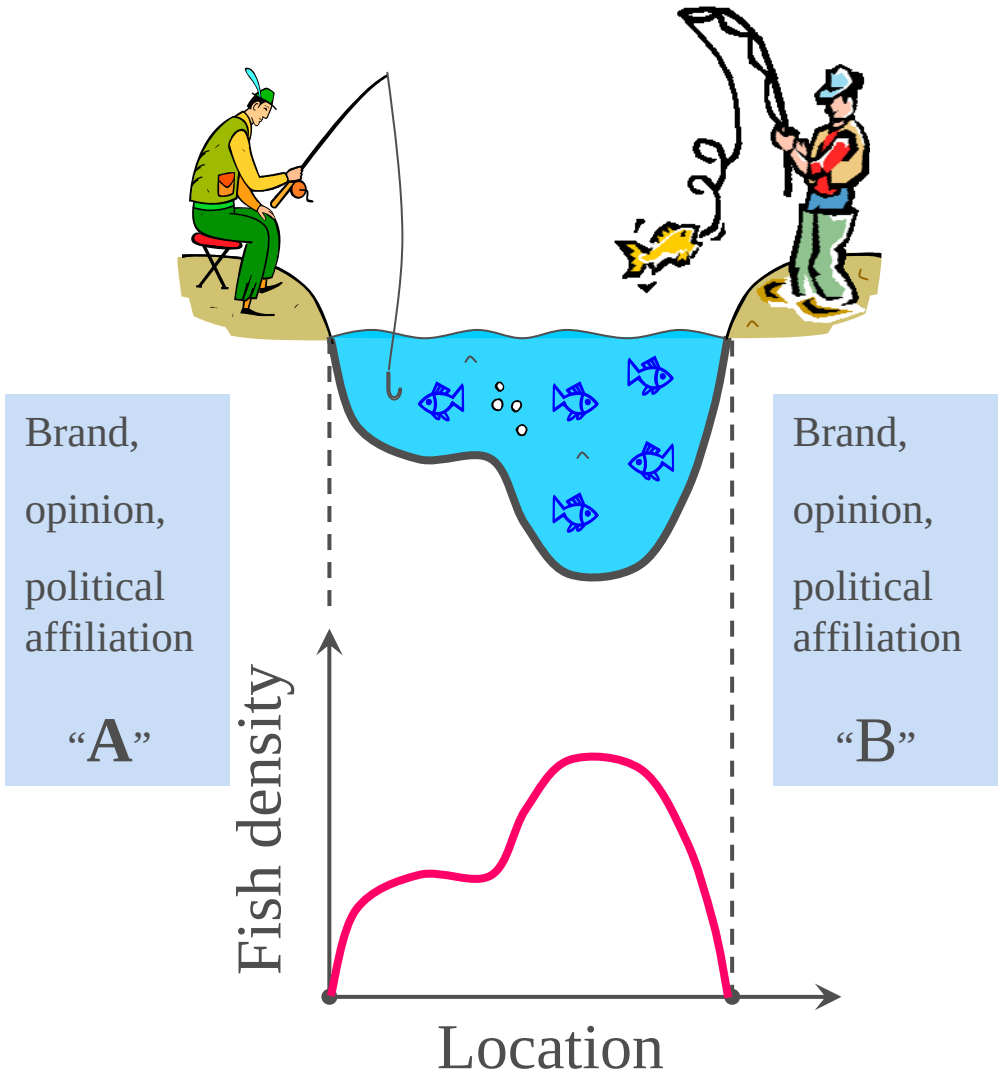
Purpose: building a systematic way to apply methods of statistical physics to practical problems, usually solved by traditional statistical means;

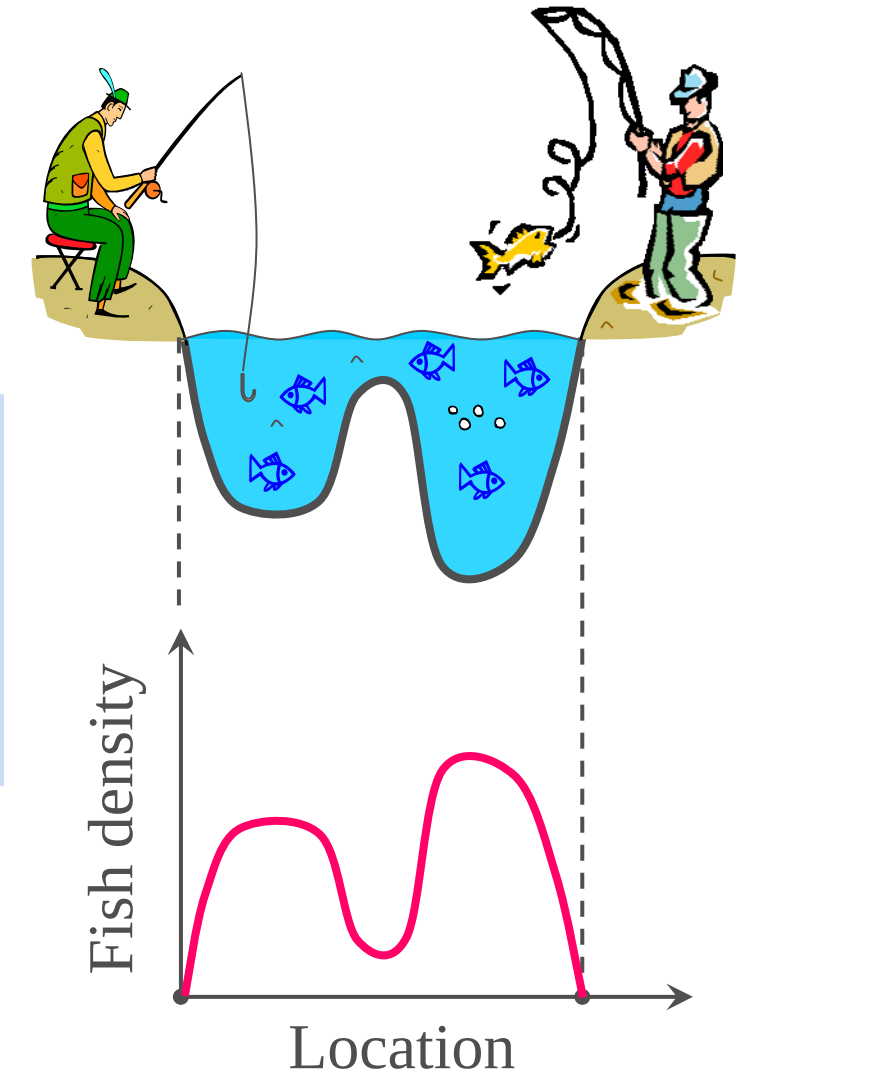
Basic analogy: a mental-physical space, in which processes of making decisions and acting are taking place (space of "willingness to buy", "propensity to participate", etc.), where laws and methods of statistical physics are presumably work for mass phenomena;

Relation with sociohysics: part of it, dealing with described types of processes. Some published works could be considered within a realm of mediaphysics, some are definitely not.

Relation with statistics is twofold. First, it distinguishes two types of data (data about one John Smith vs. data about mass processes), what in statistics are treated in the same way. Second, it has a lot in common with statistics because is related with statistical techniques of estimation of concrete parameters (mathematically it is partly shown below).

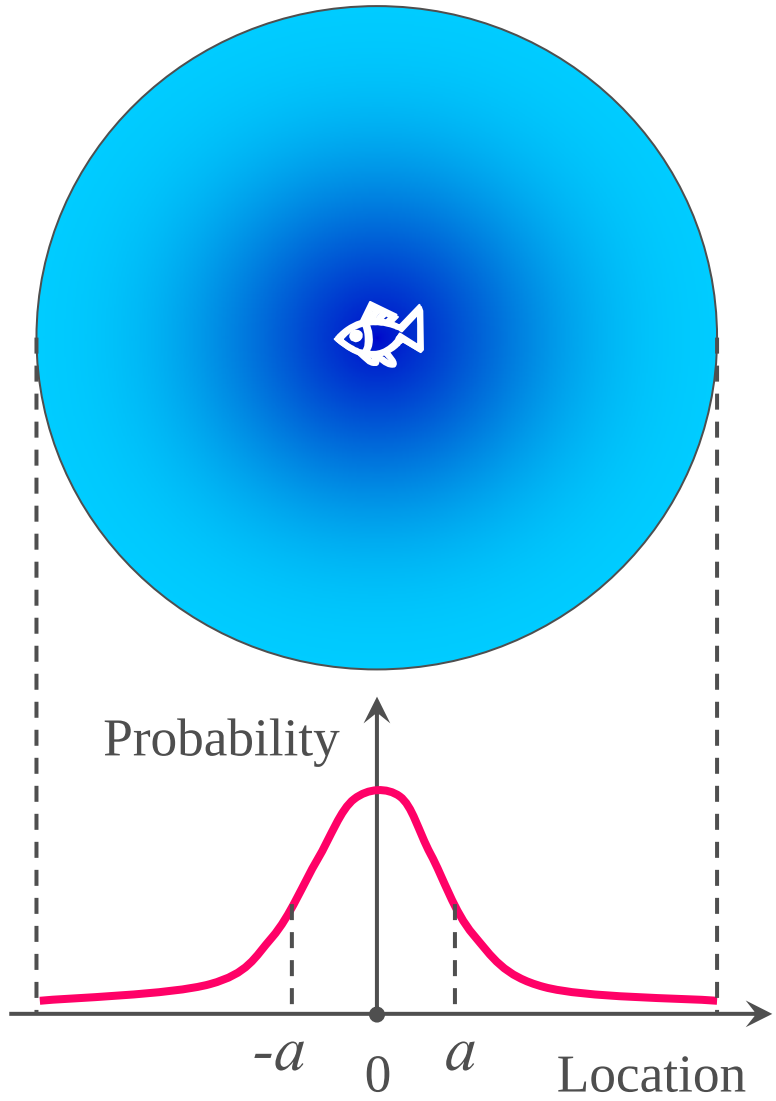
Competitive & Non-competitive Factors





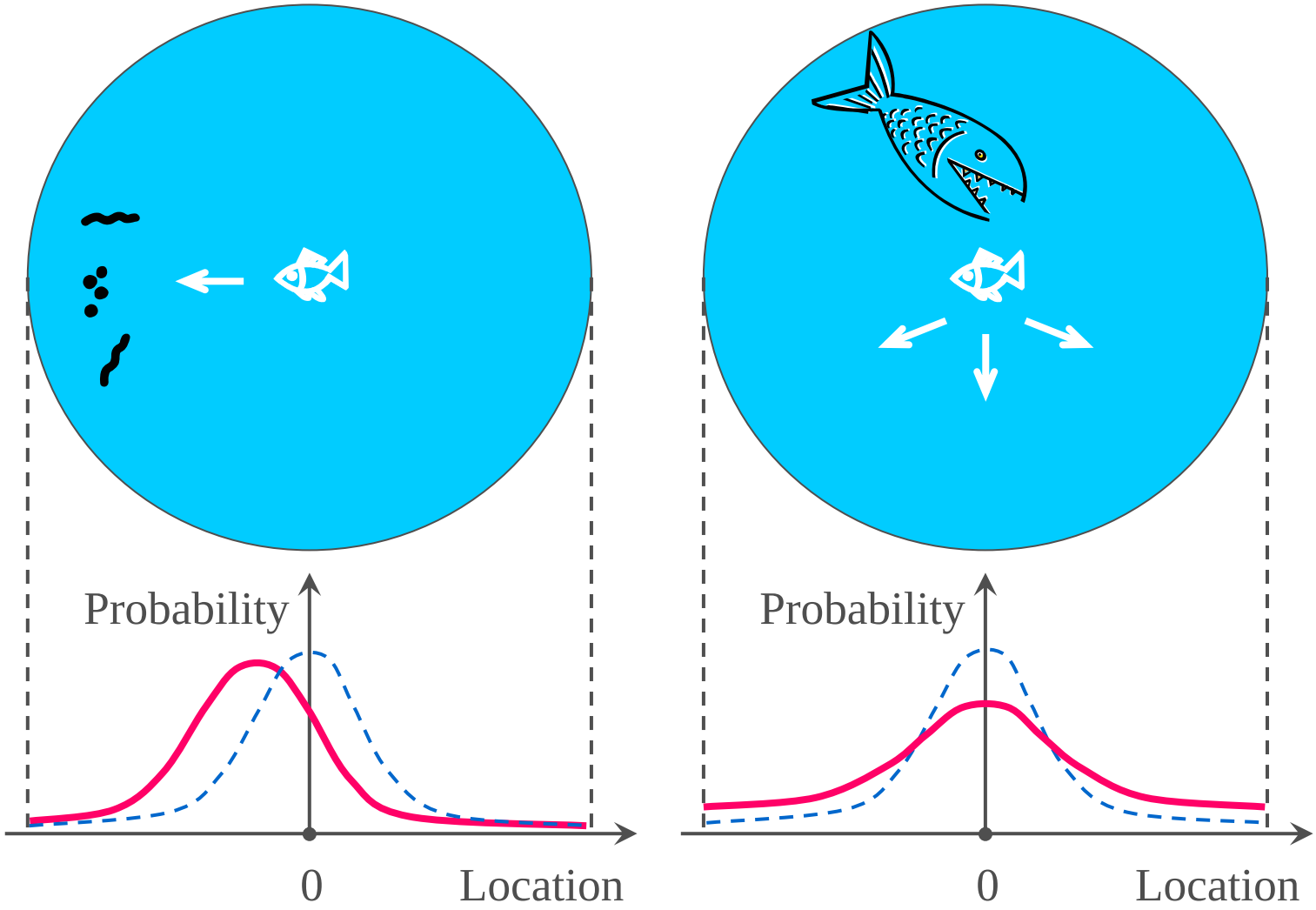
Single-Fish “Free” Swimming
Relocation Probability & Fish Types

- “Free” (random) swimming:
  - No attractors or repellents
  - No space boundaries or barriers
  - The higher distance, the lower probability to be there
- Characteristic value:
  - Root-mean-square displacement a (measure of fish’s “Flexagility”)
    - Flexibility = Often directional changes
    - Agility = Fast move in a given direction
  - Low Flexibility + High Agility = High Flexagility
  - High Flexibility + Low Agility = Low Flexagility
  - Displacement $a(i)$ for fish type i

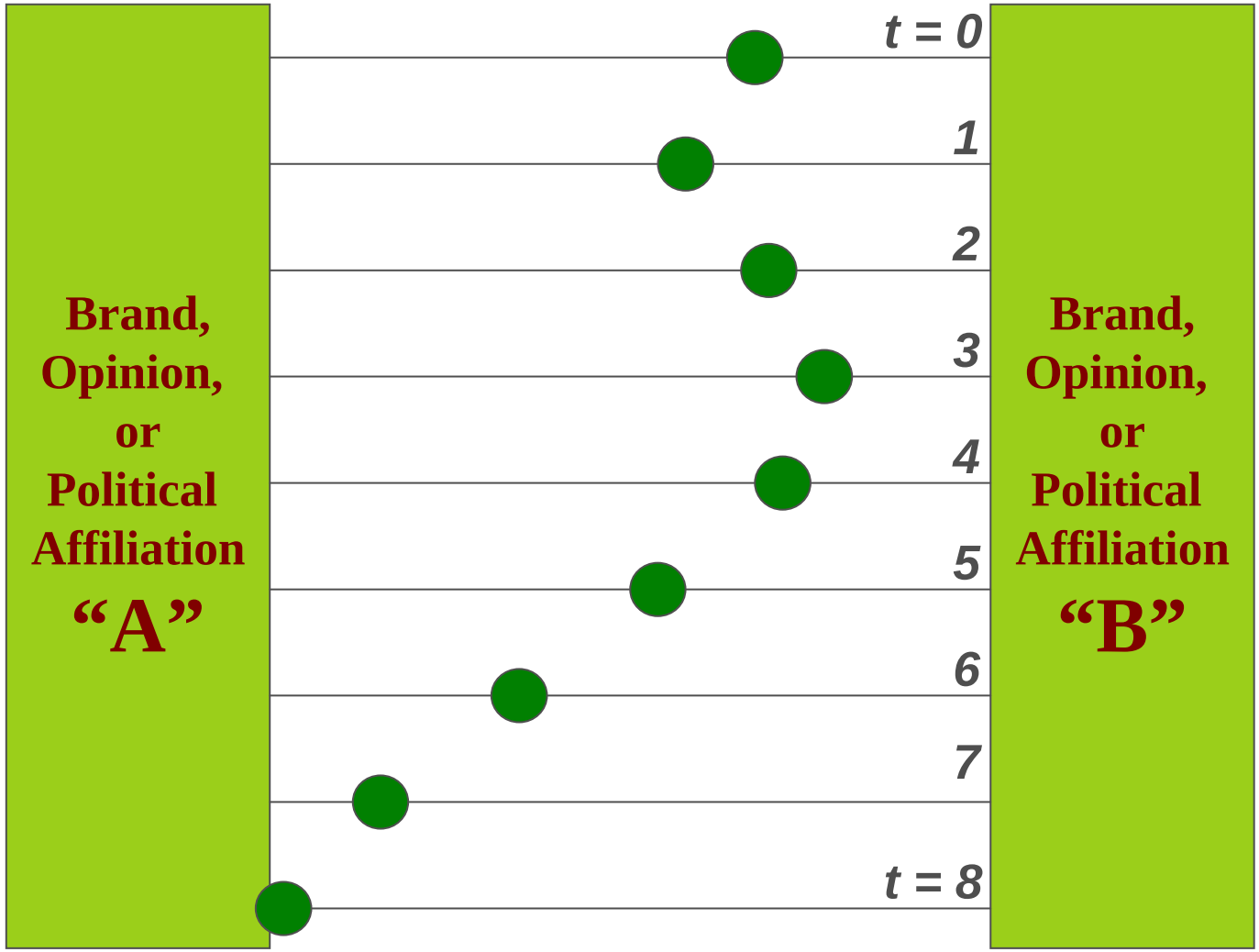


Single-Fish Swimming With Stimulus

Attractors, Repellents, Barriers & Boundaries

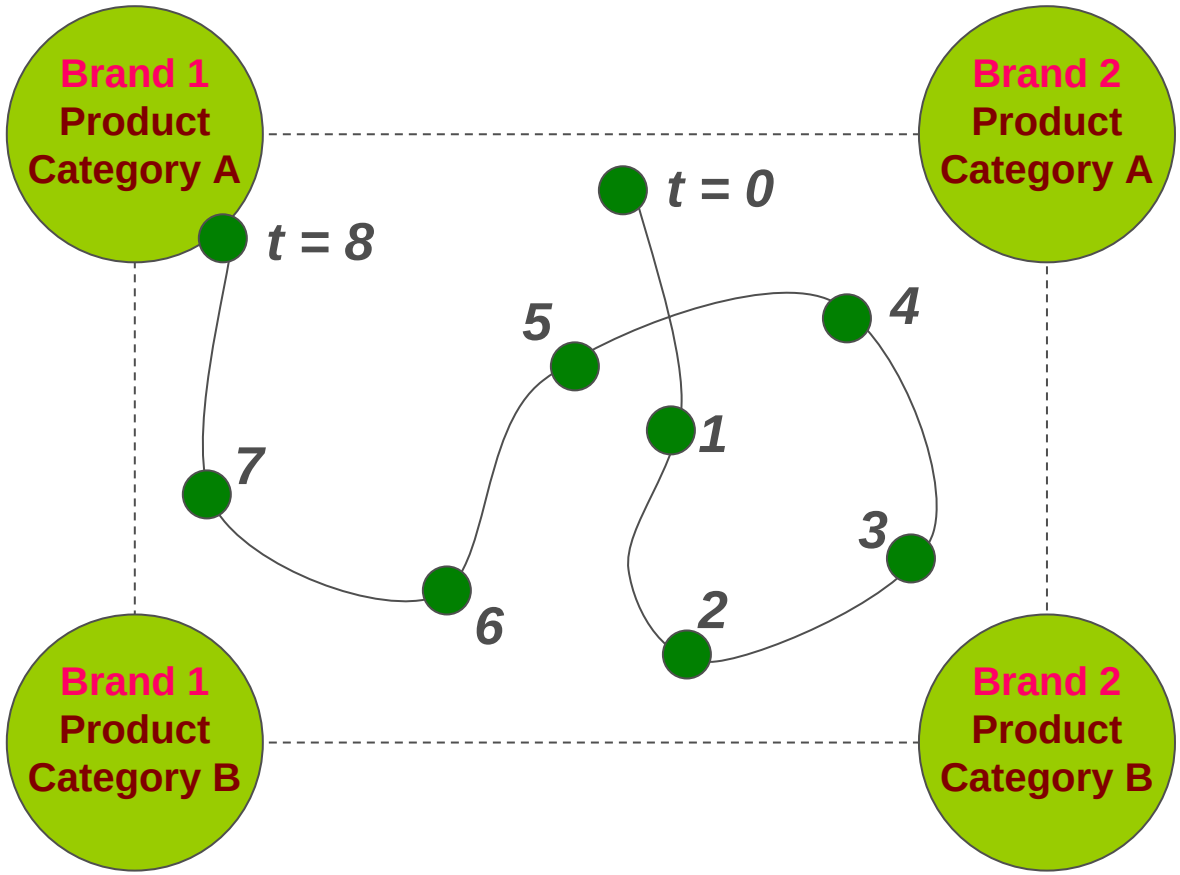


Single Person
Mindsets Floating Between 2 Opinions



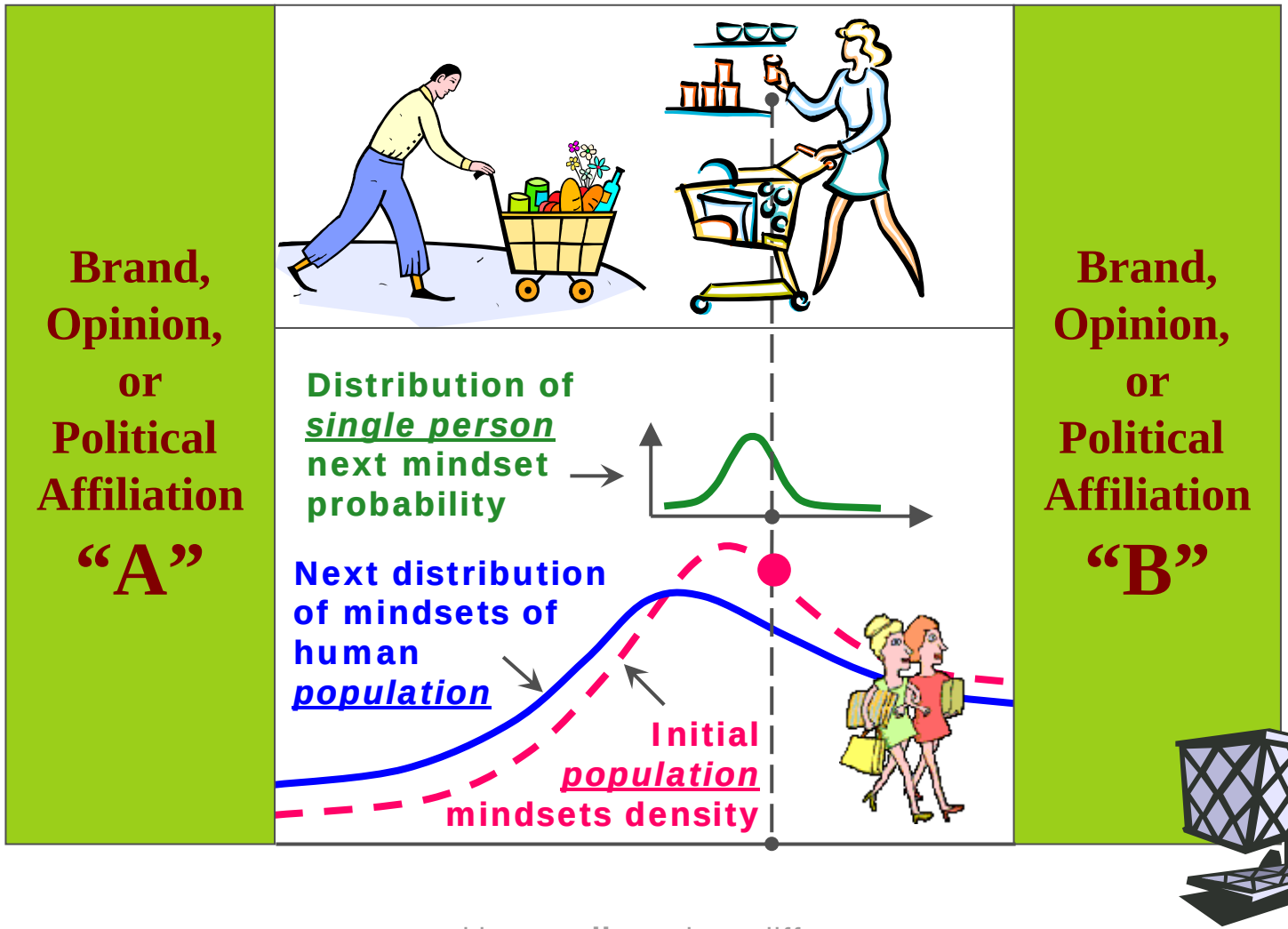
Single Person

Mindsets Floating Between Many Opinions



Human Population

Distribution of Mindsets and Its Dynamics



Mediaphysics

Real Life Features and Objectives for Complex Social Systems

- Two or many *alternatives* to choose from
  - Brands, Categories of products, Political affiliations, Faiths, Opinions
- Applied *factors*
  - *External:* Economic, Social, Marketing, Propaganda, Natural, etc.
  - *Internal:* Opinion influential, Population heterogeneity
- Real-life specialties to worry about
  - Population *connectivity and inertia* in opinion formation
  - *Partly unobserved* persons (not yet buyers etc.)
- Outcome (Sales, Voters, Followers)
  - *Dynamics of population distribution* between alternatives
  - *Short & Long-term Forecast*
  - Marketing & Propaganda Optimization

4. Mediaphysics: Methodology

http://arxiv.org/abs/physics/0506217

Methodology
Green's Function

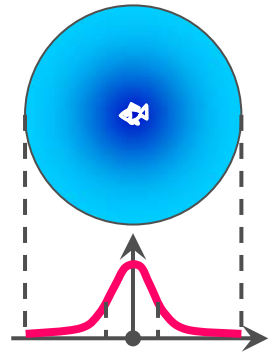
- Green's function $G_t(\boldsymbol{q}_t, \boldsymbol{q}_0)$ is conditional probability to move from state (of person's mind) $\boldsymbol{q}_0$ to $\boldsymbol{q}_t$ at time t
  - $\boldsymbol{q}_t$ is one- or multi-dimensional generalized coordinate
- Next step recurrence relation: $$G_{t+1} = \hat{Q} G_t$$
  - Transfer operator: $$\hat{Q} = \exp\left[-W_{t+1}(\vec{q})\right] \cdot \hat{g}$$
    - Applied fields (motivations): $W_t(q)$
    - Connectivity operator for mindsets of "free" persons (isolated from external factors and opinion influences, $W_t(q) \equiv 0$)

$$\hat{g} G(\vec{q}, \vec{q}_0) = \int g(\vec{q}, \vec{q}') G(\vec{q}', \vec{q}_0) d\vec{q}'$$

Methodology
Connectivity & Personal "Flexagility"

- Connectivity operator for free-person mindsets

$$\hat{g}G(\vec{q},\vec{q}_0)=\int g(\vec{q},\vec{q}')G(\vec{q}',\vec{q}_0)d\vec{q}'$$



  - Kernel $g(\boldsymbol{q},\boldsymbol{q}')$ of the operator is Markov chain conditional probability for single person in the absence of any motivations (fields)
  - For Gaussian model in D-dimensional location space $\boldsymbol{z}$:

$$g(\vec{z},\vec{z}')=\left(\frac{D}{2\pi a^2}\right)^{D/2}\exp\left[-\frac{D(\vec{z}-\vec{z}')^2}{2a^2}\right] \quad \rightarrow \quad \hat{g}\approx 1+\frac{a^2}{2D}\Delta_{\vec{z}}$$

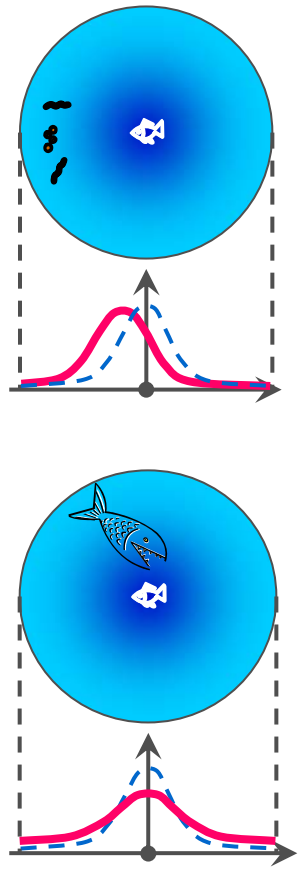
  - Mean-square displacement (a measure of personal flexagility):

$$a^2=\int \vec{z}^{\,2}g(\vec{z},0)d\vec{z}$$

Methodology
Motivations (External Factors & Influences)

- **Total field** (in the simplest form):

$$W(\vec{z},t) = W_0(\vec{z},t) + W_C(\vec{z},t) + W_F(\vec{z},t) + W_I(\vec{z},t)$$



- W_0 & W_C are field contributions from own & competitor's advertising activities (for many different national and local channels: TV, Radio, Magazines, Newspapers, Internet, Outdoor), i.e. each field is, in turn, a function of many parameters;
- W_F is from general (non-advertising) economic and social factors (Dow Jones indexes, average prices, etc.)
- W_I is influential part (interpersonal relations and opinion exchange)

Methodology
Motivations in the simplest form

- Own advertising activities (n_0 channels)

$$W_0(\vec{z},t) = z^{\nu_0}\,\beta_0\left[1+\sum_{k=1}^{n_0} B_{0k}\; b_{0k}(t)\right]$$

- Competitor's advertising activities (n_C channels)

$$W_C(\vec{z},t) = (1-z)^{\nu_C}\,\beta_C\left[1+\sum_{k=1}^{n_C} B_{Ck}\; b_{Ck}(t)\right]$$

- General economic and social factors (n_F factors)

$$W_F(\vec{z},t) = \left|C_F - z\right|^{\nu_F}\,\beta_F\left[1+\sum_{k=1}^{n_F} B_{Fk}\; b_{Fk}(t)\right]$$

- Influential part in the second virial approximation

$$W_I = \beta_I G(z,t)$$

Methodology
Population Distribution Equations

Dynamic equation

- General form: $$-\frac{\partial G_t}{\partial t} = (1-\hat{Q})G_t$$

- Schrödinger-type equation for Gaussian model in D-dimensional position space:

$$\frac{\partial G_t}{\partial t} = -W_t \cdot G_t + \frac{a^2}{2D} \cdot \Delta_{\vec{z}} G_t$$

Normalization for total population:

$$\int G_t(\vec{q}_t, \vec{q}_0) d\vec{q}_t = N(t)$$

- Population $N(t)$ can be time dependent (grow or shrink)

Methodology
Links to Some Statistical Techniques

- Target variable (e.g. sales) *S(t)* and *m* factors: $\vec{b}(t) \equiv \{b_1(t), b_2(t), ..., b_m(t)\}$
  - Regression with lags: $S(t) = \theta(t) + \sum_{i=0}^{i_0} \left(\vec{C}_i \cdot \vec{b}(t-i) \right)$
    - Constants $\vec{C}_i$ can be interrelated (e.g. to obey gamma distribution)
  - Random coefficient mixed models (yield analysis *Demidenko E., Mandel I. (2005) Yield Analysis and Mixed Model. Joint Statistical Meeting Proceedings*)

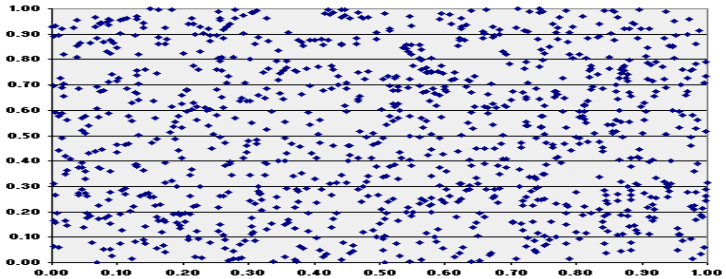
$$S(t) = \theta(t) + \sum_{i=0}^{i_0} \left([\vec{C}_i^0 + \delta\vec{C}_i(t)] \cdot \vec{b}(t-i) \right)$$

  - Mediaphysics approach:

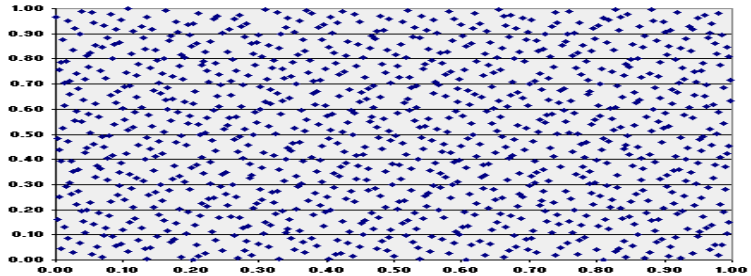
$$S(t) = \theta(t) + \sum_{i=0}^{i_0} \left([\vec{C}_i^0 + \delta\vec{C}_i(t)] \cdot \vec{b}(t-i) \right) + A \cdot \sum_{i=0}^{i_0} (\Delta_{\vec{z}} G_{t-1-i})_0$$

Methodology
Problems of statistical estimation

Quasi-Monte Carlo generation– an universal way to solve integration and optimization problems. The main idea is to create such a sequence of random points in a space of parameters to be estimated, that it is uniformly distributed within a space.



Monte Carlo, 1000 points - clusters



Quasi Monte Carlo, 1000 points – no clusters

•The basic idea of the key **van der Corput** one-dimensional sequence (1935) is to "fill the largest gapes" between already generated points;

•Later many multi-dimensional sequences were proposed. But a problem is that when dimension is raising, generated points become correlated. So called **Halton** sequences (above) degrades at D=12, **Fuare** at D=24. However, **Sobol** sequences shows no degradation signs up to D=260;

•Convergence of usual Monte Carlo is O~1/(N^0.5), whereas quasi Monte Carlo usually has O~1/N, where N is number of simulations, i.e. much faster;

•We are going to use Sobol sequences with some modern modifications togetehre with special methods of analyzing sensitivity of optimal solution

5.Mediaphysics: Implementation

Implementation
Short-, Long- and Infinite-term Effects

- *Short-term* represents immediate effects and effects during a few time steps of data observations
  - Duration of each time step depends on the phenomenon: e.g. hours, days, weeks or months
- *Long-term* represents accumulated effects during many time steps
- *Infinite-term* represents irreversibility of many social effects
  - According to Heraclitus: None can step into the same river twice precisely

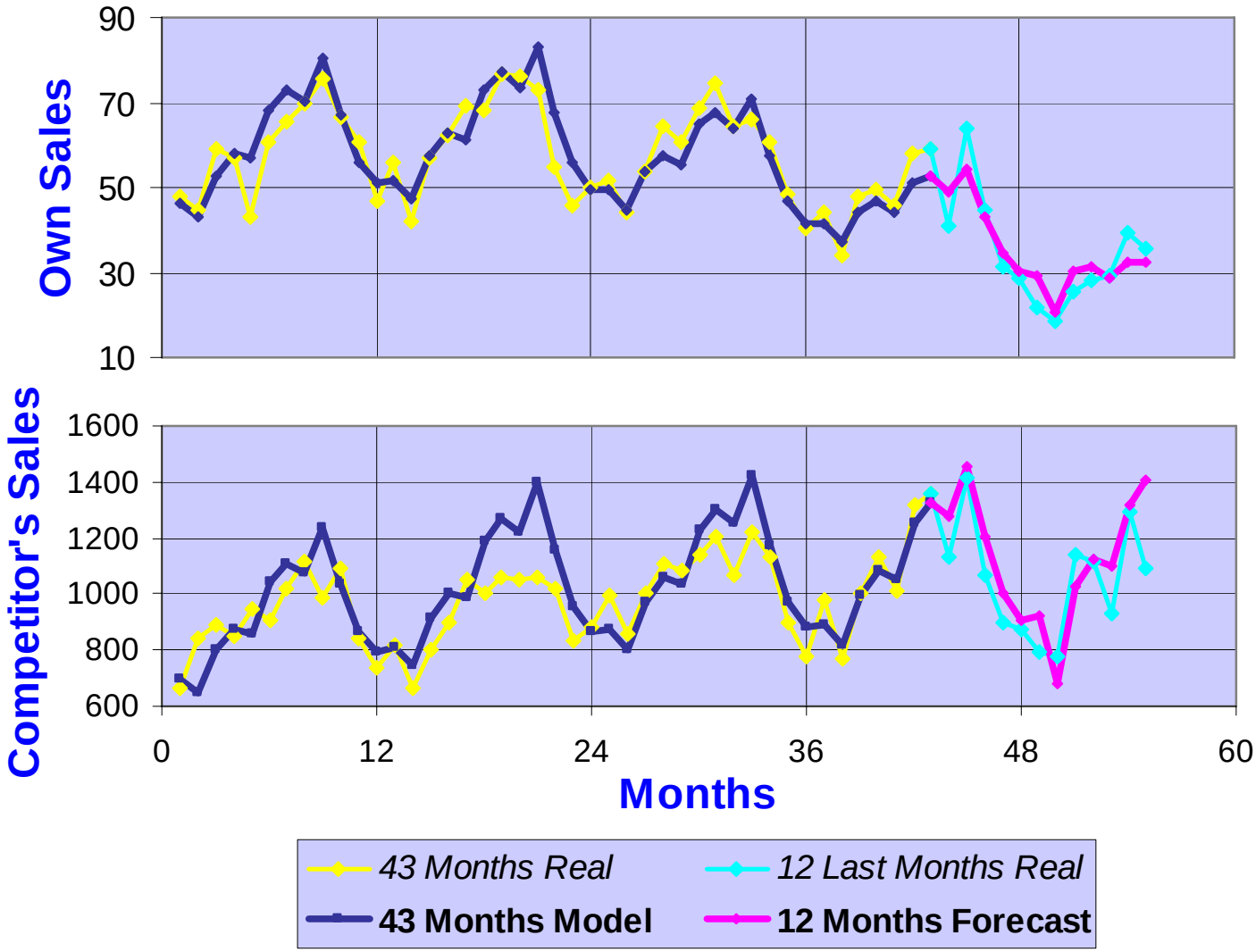
Implementation

Analytic & Numeric Calculations. Simulation & Optimization.

- Short-term effects can be analyzed *analytically*
- Long- and Infinite-term effects can be analyzed using *numeric calculations* (solution of the Mediaphysical equations)
- *Simulation* procedures can be incorporated in the methodology to reflect some aspects of complex realities
- *Optimization procedure* (dealing with non-linear behavior with many local extremes and completely non-analytical definition) has to be used for the best fitting of model

Implementation

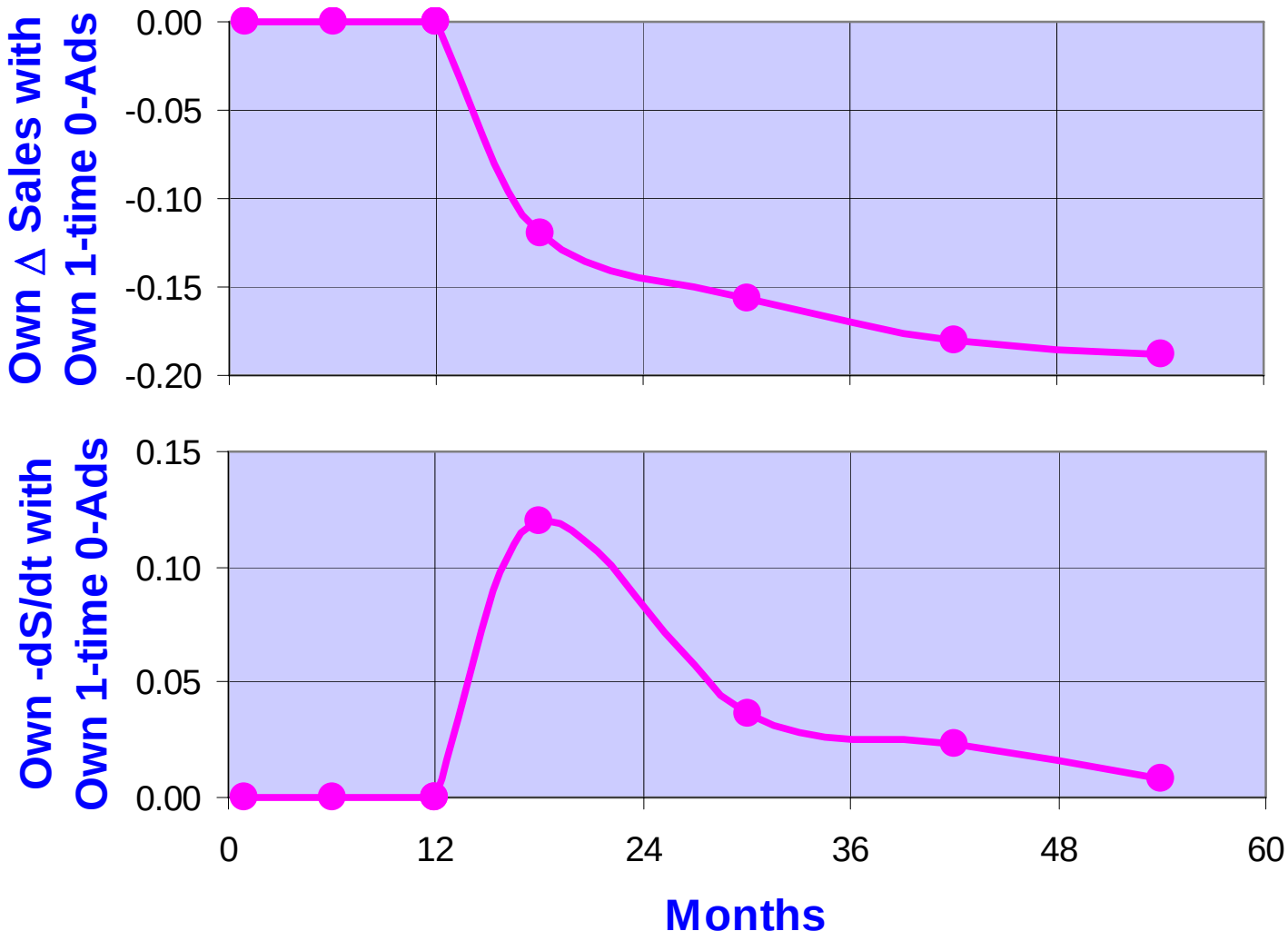
Real Data Model (Automobile Company) and Forecast for *Own and Competitor's sales*



Implementation

Real Data Model

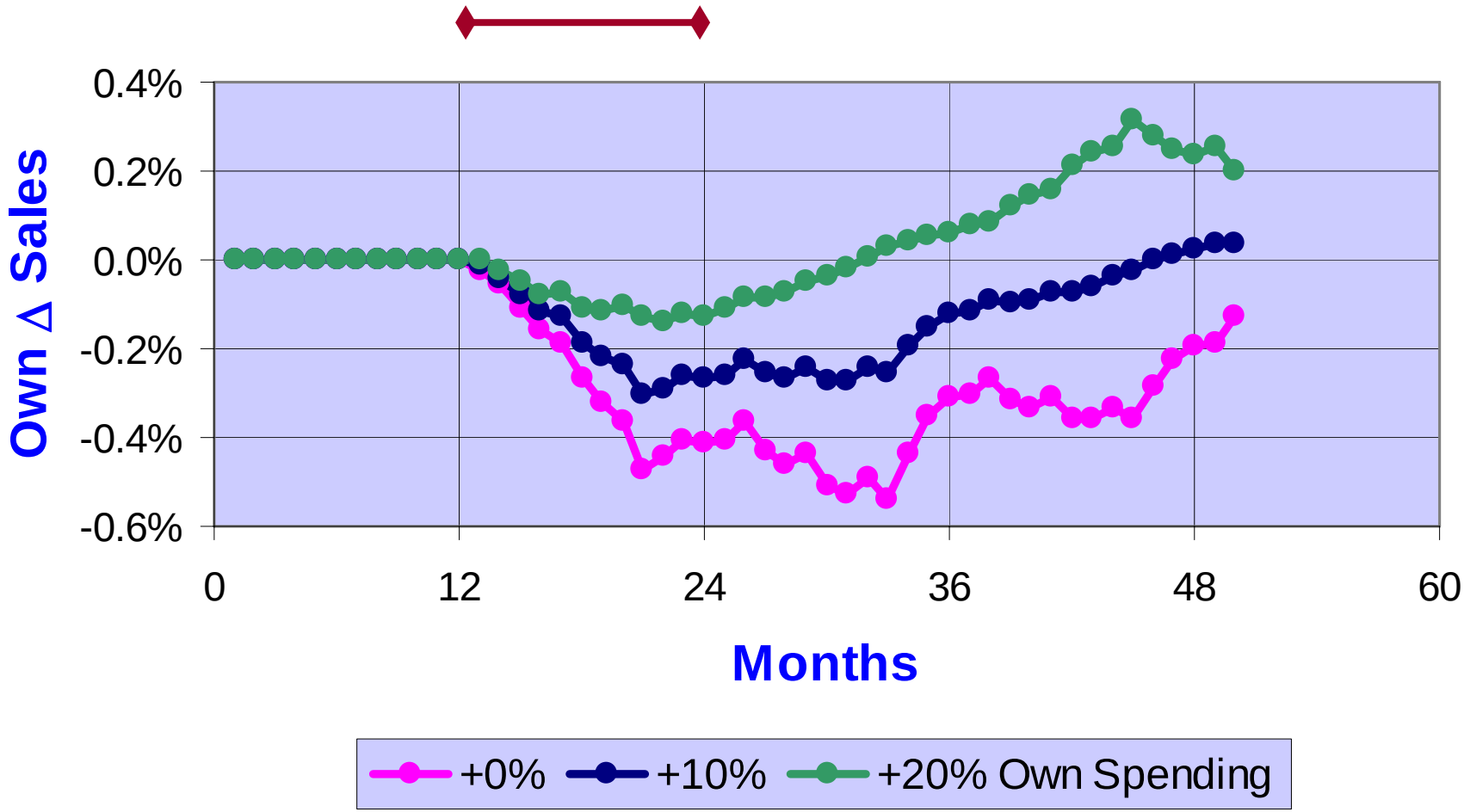
Long-term Effects of 1-time 0-Advertising



Implementation

What-If Analysis: *Monthly Effects*

+10% Competitor's Spending (for 13-24 months)



Implementation

What-If Analysis: *Cumulative Effects*

+10% Competitor's Spending (for 13-24 months)