

[simula . research laboratory]

Research on Software Cost Estimation

Stein Grimstad (steingr@simula.no)

Department of Informatics, University of Oslo, 5.11.2008

Agenda

- Basic information about software cost estimation research
- Experiments
- Surveys (discussion)
- Break
- Case-studies (discussion)

Terminology

- **Software development cost estimate** is understood as a prediction of the effort most likely required to implement a software development project
 - Produced by estimation processes
 - Probability of not exceeding the actual cost
 - Based on implicit/explicit assumptions
 - Contains uncertainty
 - Purpose is realism
- **Software development cost estimation error** is understood as a measurement of the difference between estimated and actual cost usage of a software development project

Why Study Software Cost Estimation?

- Software effort estimation is an essential part of software projects
 - Project feasibility analyses, bidding, budgeting, project steering and planning
- The consequences of inaccurate estimates can be dramatic
 - People get fired, companies fail to win contracts or lose money, products are delayed, and software of low quality
- Software projects commonly overrun their effort estimates
 - Average overrun is reported to be about 30% of estimated effort [*]
- I often encountered estimation related problems when I worked as a consultant
 - It is, e.g., often hard to know what clients want when they ask for estimates

[*] "A Review of Surveys on Software Effort Estimation", Moløkken and Jørgensen, ISESE, 2003

How Large is the Potential for Improvement of Software Cost Estimates?

- It is unrealistic to expect perfect cost estimates
 - Inherent uncertainty in software development projects
 - Complex and dynamic interaction of factors
- Still, it is likely that estimates can be improved
 - Cost estimates are overoptimistic and inconsistent
- However, even small improvements will be very valuable

Two Main Categories of Software Effort Estimation Methods

Expert estimation is typically used as a label for estimation methods in which a significant part of the estimation process (particularly the final step, i.e., the “quantification step”) is based on **intuition**

Formal estimation model is typically used as a label for estimation methods where a substantial part of the estimation (particularly the “quantification step”) is based on the use of **mechanical processes**, e.g., the use of a formula derived from historical data using regression analysis

Why Study Expert Estimation?

In our research, the main focus is on expert estimation

- Expert estimation is by far the most used estimation method in the software industry (70-80%) [*]
- Available evidence does not suggest that expert estimation should be replaced by formal estimation models [*]
- Most (meaningful) formal estimation models partly rely on expert judgment of input

[*] "Estimation of Software Development Work Effort: Evidence on Expert Judgment and Formal Models", Accepted for Int. J. of Forecasting, 2007. Jørgensen, 2007

Experiment: Estimation Irrelevant Information

- Other studies show that irrelevant information in the input to forecasting processes could be an important cause of inaccuracy in forecasting processes that relies on human judgment
- Requirement specifications and other information provided in software estimation situations typically include
 - a lot of estimation irrelevant information
 - some misleading information (on purpose or accidentally)
 - much information of low importance for the estimation work
- Several studies indicate that the quality of the requirement specifications impact estimation error, but at the time when we conducted these experiments, we were not aware of any study that investigate the impact of irrelevant information

EXPERIMENTS

Estimation Irrelevant Information

We conducted two experiments with software professionals as subjects (76/92 subjects) to investigate the impact of irrelevant information

- Group A received the original programming task specification
- Group B received the same specification, with neutral estimation irrelevant information about the end users desktop applications, web design, user passwords, future systems, etc. added
- Independent experts validated that they perceived our irrelevant information as irrelevant for estimation purposes

No estimation irrelevant information (Group A, experiment 2)

Simula organizes several seminars each year. At present, participants register for the seminars by sending an email to a given contact person.

Simula wants you to develop a simple web system (one web page) for registration of participants. The system will handle registration of all Simula's seminars. The participants will register on the web by submitting their email address and a registration code that uniquely identifies the seminar (this code is sent to them by email). The only functionality in the system is to store the email address and the registration code in a database. All queries will be done manually (in sql), and there is no need for any validation of submitted data. There are no security requirements.

The system will run on a webserver that has Tomcat, Java and MySql installed and running. Assume that you are familiar with the relevant technologies. You are free to use development tools of choice.

Estimation irrelevant information (Group B, experiment 2)

Simula organizes several seminars each year. At present, participants register for the seminars by sending an email to a given contact person.

Simula has ordered a web-based registration system. This system will let participants choose seminar from a list, and then register name, company and contact information. An email that confirms the registration will be sent to the participants. The system will keep track of the number of participants for each seminar and automatically close registration when the seminar is fully booked. There will also be a management module where employees at Simula can log in and query and manipulate conferences. This system will run on a Weblogic server application server, use a Sybase database and be fully integrated with Simula's other web pages. However, due to economical priorities, this system will not be developed before the end of 2006. Simula therefore needs an intermediate system for registration. This system will be thrown away when the new system is ready.

Simula wants you to develop a simple web system (one web page) for registration of participants. The system will handle registration of all Simula's seminars. The participants will register on the web by submitting their email address and a registration code that uniquely identifies the seminar (this code is sent to them by email). The only functionality in the system is to store the email address and the registration code in a database. All queries will be done manually (in sql), and there is no need for any validation of submitted data. There are no security requirements.

The system will run on a webserver that has Tomcat, Java and MySql installed and running. Assume that you are familiar with the relevant technologies. You are free to use development tools of choice.

Estimation Irrelevant Information (Paper I)

Results experiment 1:

Group	N	Mean	Median	Min	Max	Stdv
No irrelevant information	37	19.7	15.0	4	70	15.9
Irrelevant information	38	39.3	27.5	4	250	47.2

Results experiment 2:

Group	N	Mean	Median	Min	Max	Stdv
No irrelevant information	41	16.0	8	0.5	120	22.7
Irrelevant information	47	18.7	10	1.0	100	19.8

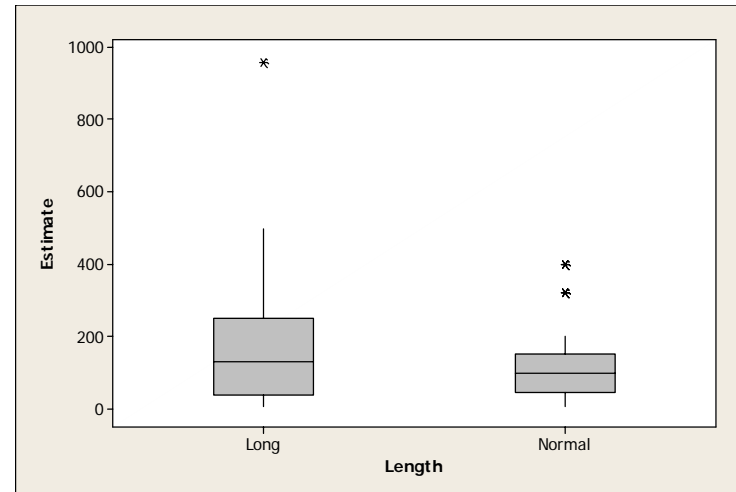
Note that in both experiments, irrelevant information increased the estimates. However, we have done other studies have shown that irrelevant information can lead to reduced estimates as well, e.g. by choosing words “loaded” towards small tasks.

Manipulate document size

- IFI-students estimated work-effort for the same programming task
 - Group A: Received the original specification, one page long
 - Group B: Received a version that had identical text, but was seven pages long. The line spacing was increased, the margins were wider, the font was larger and there were more space between the sections



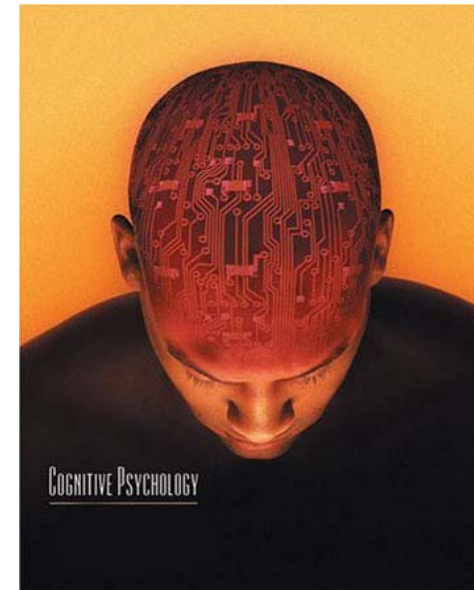
Results



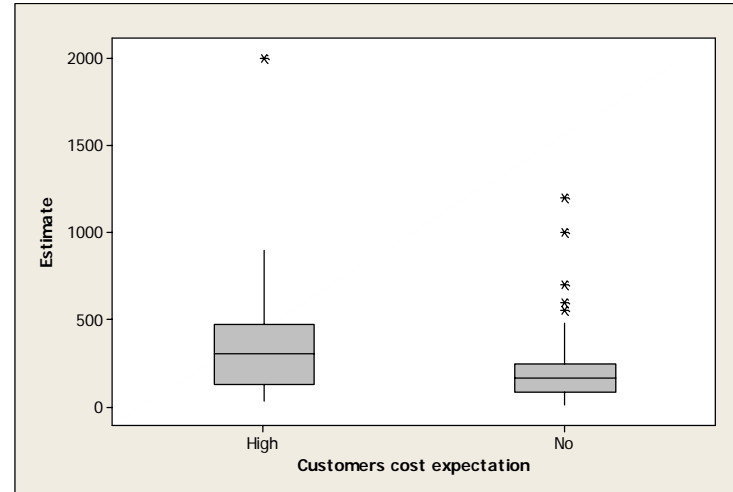
	Long	Normal	Difference
Mean	170	117	45%
StDev	173	98	77%

Information regarding client expectations

HIGH group: *“The customer has indicated that he believes that **1000 work-hours** is a reasonable effort estimate for the specified system. However, the customer knows very little about the implications of his specification on the development effort and you shall not let the customer’s expectations impact your estimate. Your task is to provide a realistic effort estimate of a system that meets the requirements specification and has a sufficient quality.”*



Results



	High	None	Difference
Mean	385	233	65%
StDev	376	242	55%

SURVEYS

Excercise

- Read the study design in the paper that I handed out (Section 3)
 - Identify shortcomings especially related to selection of participants

CASE STUDY

PREDICTION MARKETS

- ✓ Prediction markets is a tool that enables access to, and aggregation of, the estimation relevant information in the organization
- ✓ Prediction markets are implemented as electronic markets where all employees can participate, and the prizes are related to future events

The screenshot displays the Intrade website interface. At the top, the logo "intrade™ The Prediction Market" is visible, along with a quote: "When people ask me who will win the next election, I say 'Let me look at the price on Intrade. It is the best forecast I know.'" and the name "Koleman Strumpf, Professor of Economics, University of Kansas". There are fields for "User Name" and "Password", and links for "Live Help", "Login Help", and "Log I".

The main navigation bar includes "Home", "Signup", "Markets", "Trade", "Practice", and "Forum". Below this, a "READY TO START?" section features a green "Open An Account" button and a "Welcome Video" player with a "More Videos" link.

The "Introduction & Example Trade" section includes links for "Register For a Webinar", "Intrade News", and "PDA Interface".

The "MARKET SEARCH:" section has a search bar and a "Search" button. Below it, a prediction market is displayed for the question: "Will project AcmeWeb release to prod before 24.12.2008?".

The "Selected Prediction Markets" section shows the following details:

- Last Price: 61.2
- You can buy this at 61.2 (with a green "Buy" button)
- You can sell this at 61.1 (with a red "Sell" button)
- Closing Price: 61.4 (indicated by a black square)
- Buttons for "Explain" and "Trade Now"
- A line graph showing price movement from 2007 to 2008, with the closing price at 61.4.

Use of prediction markets

The market (question)

- A time-constrained question related to a future question or parameter value
- E.g. "When will our new ticket system be released?"

The contracts (outcomes)

- Each contract represents a unique outcome
- E.g. : A:[no later than week 46], B:[week 47-50], C:[after week 50]

The market place (application)

- A web application where the estimators can buy and sell contracts
- The market place generates buying and selling prices based on the trading
- The estimators buy whatever they find cheap, and sell when they get a good price

How is the contracts priced?

The price reflects the probability that the event related till the contract will occur

Contract prices (an example)

A: 16 kroner = 16% probable

B: 27 kroner = 27% probable

C: 57 kroner = 57% probable

Sum is always 100

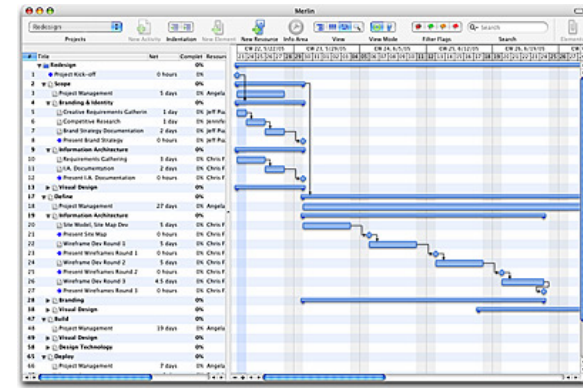
Prediction markets has been successfully applied by several IT organizations (Google, Siemens, IBM, ...)



#Fauls

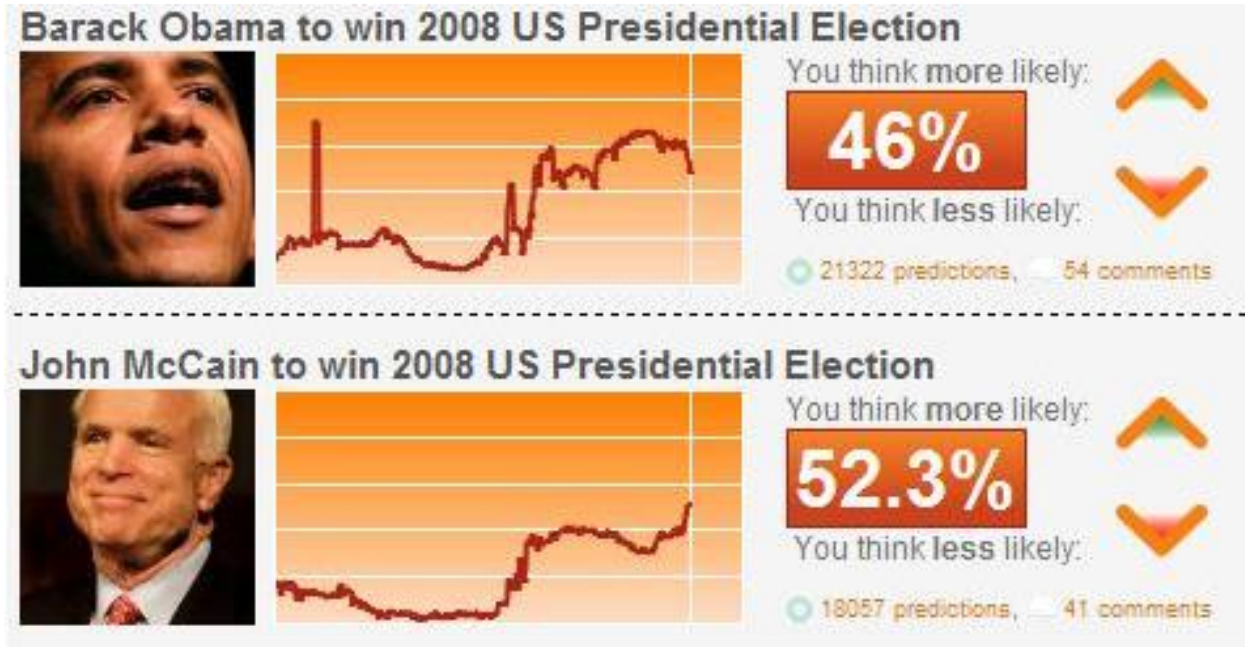


Cost estimates



Duration

...and for much else



OUR EXPERIENCES

We have investigated the use of prediction markets as an estimation tool in small, agile projects



1) Will the projects embrace the technology?

2) How do we use the technology efficiently?

3) Are the predictions accurate? (if so, why?)

Study 1 : ”# of errors found during system testing”

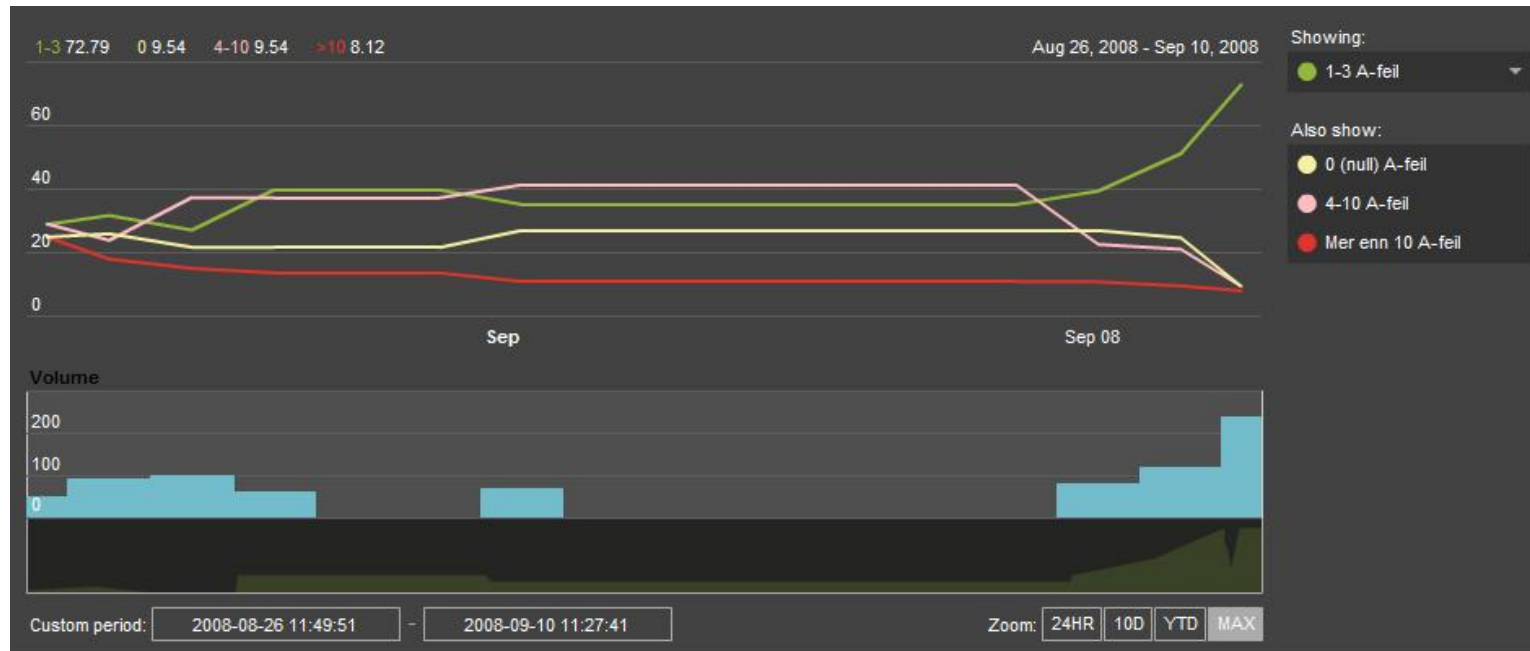
Project

- Part of a large project, 7 people
- Java- and web-technology
- Agile development (scrum)
- Duration : app. 2 years

Market

- ”How many A-bugs will be found during system testing”
- Four outcomes (0, 1-3, 4-10, more than 10)
- People: project leader and developers
- Duration: 3 weeks

Results



7 out of 7 traded

A total of 45 trades

The trades are done in small intervals

Study 2 : "Will the product be accepted?"

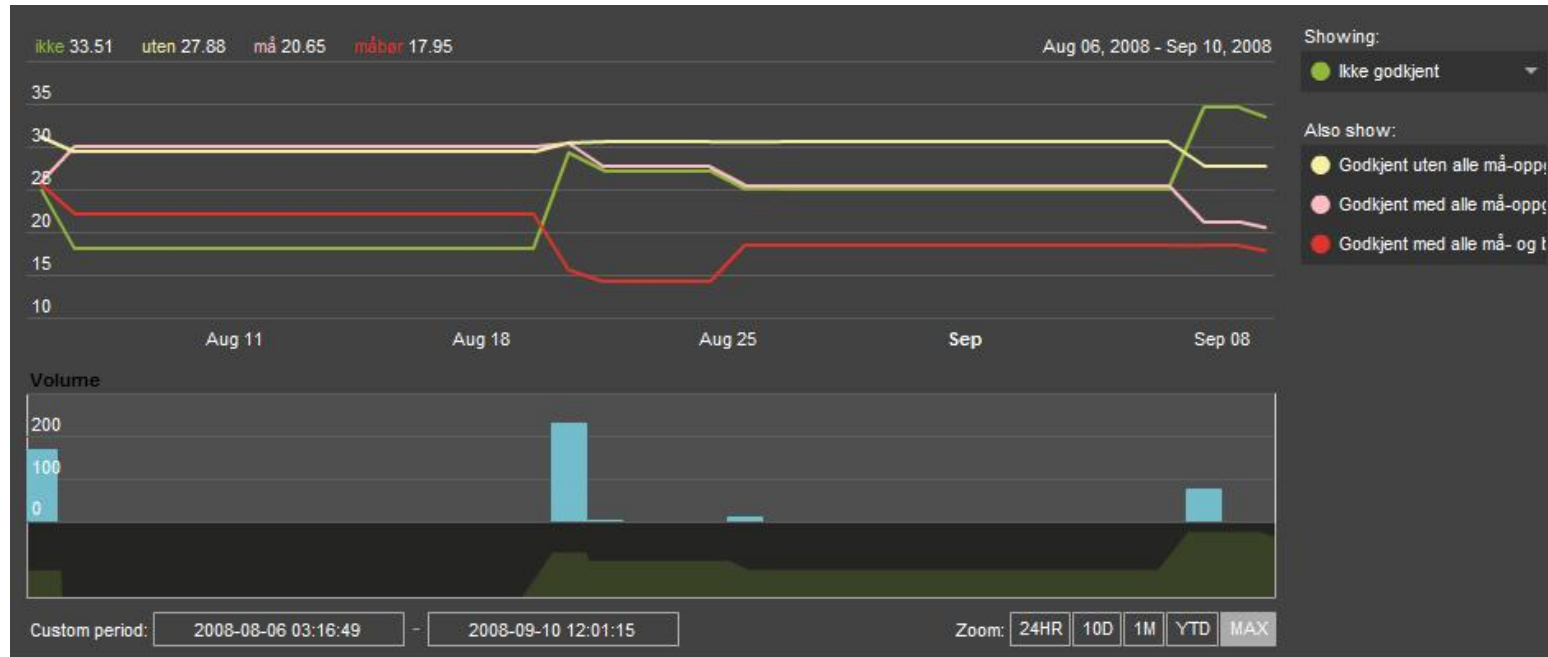
The Project

- Pay per hour, 8 people
- Ruby on Rails
- Agile (scrum)
- Duration: App. 6 months

The Market

- "On the agreed date for delivery: Will the product be accepted?"
- Four outcomes
- People: project leader and developers
- Duration: 8 weeks

Results



8 of 8 traded

App 50 trades (market still open)

The trades are done in small intervals

LESSONS LEARNT

Carefully define markets ("questions") and contracts ("outcomes")

The Market must be possible to settle

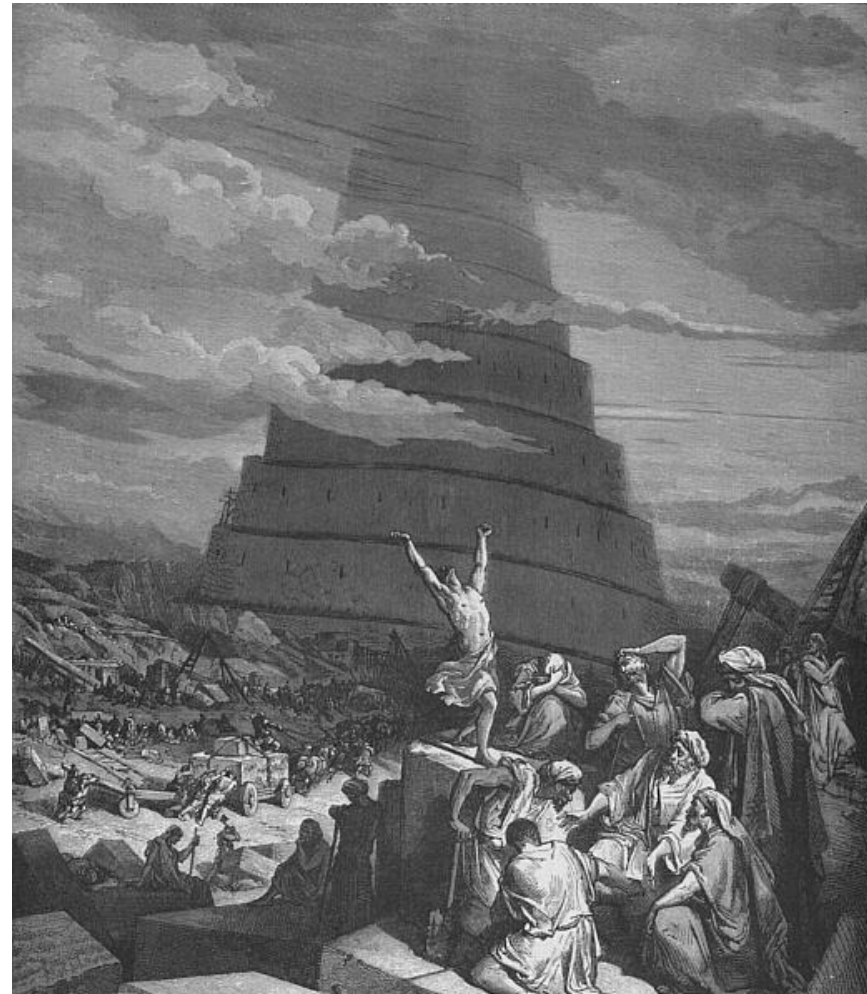
- NOT GOOD: Will age of Conan be a success?
- BETTER: Will Age of Conan have 400.000 paying customers by 01.01.2009"

The contracts must not overlap

- NOT GOOD: [Week 46], [46-50], [Week 50 or later]
- BETTER: [Week 46 or before], [week 47-50], [after week 50]

A precise terminology is important

- E.g.: What to do with errors that are wrongly reported?
- Imprecise terminology is one of the big problems in estimation (e.g. what is an estimate?)



Extensive training is needed

An instruction mail was not sufficient

- Several struggled to understand the concept
- The software is immature and far from easy to use

Several papers stress the importance of training

- Inexperiences people perform worse than those that have experience

Google reports positive experiences with using entertainment markets for training (e.g. "Who will win the WC in football?")



Incentives are important, but the "winner takes all" was a big mistake!

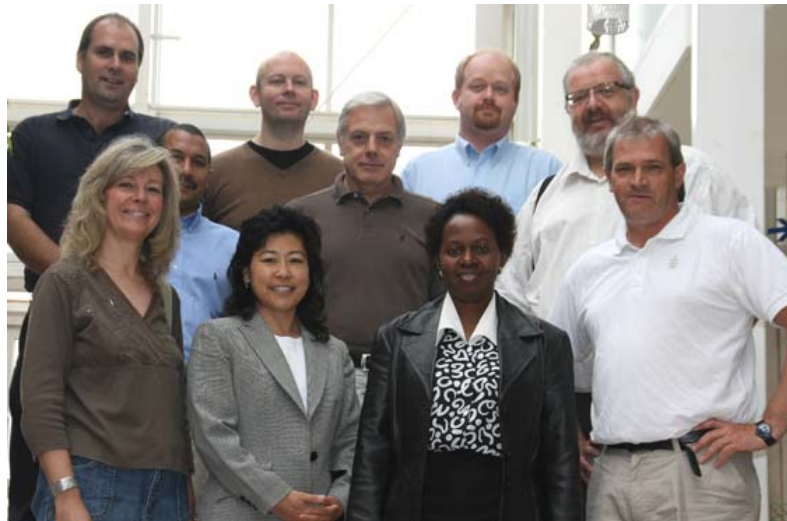
- People lost interest when they had small chances of winning
- The market can be destroyed when trades are not motivated by realism
- A large prize can impact performance

A lottery seems to be a good alternative to "the winner takes it all"



People with different background should participate

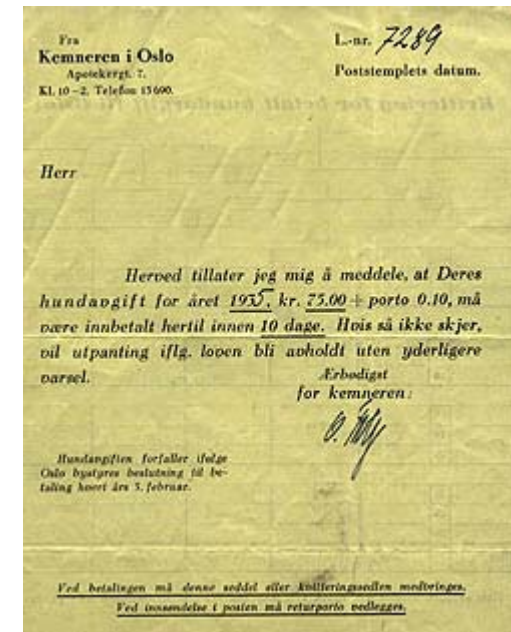
- All the people involved talked to each other / discussed the market
- Everyone had pretty much access to the same information. However, they used the information differently
- The market would probably have been better if the clients had participated.



Reminders are necessary!

In both markets, the trades were done in narrow time intervals

- The market was not often updated, this can make interpretation hard
- The market will be more random
- Several people asked us to send reminders



Several papers reports positive experience related to combining reminders with weekly status meetings

Potentially irrelevant factors can impact the trading

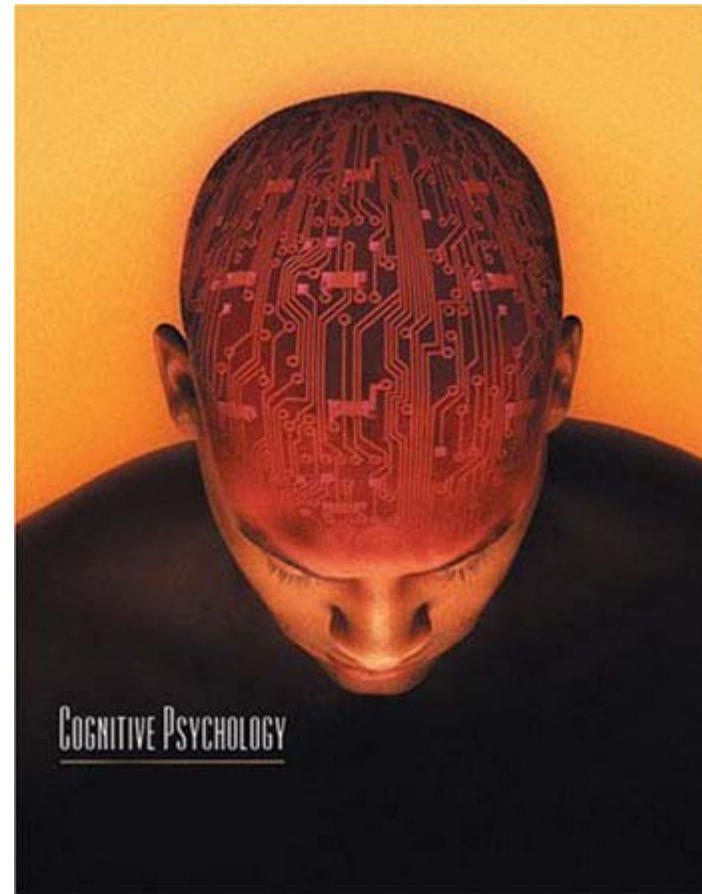
Over-optimism

Wishful thinking

Irrelevant context information

Framing

Location bias



Excercise

One hypothesis is that prediction markets work (partly) because people think differently when there is (own) money involved (play money or real money seems to make no difference).

Design an experiment that test this hypothesis.

Q & A ?