

# Does the prioritization technique affect stakeholders' selection of essential software product features?

Hans Christian Benestad\*

Jo Hannay\*\*

Expertware\*

Oslo, Norway

Simula Research Laboratory\*\*

Oslo, Norway

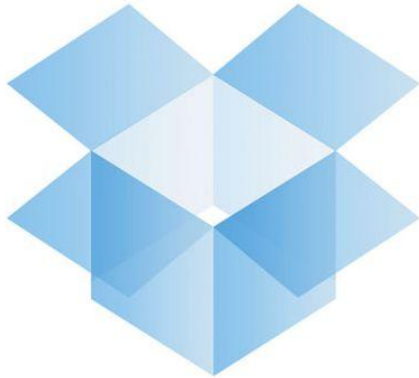
ESEM, Lund

Sep 20<sup>th</sup> 2012



# Features are *essential* when they define the success of a product

Organizations that consistently recognize *essentials* have a remarkable competitive advantage



I need to sync my files



I need to communicate effortlessly



I need to decode my TV-signals

# **Prioritization is inherently unreliable**

**Prioritization is an attempt to predict business value**

**Predicting business value is a type of estimation**

**Estimation heavily relies on human judgement**

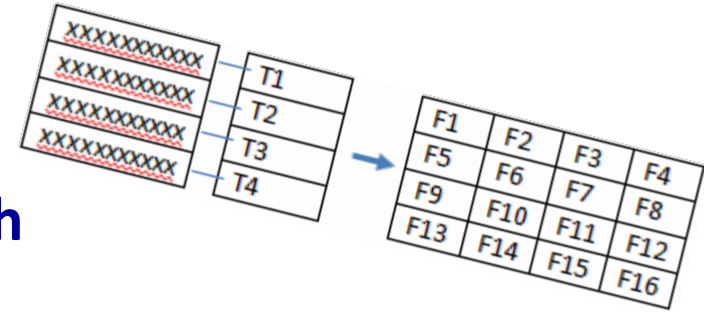
**Human judgement is inherently unreliable**

**Our research attempts to find the prevalent factors that bias priorities, and aims at guidelines to counteract those factors**

# In two controlled experiments, we compared the said number of essential features, using four prioritization techniques

Subjects were asked to prioritize 16 features

Four techniques were compared, one for each treatment group



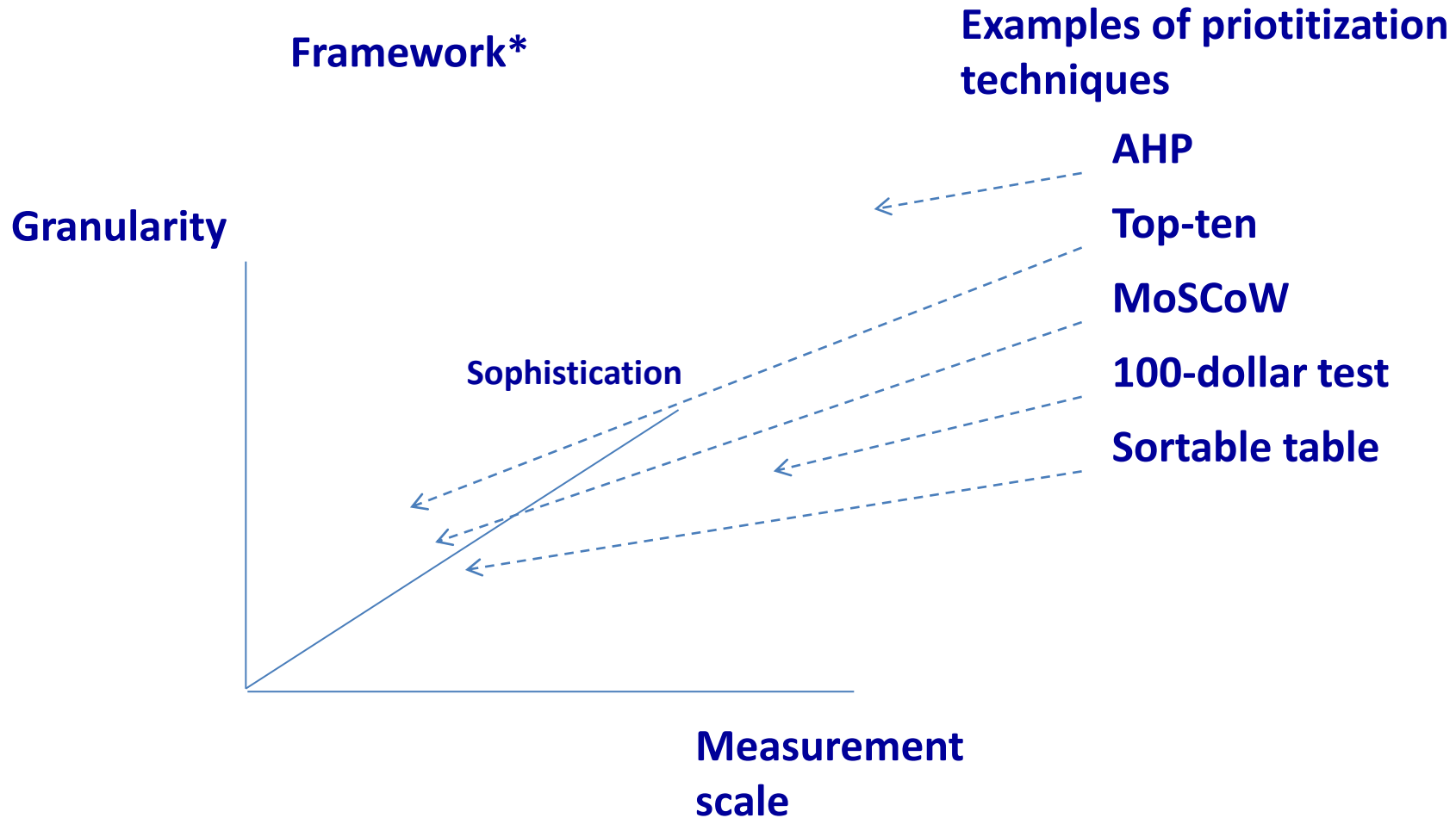
The definition of «essential» was kept constant

Field context: 44 business experts prioritized product queue elements in a large agile project

Artificial context: 94 research colleagues & business contacts prioritized cell phone features

Looking for robust, domain independent effects

# Prioritization techniques vary with respect to Measurement scale, Granularity and Sophistication



*\*Bernander & Andrews, 2005*

# Our research goal led us to focus on ordinal techniques, with two granularity levels and two levels of cognitive support

The question of essential/not essential is ordinal in nature



Ordinal techniques were sufficient

We wanted to know *whether or not* granularity affects the judgments



Two experimental levels were sufficient

We wanted to know *whether or not* cognitive support affects the judgments



Two experimental levels were sufficient

# Resulting in a 2x2 experimental design, investigating the effect of granularity and cognitive support on said number of essentials

Ordinal categories (2 vs 4)

X

Cognitive support (low vs high)



Number of essentials

	Low	High
4	T1: Simple dropdown	T2: Drag into bins
2	T3: Sort & pick	T4: Pairwise comparison & pick

*Categories*

*Cogn. support*

# Tool setup for T1: Low cognitive support, high granularity

## Simple Dropdown.

Categorize the features according to their influence on your buying decision for a new cell phone.

Wi-Fi	Insignificant- Does not influence on my buying decision
Physical QWERTY keyboard	Insignificant- Does not influence on my buying decision
Quality of speech better than average	Insignificant- Does not influence on my buying decision
Battery capacity better than average	Limited -Useful but I can buy a phone without it
Build quality better than average	Limited -Useful but I can buy a phone without it
Physical size smaller than average	Limited -Useful but I can buy a phone without it
E-mail client better than average	Essential - I would not buy the phone without it
Usability better than average	Essential - I would not buy the phone without it
Web browser better than average	Essential - I would not buy the phone without it
Display better than average	Significant - It would be difficult for me to buy a phone without it
Still camera better than average	Significant - It would be difficult for me to buy a phone without it
Storage capacity larger than average	Significant - It would be difficult for me to buy a phone without it
Video camera better than average	not set
GPS navigation	not set



# Tool setup for T2: High cognitive support, high granularity

## Drag into Bins.

Categorize the features according to their influence on your buying decision for a new cell phone.

Drag and drop the features into the categories to the right

- Display better than average
- Storage capacity larger than average
- GPS navigation

**Essential** I would not buy the phone without it

- Wi-Fi
- Physical QWERTY keyboard
- Quality of speech better than average
- Battery capacity better than average
- Still camera better than average
- Video camera better than average

**Significant** It would be difficult for me to buy a phone without it

- Build quality better than average
- Physical size smaller than average
- E-mail client better than average

**Limited** Useful but I can buy a phone without it

- Usability better than average
- Web browser better than average
- Availability of 3.party application better than average









**Insignificant** Does not influence on my buying decision

# Tool setup for T3: Low cognitive support, low granularity

## Sortable Table.

Rank the features according to their influence on your buying decision for a new cell phone.

Adjust the ranking by dragging features to a new position

1	 Wi-Fi	
2	 Physical QWERTY keyboard	
3	 Quality of speech better than average	
4	 Battery capacity better than average	<b>Essential</b>
5	 Build quality better than average	<b>Non-essential</b>
6	 Physical size smaller than average	
7	 E-mail client better than average	
8	 Usability better than average	

# Tool setup for T4: High cognitive support, low granularity

## Pairwise comparisons & Ranking.

Rank the features according to their influence on your buying decision for a new cell phone.

Indicate on a continuous scale from 1 to 9 how strong your preference is for one feature over the other. Click right arrow when you are done with all pairs

Adjust the ranking by dragging features to a new position

Wi-Fi 9.0 (Extreme) Physical QWERTY keyboard 2.2 (Moderate) 9.0 (Extreme)

Physical QWERTY keyboard 9.0 (Extreme) Quality of speech better than average 2.9 (Moderate) 9.0 (Extreme)

Quality of speech better than average 9.0 (Extreme) Battery capacity better than average 4.3 (Strong) 9.0 (Extreme)

Battery capacity better than average 9.0 (Extreme) Build quality better than average 3.1 (Moderate) 9.0 (Extreme)

Build quality better than average 9.0 (Extreme) Physical size smaller than average 1.0 (Equal) 9.0 (Extreme)



- 
- 
- 
- 
- Essential**
- Non-essential**
- 
- 
-

# Theoretical propositions

**A non-biased, rational result would indicate no difference between techniques on said number of essentials**

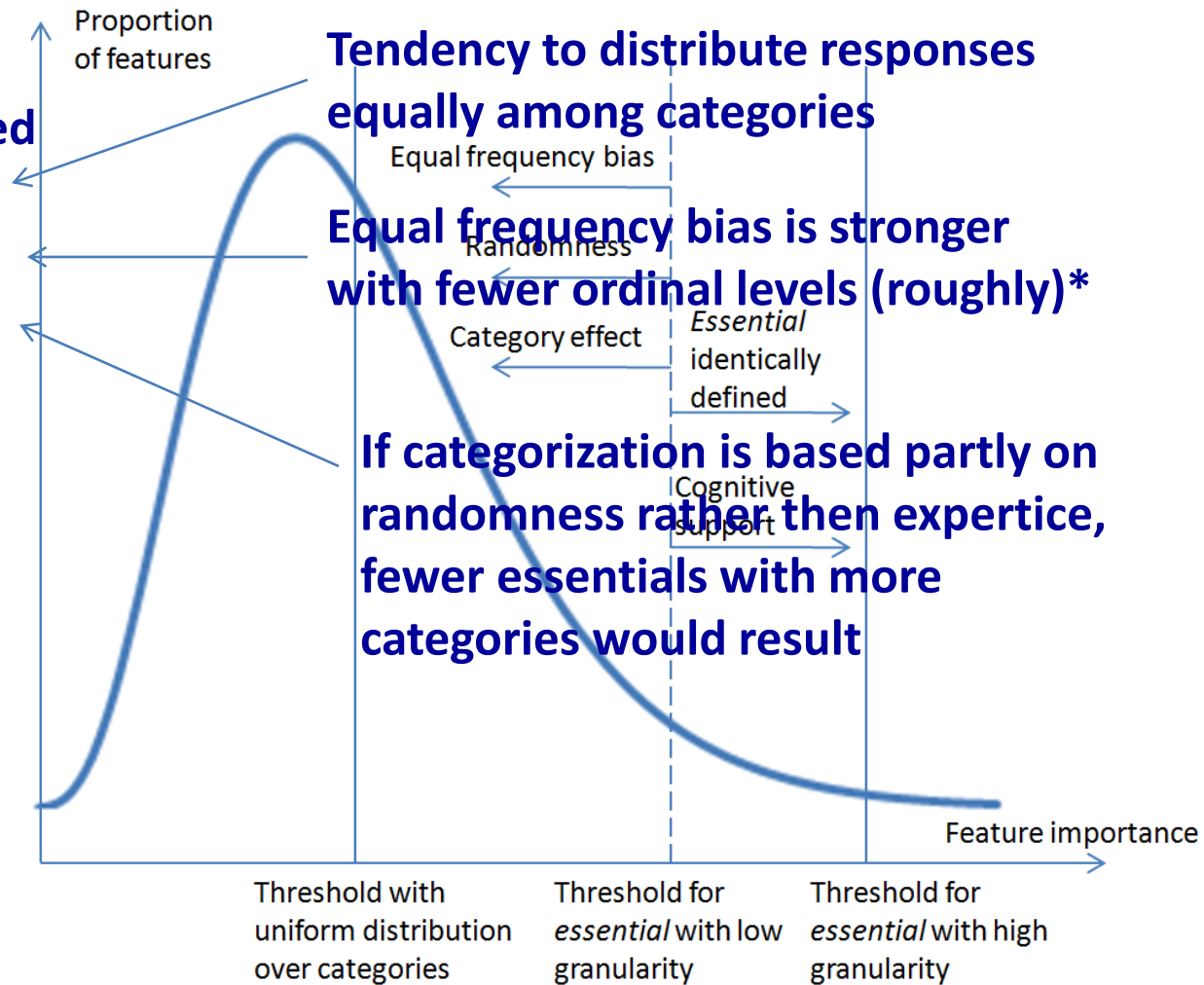
**Definition of «essential» identical across techniques:**

- **Artificial context: «I would not buy the phone without the feature»**
- **Field context «The product vision would not be fulfilled without the feature»**

# Known cognitive effects led us to expect differences

With more ordinal levels, fewer essentials were expected

- Equal frequency bias
- Category effect
- Pure randomness



Possibly counteracted by

- Stable definition of «essential»
- Increased cognitive support

\*Parducci&Wedell, 1996, J.Experimental Psychology

# We executed the experiment using a web-based prioritization tool

Flexibility in time and space

Certain validity threats must be controlled

Made a decently sized field experiment viable

The screenshot shows the EstimationWeb website. At the top left is the logo, a blue cube with a white plus sign, followed by the text "estimationweb" in a bold, sans-serif font and "advancing experts" in a smaller font below it. In the top right corner, there is a "Log out" button. Below the logo is a navigation bar with four items: "Home", "Agile planning" (which is highlighted with a mouse cursor), "Value and priority", and "Estimates". Underneath the navigation bar is a sub-header: "Schedule development and assess the uncertainty associated with the plan". Below this is a user login status: "You are logged in as: demo@estimationweb.com" and a row of small flags representing different languages. The main content area is divided into three columns. The left column contains two paragraphs of text. The middle column features a cartoon illustration of a man with a large, spiky hairstyle, wearing a suit and tie, with his hands raised in a gesture. The right column contains three stacked colored boxes with text: an orange box with "Value and priority of candidate features", a blue box with "Estimates of costs and effort involved in developing candidate features", and a green box with "Planning and uncertainty assessments, based on estimates and priority input".

estimationweb  
advancing experts

Log out

Home Agile planning Value and priority Estimates

Schedule development and assess the uncertainty associated with the plan

You are logged in as: demo@estimationweb.com

Developed with large software development projects specifically in mind, EstimationWeb is intended to be useful for planning any form of creative engineering endeavor which inevitably uses the judgment of experts as planning input.

By offering an optimal and consistent environment for collaborative expert knowledge elicitation, EstimationWeb is designed to improve three important steps in expert-based planning.

Value and priority of candidate features

Estimates of costs and effort involved in developing candidate features

Planning and uncertainty assessments, based on estimates and priority input

# Results



**In the artificial experiment, fewer categories and less cognitive support gave significantly more essentials**

Statistic	Overall	T1	T2	T3	T4
Mean	4.1	3.4	2.9	5.8	4.1
Stddev	2.7	1.8	1.8	3.1	2.9
Range	0-12	0-7	0-7	1-11	0-12

**Effect size Granularity: 0.73**  
**Effect size Cogn.support: 0.43**

	Df	Sum sq	Mean sq	Sq F value	Pr(>F)
Gr	1	7458	7458	11.18	0.0012**
Cogn	1	3812	3812	5.71	0.019*
gr:cogn	1	574	574	0.86	0.36
Residuals	90	60057	667		

**Two-way Anova with rank transformed response:**  
**Granularity**  
**Cognitive support: 0.43**

# In the realistic experiment, fewer categories gave dramatically more essentials

Statistic	Overall	T1	T2	T3	T4
Mean	5.8	3.0	2.9	7.8	9.50
Stddev	3.75	2.22	2.18	2.15	2.78
Range	0-14	0-7	0-7	5-10	5-14

Effect size Granularity: 2.40  
 Effect size Cogn.support: 0.50  
 (but opposite direction)

	Df	Sum sq	Mean sq	Sq F value	Pr(>F)
Gr	1	4420	4420	64.20	<0.001***
Cogn	1	55	55	0.80	0.38
gr:cogn	1	110	110	1.60	0.21
Residuals	40	2754	69		

Two-way Anova with rank transformed response:

Granularity  
 Cognitive support

# **In sum, strong evidence on the effect of granularity, not possible to conclude on the effect of cognitive support**

## **Why is the granularity effect much larger in a field setting?**

- **The software features might have a more heavily skewed importance distribution -> Larger «category effect»**
- **Subjects might know more about their cell phones preference than the system-to-be -> Larger randomness effect**

## **Why is the direction of the «cognitive support»-effect different between the experiments?**

- **I honestly don't know**

# External validity

**Statistical inference can be used to generalize to the specific populations investigated, but not automatically to other contexts**

**The two experiments combined indicate that the effects are indeed domain independent and possibly instances of robust psychological effects. Hence similar effects are likely to occur in other software projects.**

# Recommendations for project managers

Assuming that importance distribution of features is right skewed (few features are really essential), using more ordinal levels is safer

- «Essential» perceived more extreme

Assuming that importance distribution is left skewed (most features are actually essential), using less levels is safer

- «Essential» perceived less extreme

Use tools and group processes to help the team make reliable assessments

# Conclusion

- A decently scaled experiment has added to empirical evidence on prioritization bias
- Robust and large effect of granularity on said number of essentials
- Project managers can use the results directly



Questions