

# 1 Scientific roots and theories combined.



- Paul A. Samuelson: Revealed Preference Theory (Nobel Prize in Economics 1970)
- A. Michael Spence and Joseph Stiglitz: Market Signalling and Screening by Self-Selection (Nobel Prize in Economics 2001)
- Modelling as a “Gestaltproblem”. Causal models versus noise models.
- Ideas from ensemble methods:  
e.g. Albert Einstein’s work about the molecular Brownian motion. (Nobel prize in physics 1921).
- Ehrenberg’s repeat-buying theory.

# 1.1 Paul A. Samuelson: Revealed Preference Theory (Nobel Prize in Economics 1970)









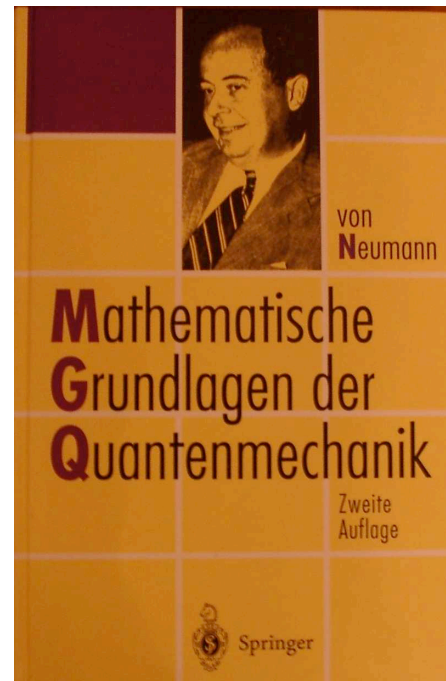


**Purchases reveal preferences.**

## 1.2 A. Michael Spence and Joseph Stiglitz: Market Signalling and Screening by Self-Selection (Nobel Prize in Economics 2001)

### Signalling (M. Spence):

- The choice of John von Neumann's Quantenmechanik (quantum mechanics) signals experience and expert knowledge.
- Incentive compatibility: voluntary and truthful.



### Screening (J. Stiglitz):

- A professor offers a menu of course books each of which requires a certain level of experience and expert knowledge.
- Selection of a book by a student reveals the student's level.
- Incentive compatibility: voluntary and truthful.

**Self-selection processes build homogenous clusters of readers.**

**Self-selection helps with the cold start problem.**

## 1.3 Modelling as a “Gestaltproblem”: Causal models versus noise models



Economics and management science versus natural sciences.

## 1.4 Ideas from ensemble methods.

Ensemble methods are methods which measure the behavior and the interactions of a large set of particles only by observing of the states of the aggregate (e.g. pressure, temperature and volume of a gas).

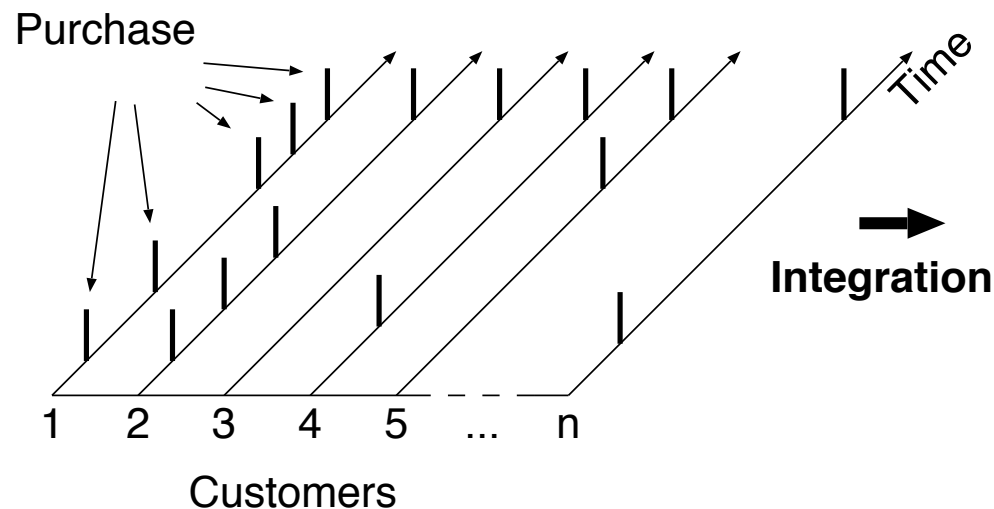
### Examples:

- Ludwig Boltzmann's Thermodynamics.
- Albert Einstein's work about the molecular Brownian motion.  
(Nobel prize in physics 1921).
- Ilya Prigogine: Nonequilibrium Statistical Mechanics.  
(Nobel prize in chemistry 1977).

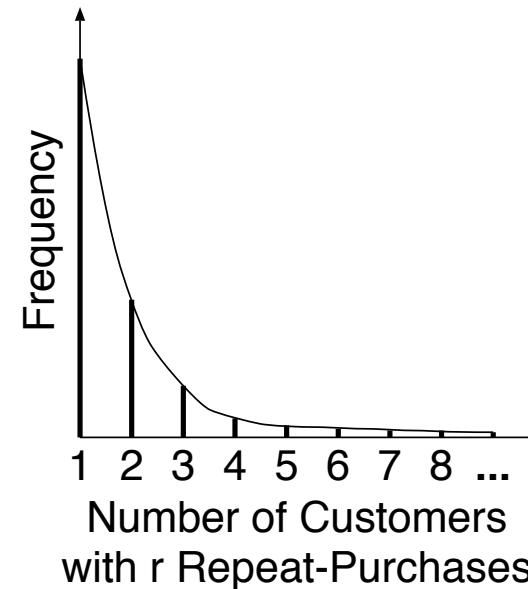


## 1.5 Ehrenbergs Repeat-Buying Theory.

**Independent Stochastic Purchase Processes  
for Product  $x$  or Productclass  $x$**



**Frequency Distribution of  
Repeat Purchases  
of Period  $t$**



**Independent Poisson processes act as models of noise and filters:  
Outliers are recommendations.**

## 2 Statistical models.

## 2.1 The LSD Model: A Mixture of Poisson Processes

The logarithmic series distribution describes, how many consumers have bought a specific product 1, 2, 3, ... times in the observation period (without taking non-buyers into account):

$$P(r \text{ purchases}) = \frac{-q^r}{r \ln(1-q)}, \quad r \geq 1 \quad (1)$$

Mean Purchase Frequency:

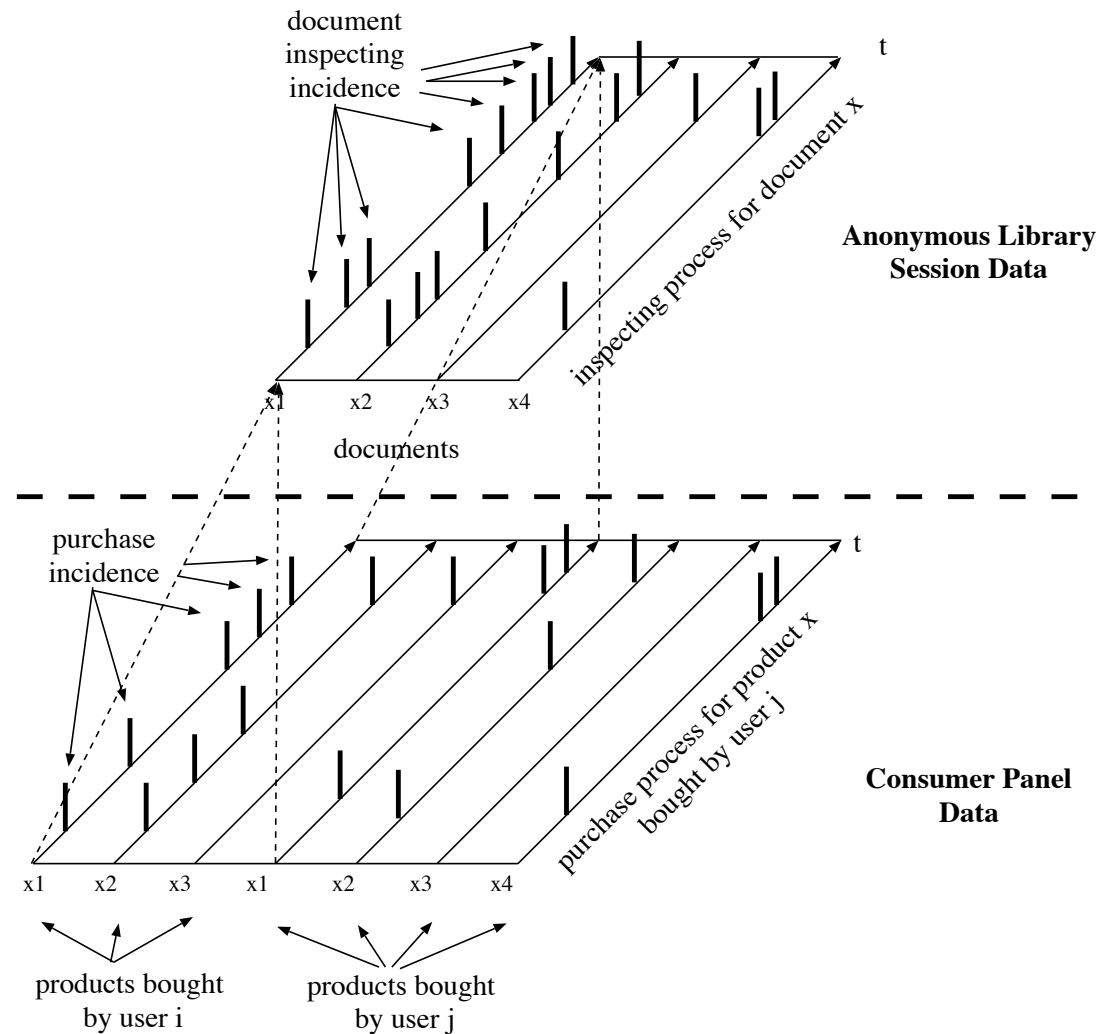
$$w = \frac{-q}{(1-q) \ln(1-q)} \quad (2)$$

Variance:

$$\sigma^2 = \frac{-q \frac{1+q}{\ln(1-q)}}{(1-q)^2 \ln(1-q)} \quad (3)$$



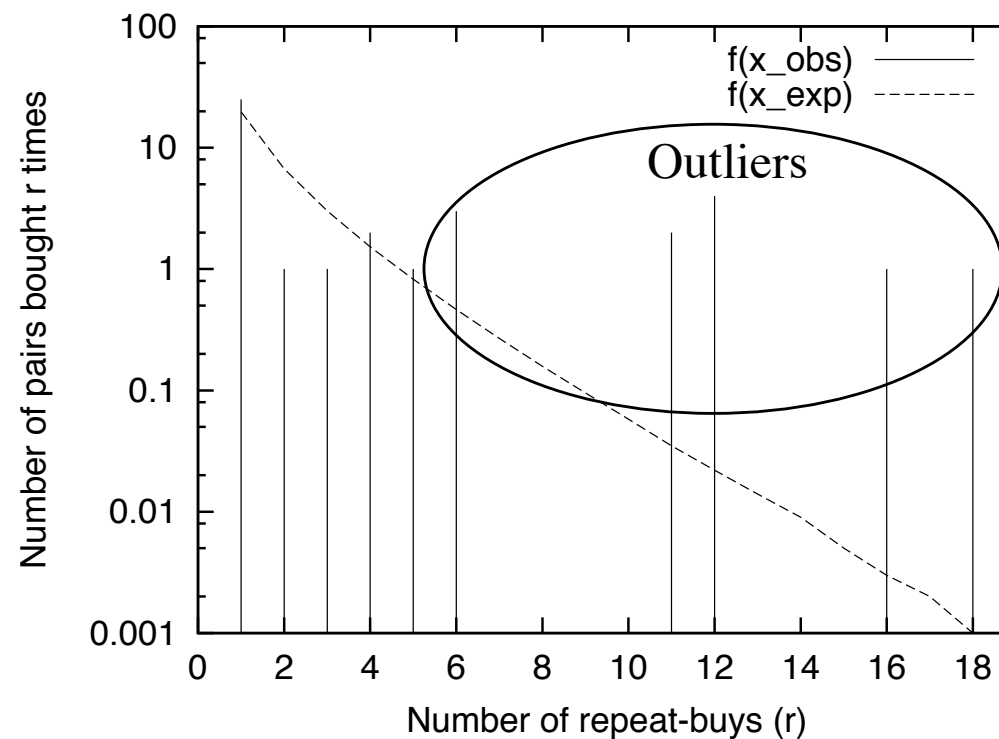
## 2.1.1 Stochastic Processes



From Consumer Panels to Anonymous Library Users

## 2.1.2 Generating Recommendations

- Products which are bought together more frequently than expected by the stochastic model
- violate independence assumptions of the model



Observed Distribution Function

### 3 The Cold Start Problem: 3 Routes of Attack

1. **Forecasting:** User Profile, User History, ...

**Problems:** New Users and new books.

2. **Small Sample Statistics:** Tests based on multinomial distributions (Geyer-Schulz, Neumann)

**Problem:** General solution depends on efficient computation of the partition function (number theoretic problem).

3. **Boosting:** Provide the recommendations of a similar organization.

Self-selection leads to homogenous clusters of users within an organisation.

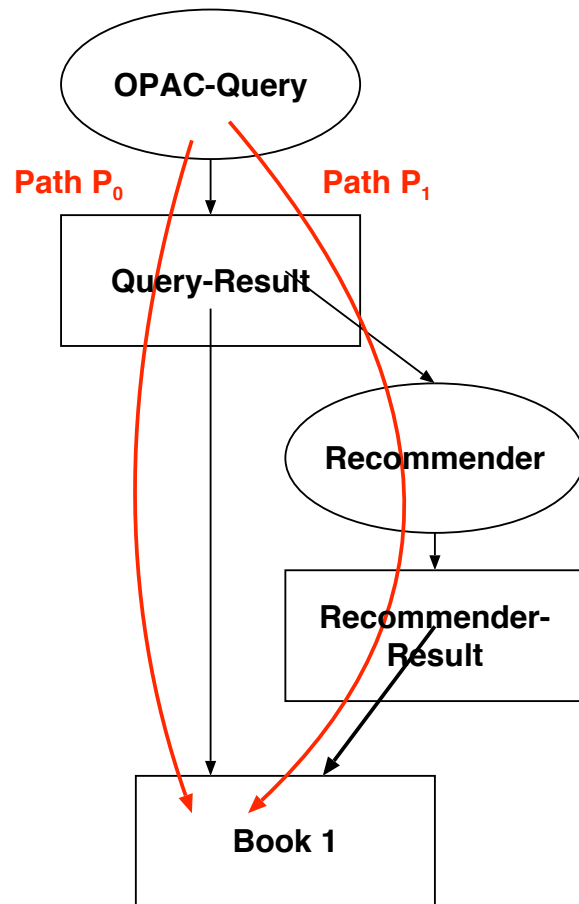
Homogenous clusters of scientific users are comparable across scientific organisations of similar type.

The use recommendations of a comparable scientific organization is a way a recommendation service provider can attack the cold start problem for new organisations.

---



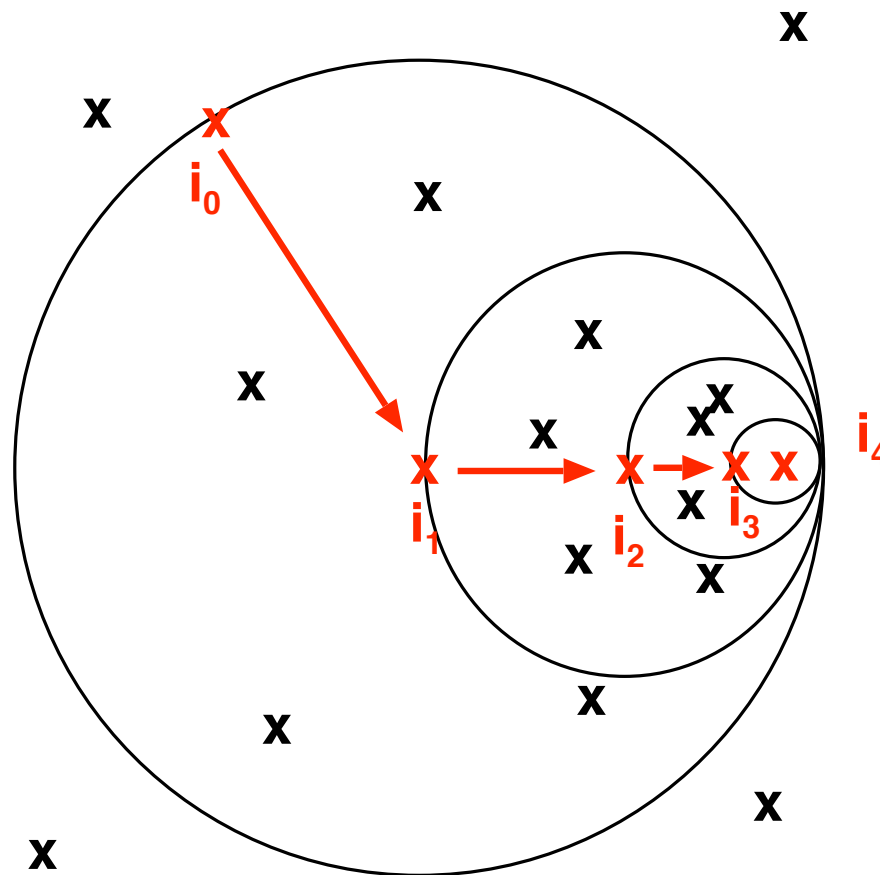
### 3.1 Boosting: University of Heidelberg by KIT



$P_0$	$P_1$	Count Before	Count After Boosting
1.0	0.0	22881	19944
0.9	0.1	817	251
0.8	0.2	416	398
0.7	0.3	230	352
0.6	0.4	57	81
0.5	0.5	163	592
0.4	0.6	44	118
0.3	0.7	9	31
0.2	0.8	6	17
1.0	1.0	50	1196

- Boosting leads to the activation of unused books.
- Boosting is a road to collection management.
- Boosting with recommendations of nearby libraries (e.g. KIT-BLB).

## 4 Restricted random walk clustering



## 5 Adaptive recommender systems

### Recommendations for

**Finding groups in data - an introduction to cluster analysis / Leonard Kaufman ; Peter J. Rousseeuw (1990)**

*Few precise hits*



*Many, but less precise hits*

- (75) Robust regression and outlier detection / Peter J. Rousseeuw ; Annick M. Leroy (1987)
- (66) Cluster analysis and data analysis / M. Jambu and M.-O. Lebeaux (1983)




UB Karlsruhe - Empfehlungen - Mozilla Firefox

Datei
Bearbeiten
Ansicht
Gehe
Lesezeichen
Extras
Hilfe



Universität Karlsruhe (TH)  
Universitätsbibliothek



Home
|
Mein Konto
|
Kataloge
|
Digitale Bibliothek
|
Liefersdienste
|
Fachgebiete
|
Infos

Recommendations for

**Automatische Klassifikation - theoretische und praktische Methoden zur Gruppierung und Strukturierung von Daten / von Hans Hermann Bock (1974)**

Few precise hits

Many, but less precise hits

(83) Clustering algorithms / John A. Hartigan (1975)

(83) Clusteranalyse - anwendungsorientierte Einführung / Johann Bacher (1996)

(83) Mathematical classification and clustering / Boris G. Mirkin (1996)

(80) Forecasting economic time series / Michael P. Clements and David F. Hendry (1998)

(77) Data analysis - scientific modeling and practical application ; with 45 tables / Wolfgang Gaul ... (eds.) (2000)

(70) Cluster analysis for applications / Michael R. Anderberg (1973)



developed by  
Schroff-Stiftungslehrstuhl für  
Informationsdienste und elektronische Märkte

Gefördert von der  
Deutschen Forschungsgemeinschaft



Recommendations for

**Finding groups in data - an introduction to cluster analysis / Leonard Kaufman ; Peter J. Rousseeuw (1990)**

*Few precise hits*

*Many, but less precise hits*

- (75) Robust regression and outlier detection / Peter J. Rousseeuw ; Annick M. Leroy (1987)
- (66) Cluster analysis and data analysis / M. Jambu and M.-O. Lebeaux (1983)
- (50) Clusteranalyse - anwendungsorientierte Einführung / Johann Bacher (1996)
- (50) Empirical methods for artificial intelligence / Paul R. Cohen (1995)
- (50) Clustern mit Hintergrundwissen / Andreas Hotho (2004)
- (50) Bayeslösungen des Ausreißerproblems / Friedrich Gebhardt (1961)
- (50) Ausreisser bei ein- und mehrdimensionalen Wahrscheinlichkeitsverteilungen / Rudolf Mathar (1981)
- (33) Clustering algorithms / John A. Hartigan (1975)
- (33) Mathematical classification and clustering / Boris G. Mirkin (1996)
- (33) Data analysis - scientific modeling and practical application ; with 45 tables / Wolfgang Gaul ... (eds.) (2000)
- (33) Modern regression methods / Thomas P. Ryan (1997)
- (33) Untersuchung zur zeitlich-räumlichen Ähnlichkeit von phänologischen und klimatologischen Parametern in Westdeutschland u / von Xiaoqiu Chen (1994)
- (33) Cluster analysis / Brian Everitt (1974)
- (33) Social Science Research Council / Social Science Research Council ()
- (25) New approaches in classification and data analysis / E. Diday ... (eds.) (1994)
- (25) Clusteranalyse - Einführung in Methoden und Verfahren der automatischen Klassifikation ; mit zahlreichen Algorithmen, FO / Detlef Steinhausen ; Klaus Langer (1977)
- (25) Tests und Schätzungen in Ausreißermodellen / Ursula Gather (1984)
- (25) Concurrence probabilities for a locally slotted packet radio network by combinatorial methods / Rudolf Mathar (1990)
- (25) Classification and dissimilarity analysis / Bernard Van Cutsem (ed.) (1994)
- (25) Fallstudien Cluster-Analyse / Helmuth Späth (1977)
- (11) Mastering data mining - the art and science of customer relationship management / Michael J. A. Berry ; Gordon Linoff (2000)
- (11) Entwicklung von Kundenbeziehungen - theoretische und empirische Analysen unter dynamischen Aspekten / Dominik Georgi (2000)
- (11) Kundenwert - Grundlagen - innovative Konzepte - praktische Umsetzungen / Bernd Günter ... (Hrsg.) (2003)
- (11) Relationship Marketing - das Management von Kundenbeziehungen / von Manfred Bruhn (2001)
- (11) Customer-Lifetime-Value-Management - Kundenwert schaffen und erhöhen: Konzepte, Strategien, Praxisbeispiele / Markus Hofmann ... (Hrsg.) (2000)
- (11) Kundenwertmanagement - Konzept zur wertorientierten Analyse und Gestaltung von Kundenbeziehungen / Gunter Eberling. Mit einem Geleitw. von Günter Specht (2002)
- (11) Den Kundennutzen managen - so beschreiten sie den Weg zur Wertschöpfungskette / Harald Münzberg (1995)
- (8) Scheduling theory / Tanaev, Vjaceslav S. (1994)

## 6 Exploration of the information space



## Literatur

- [Geyer-Schulz et al., 2003a] Geyer-Schulz, A., Hahlser, M., Neumann, A., and Thede, A. (2003a). Recommenderdienste für wissenschaftliche Bibliotheken und Bibliotheksverbünde. In Geyer-Schulz, A. and Taudes, A., editors, *Informationswirtschaft – ein Sektor mit Zukunft*, volume P-33 of *GI-Edition – Lecture Notes in Informatics (LNI)*, Bonn. Köllen.
- [Geyer-Schulz et al., 2003b] Geyer-Schulz, A., Hahsler, M., Neumann, A., and Thede, A. (2003b). Behavior-based recommender systems as value-added services for scientific libraries. In Bozdogan, H., editor, *Statistical Data Mining & Knowledge Discovery*. Chapman & Hall / CRC.
- [Geyer-Schulz et al., 2003c] Geyer-Schulz, A., Hahsler, M., Neumann, A., and Thede, A. (2003c). An integration strategy for distributed recommender services in legacy library systems. In Schader, M., Gaul, W., and Vichi, M., editors, *Between Data Science And Applied Data Analysis*, Studies in Classification, Data Analysis, and Knowledge Organization, Heidelberg-Berlin. Springer.
- [Geyer-Schulz et al., 2003d] Geyer-Schulz, A., Neumann, A., Heitmann, A., and Strobörn, K. (2003d). Strategic positioning options for scientific libraries in markets of

scientific and technical information – the economic impact of digitization. *Journal of Digital Information*, 4(2).

[Geyer-Schulz et al., 2003e] Geyer-Schulz, A., Neumann, A., and Thede, A. (2003e). An architecture for behavior-based library recommender systems. *Information Technology and Libraries*, 22(4):165–174.

[Geyer-Schulz et al., 2003f] Geyer-Schulz, A., Neumann, A., and Thede, A. (2003f). Others also use: A robust recommender system for scientific libraries. In Koch, T. and Sølvsberg, I. T., editors, *Research and Advanced Technology for Digital Libraries*, volume 2769 of *Lecture Notes in Computer Science*, pages 113 – 125, Berlin Heidelberg New York. Springer.

[Neumann, 2007] Neumann, A. W. (2007). Motivating and supporting user interaction with recommender systems. In Kovács, L., Fuhr, N., and Meghini, C., editors, *Research and Advanced Technology for Digital Libraries*, LNCS 4675, pages 428–439, Berlin Heidelberg New York. Springer.

[Neumann and Geyer-Schulz, 2007] Neumann, A. W. and Geyer-Schulz, A. (2007). Applying small sample test statistics for behavior-based recommendations. In *Proceedings of the 31st Annual Conference of the German Classification Society*. Springer. To appear.