

Introduction to Information Retrieval (36662)

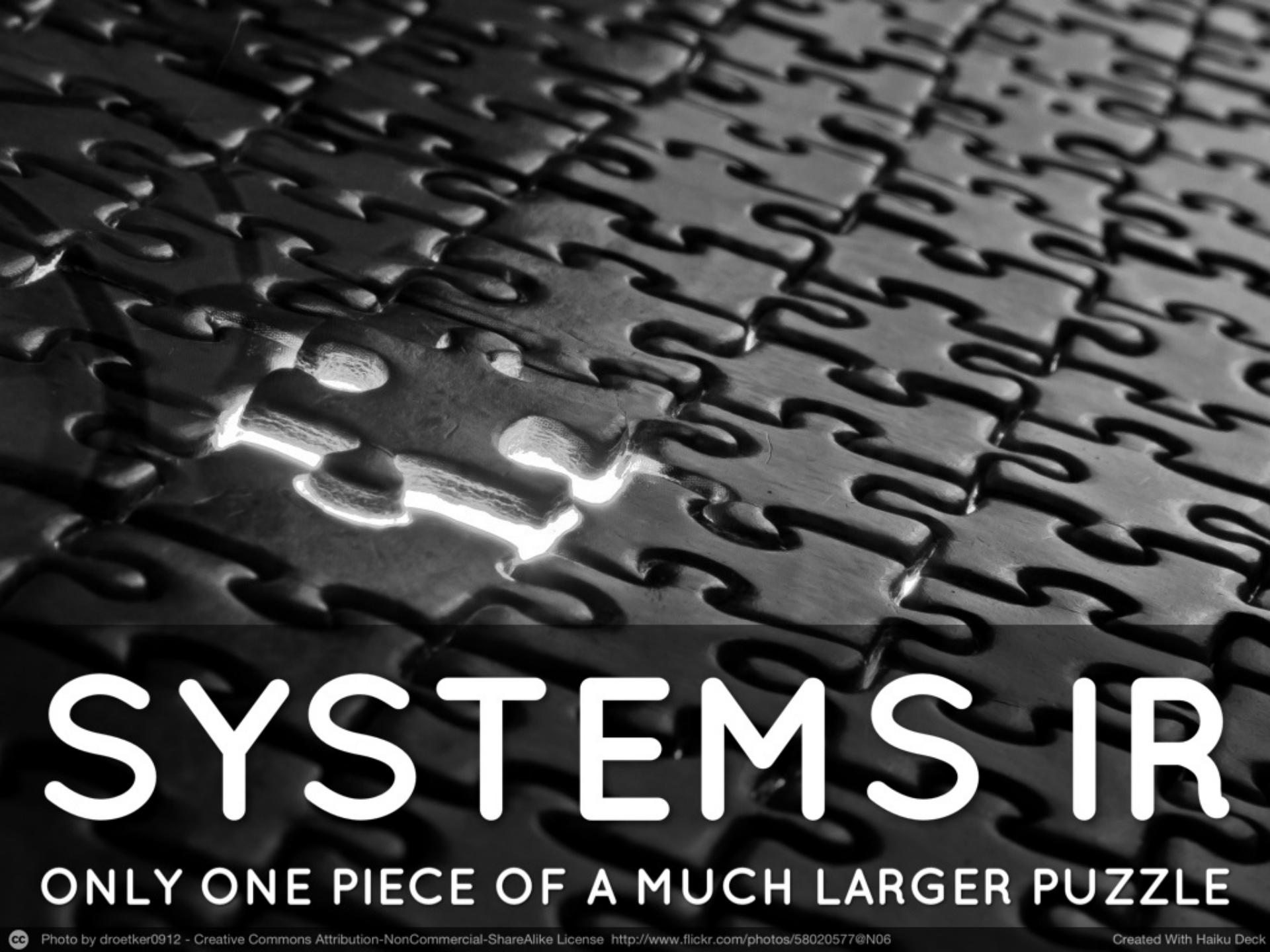


Universität Regensburg

Dr. David Elsweiler
Chair for Information Science
Faculty of Language, Literature
and Cultural Sciences
David.Elsweiler@sprachlit.uni-regensburg.de

Summarise Systems IR approach

- So far we have simplified IR to a particular scenario
- Adhoc Retrieval (1 shot retrieval, fixed relevance judgements, fixed document collection, fixed scenarios, fixed queries)
 - Easy to control
 - Easy to repeat
 - Provide us with a sound, cost effective way to test retrieval algorithms
- Hinted last time that we may not be measuring the thing we really want measure!



SYSTEMS IR

ONLY ONE PIECE OF A MUCH LARGER PUZZLE

Aims for today

- See the bigger picture
- Spot and understand some of the limitations of the systems approach (at least in isolation)
- Introduce some important models Information Seeking
- To become aware of some of the other aspects that need to be attended to when building and evaluating an IR system

Expectations from you!

- Use what you learn today to reevaluate what you know so far
 - How do you behave? What support do you need?
 - What do you think about the system's approach we have been looking at so far. What are the strengths and limitations? How do the various models change your thinking?
 - How do these models compare? Which are helpful and which are not?
 - We will discuss these at the end!

What is Information Seeking?

- Human Focused
- Built on observing / studying human users
- Often performed by Library and Information Scientists rather than Computer Scientists
- Is becoming increasingly important to understand how to improve IR
 - Retrieval model performance has plateaued in the last 10 years
 - Other aspects of the system need to be understood and improved

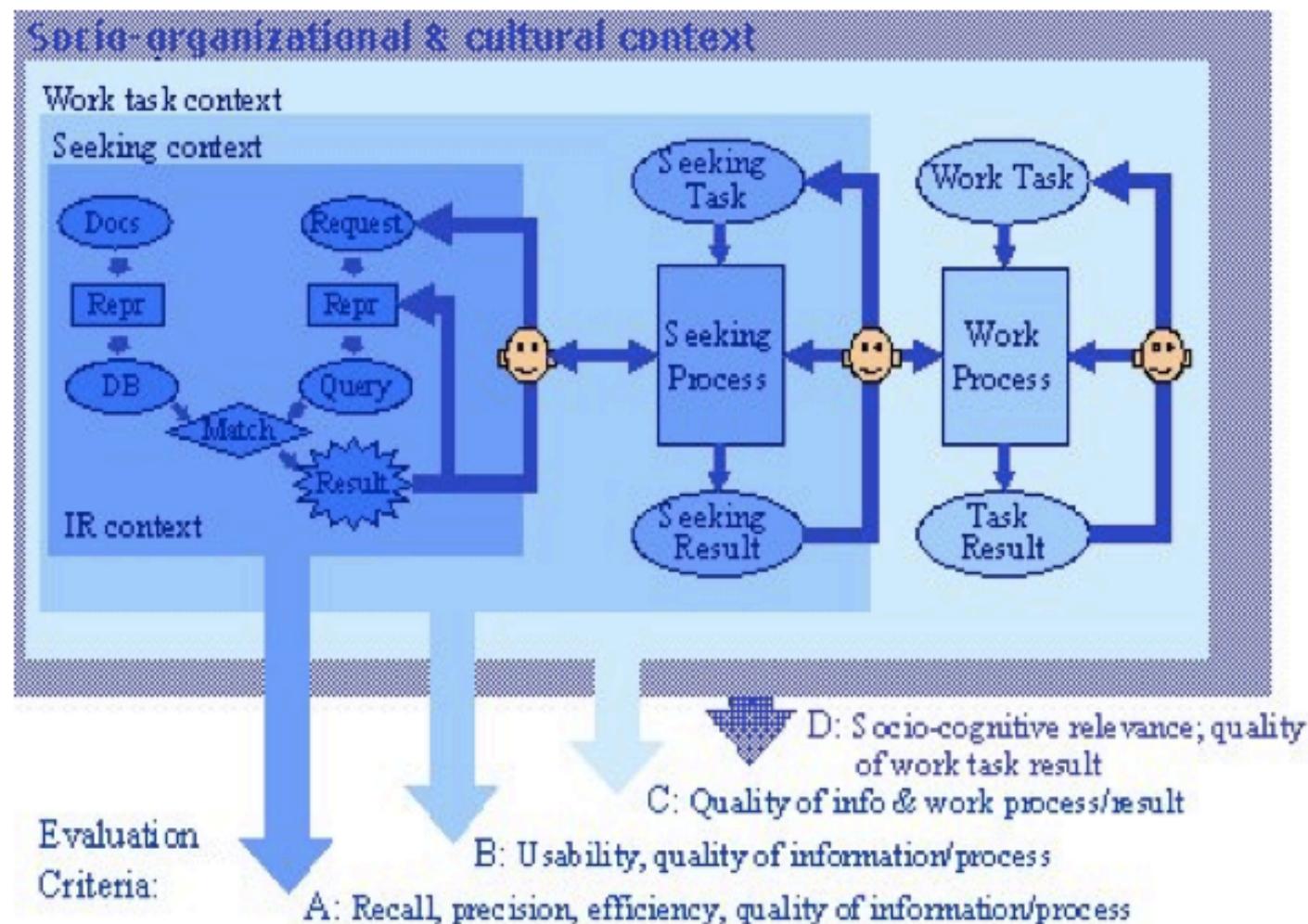
What is a model?

- Simplification of reality
- Often pictorial (can look somewhat like flowcharts)
- Created in different ways – inductive or deductive
- Used to
 - demonstrate our understanding of a problem
 - generate hypotheses
 - understand what to investigate
- Hopefully will lead to theories
 - Generally a precursor to theory

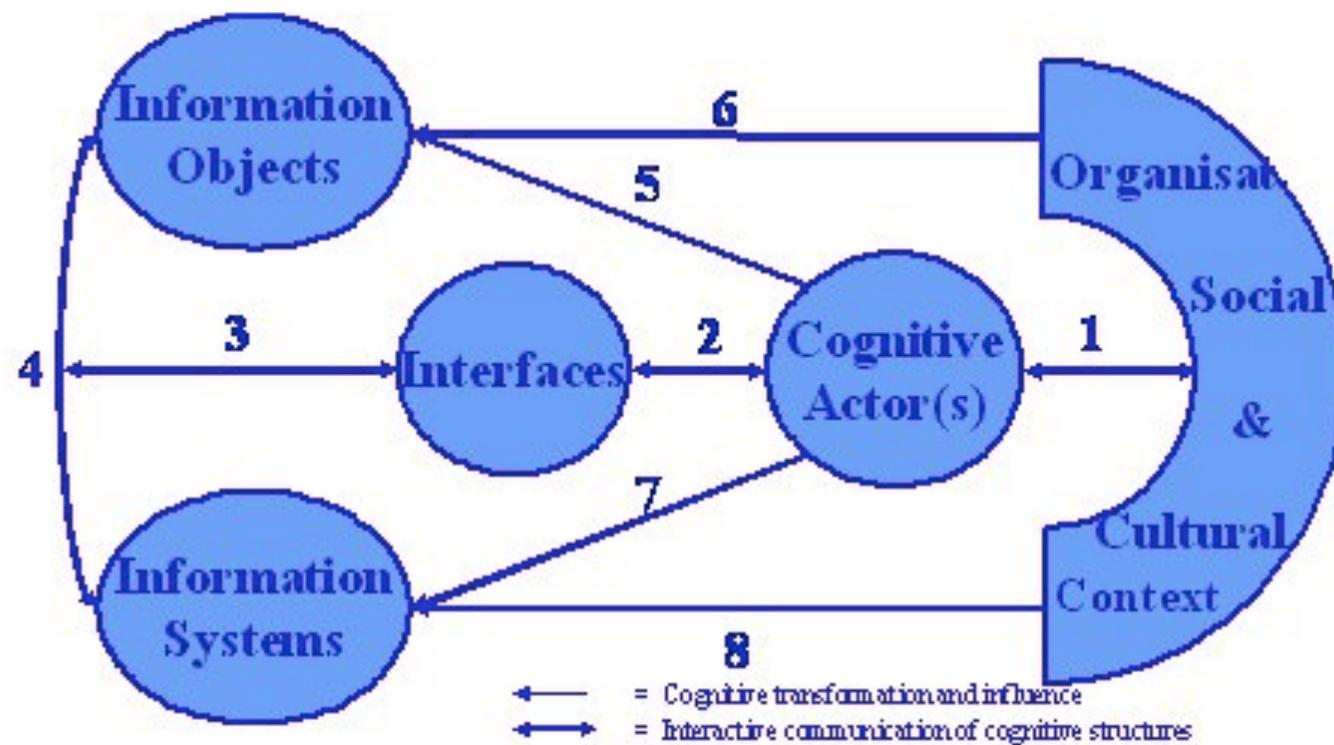
Limitations of models

- Too simplistic –try to represent very complex behaviour in a way that can be simply understood
- Can be misinterpreted
- Can be over-generalised
 - Used in contexts beyond its planned use

Nested Contexts of Use (Ingwersen & Järvelin 2003)



Components of IR and Seeking



What does this model offer?

9 Dimensions Encouraged by the Model

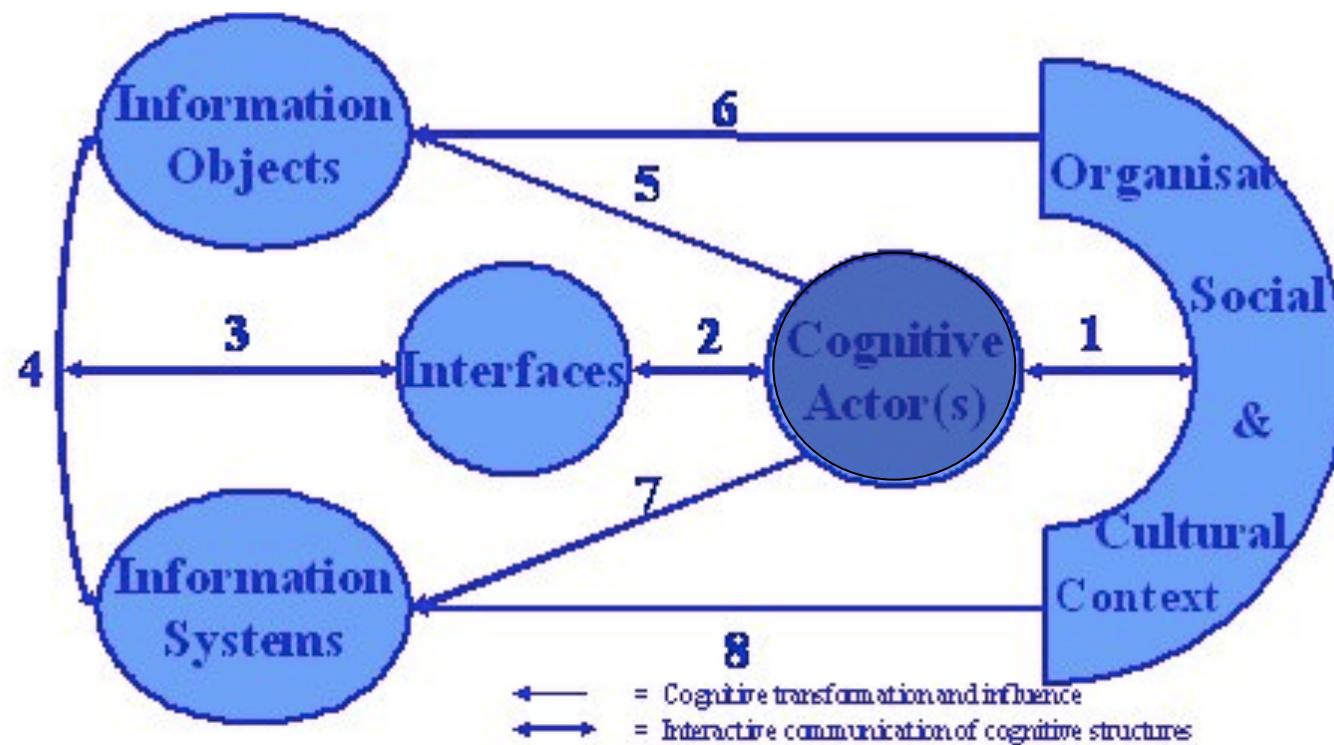
1. **Work Task:** set by the organisation, social context of work, collaboration between actors, physical and system environments
2. **Search Task:** seeking and retrieval practices as understood in the organisational / social practice
3. **Actor:** e.g. declarative knowledge and practical skills, motivation and emotion
4. **Perceived Work task:** how the actor interprets and performs the task.
5. **Perceived Search task:** how the actor translates the info need; perceives the information space etc.

9 Dimensions Encouraged by the Model

6. **Document**: document contents and collections in various languages and media
 7. **Algorithmic search engine**: think IR
 8. **Interface**: how objects, collections and their organisation are visualised and presented
 9. **Access and interaction**: strategies of information access, interaction between the actor and the interface
- All of these dimensions are complex with several variables!

Are there any other benefits?

Components of IR and Seeking



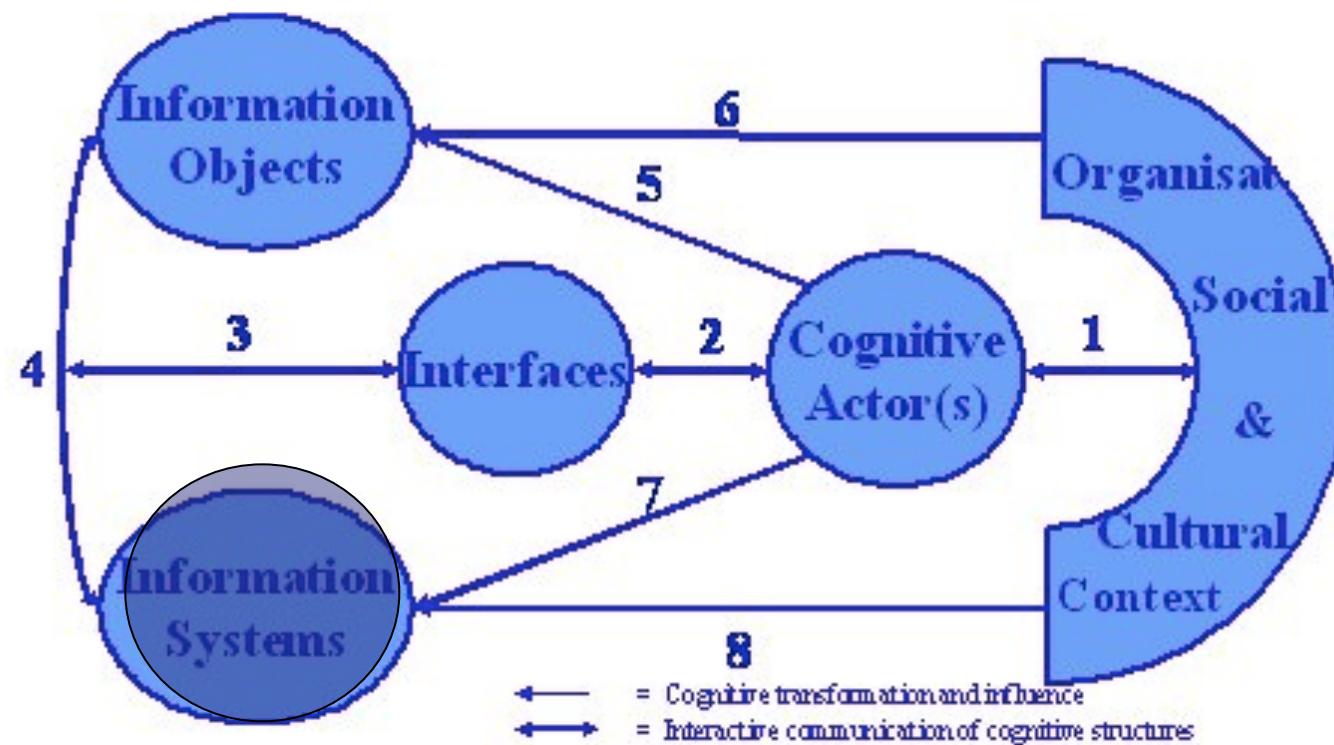
NOVICE VS EXPERT

- Experts spend less time generating queries
- take less time clicking on results
- explore results more thoroughly
- are ultimately more successful
- White & Morris, 2007; Aula et al., 2005



Different facets of society

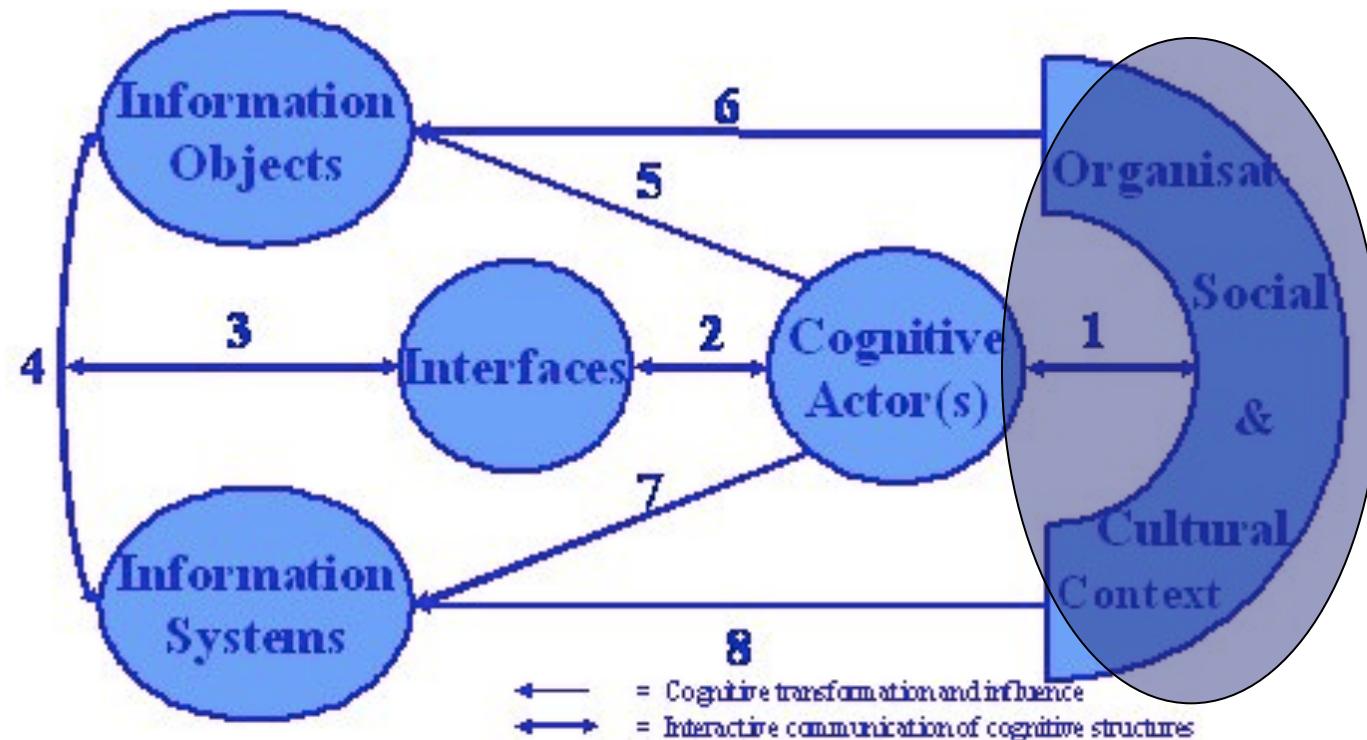
Components of IR and Seeking



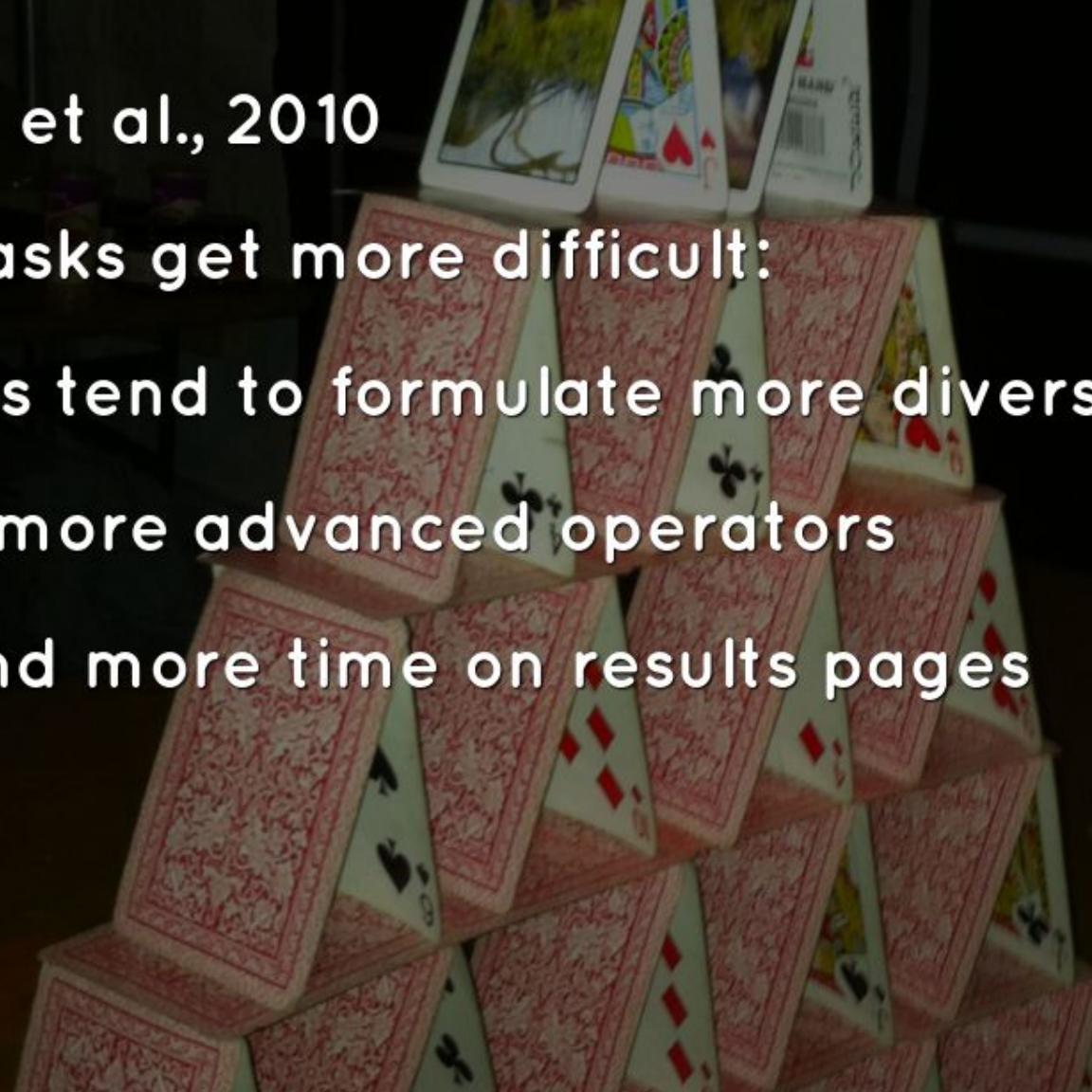
STRATEGY ADAPTED TO SYSTEM

- when search systems are less effective
- users submit more queries
- look deeper in the result lists
- Smith and Kantor, 2008

Components of IR and Seeking

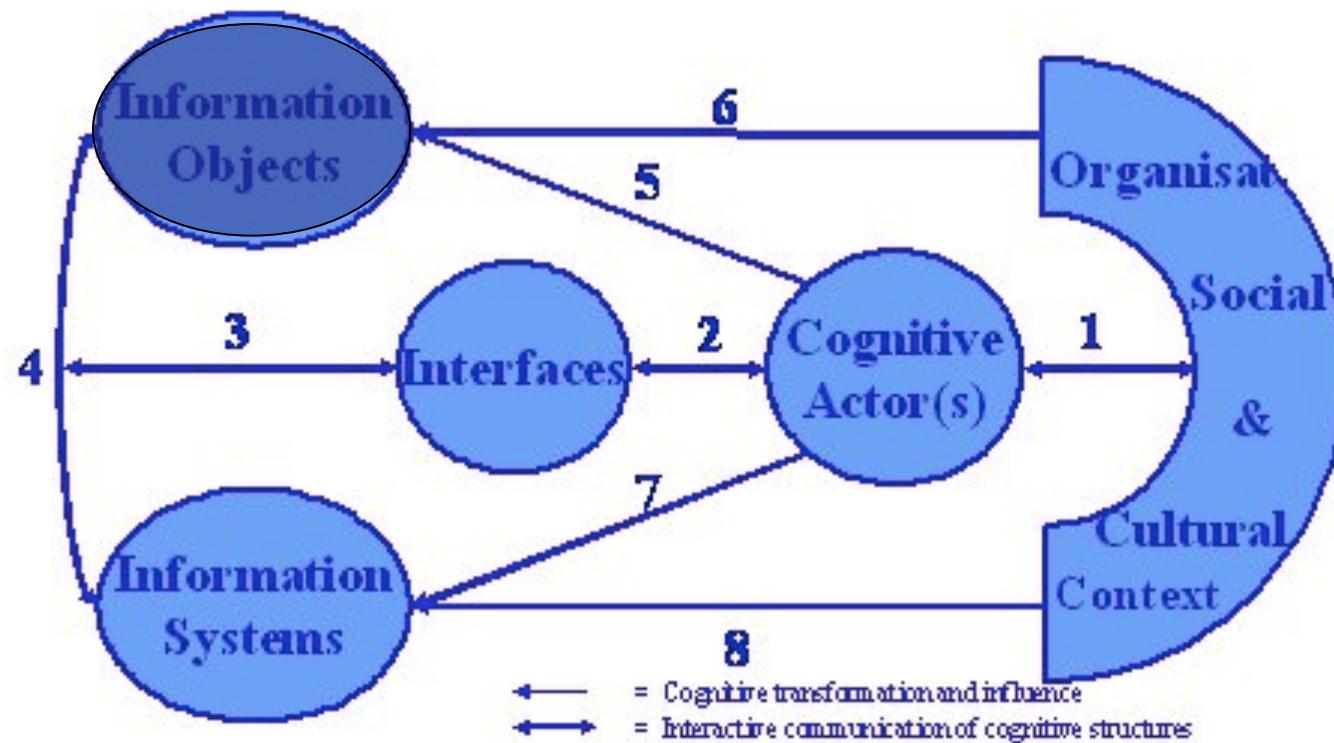


TASK DIFFICULTY

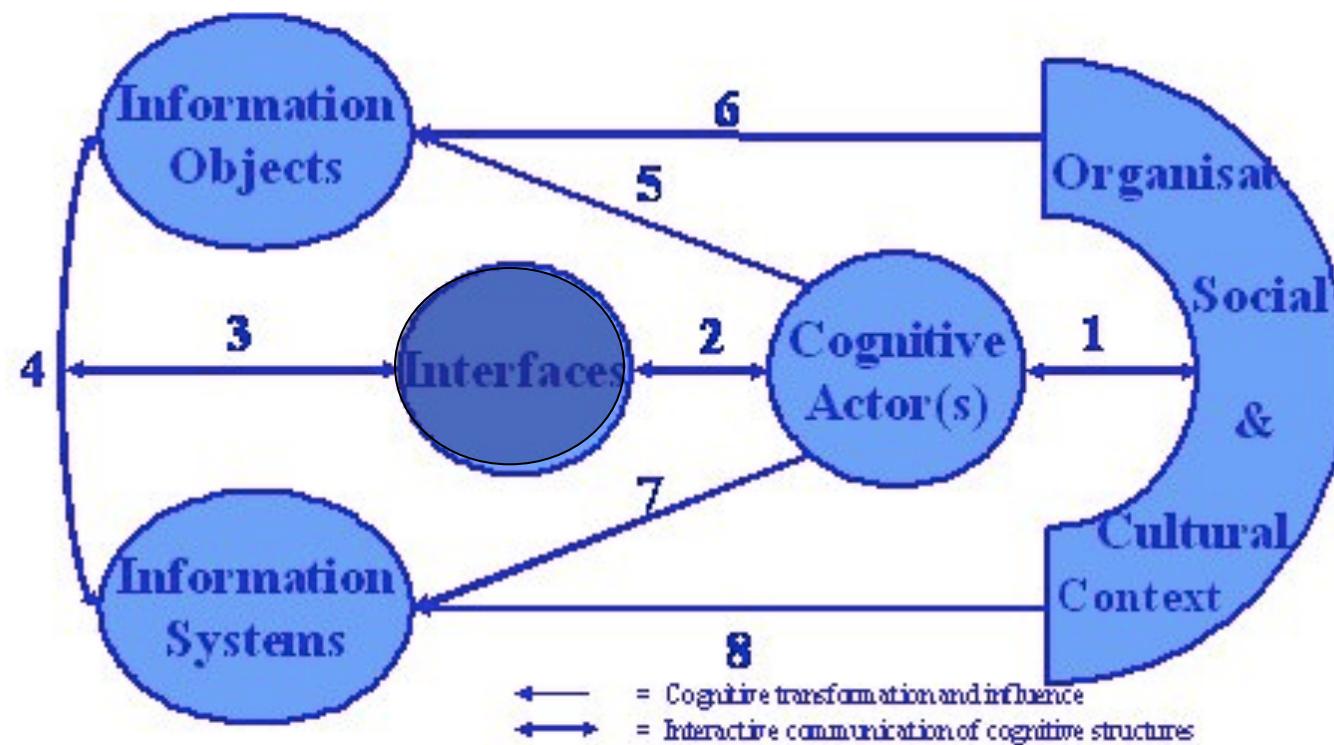


- Aula et al., 2010
- As tasks get more difficult:
- users tend to formulate more diverse queries
- use more advanced operators
- spend more time on results pages

Components of IR and Seeking

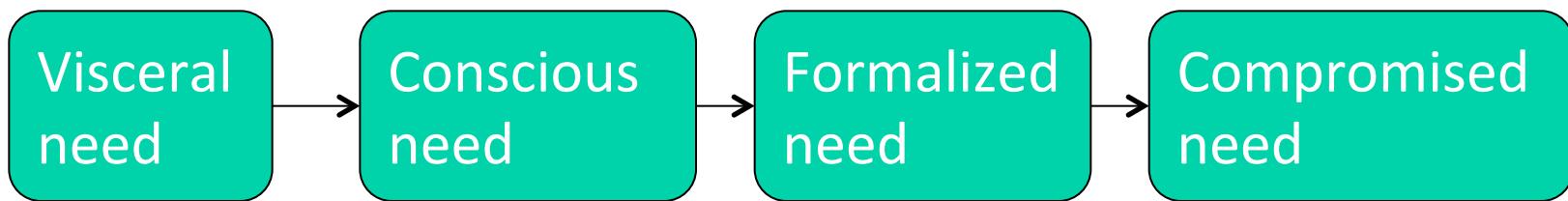


Components of IR and Seeking



Why do people search?

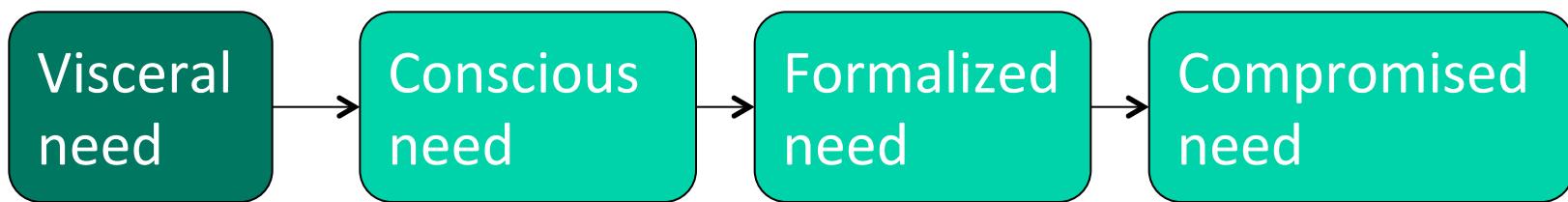
Taylor's Stages of Information Need



Thought about the how and why people ask questions to librarians in libraries

Why do people search?

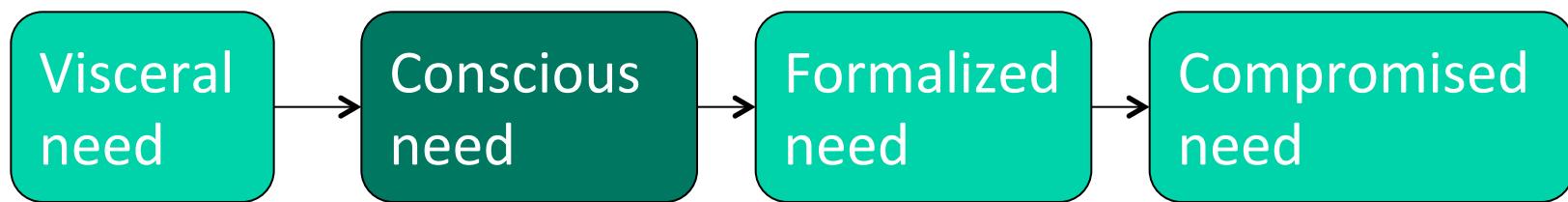
Taylor's Stages of Information Need (1968)



“A conscious or even unconscious need for information
... a vague sort of dissatisfaction ... probably
inexpressible in linguistic terms”

Why do people search?

Taylor's Stages of Information Need (1968)

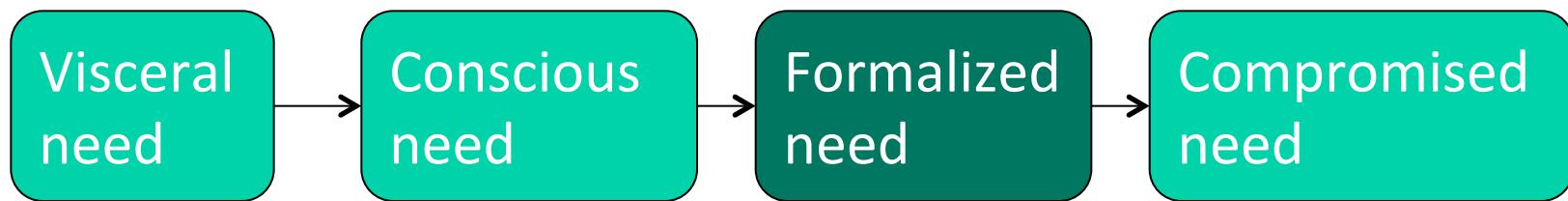


“A conscious mental description... an ambiguous and rambling statement”

Sometimes involves talking to another person about it.

Why do people search?

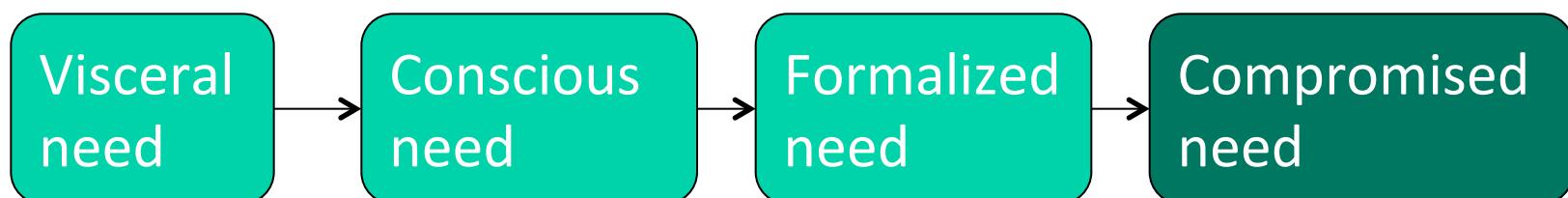
Taylor's Stages of Information Need (1968)



- At this point the enquirer might be able to construct a *formalized* “qualified and rational” statement of the need
- Not aware if it the need can be answered by any available person or information system

Why do people search?

Taylor's Stages of Information Need (1968)



- “The question is recast in anticipation of what the files available can deliver”
- Could be in the form of a question to a librarian or a query to an IR system
- A compromise between how the user originally envisages the query and the query must be to match the source.

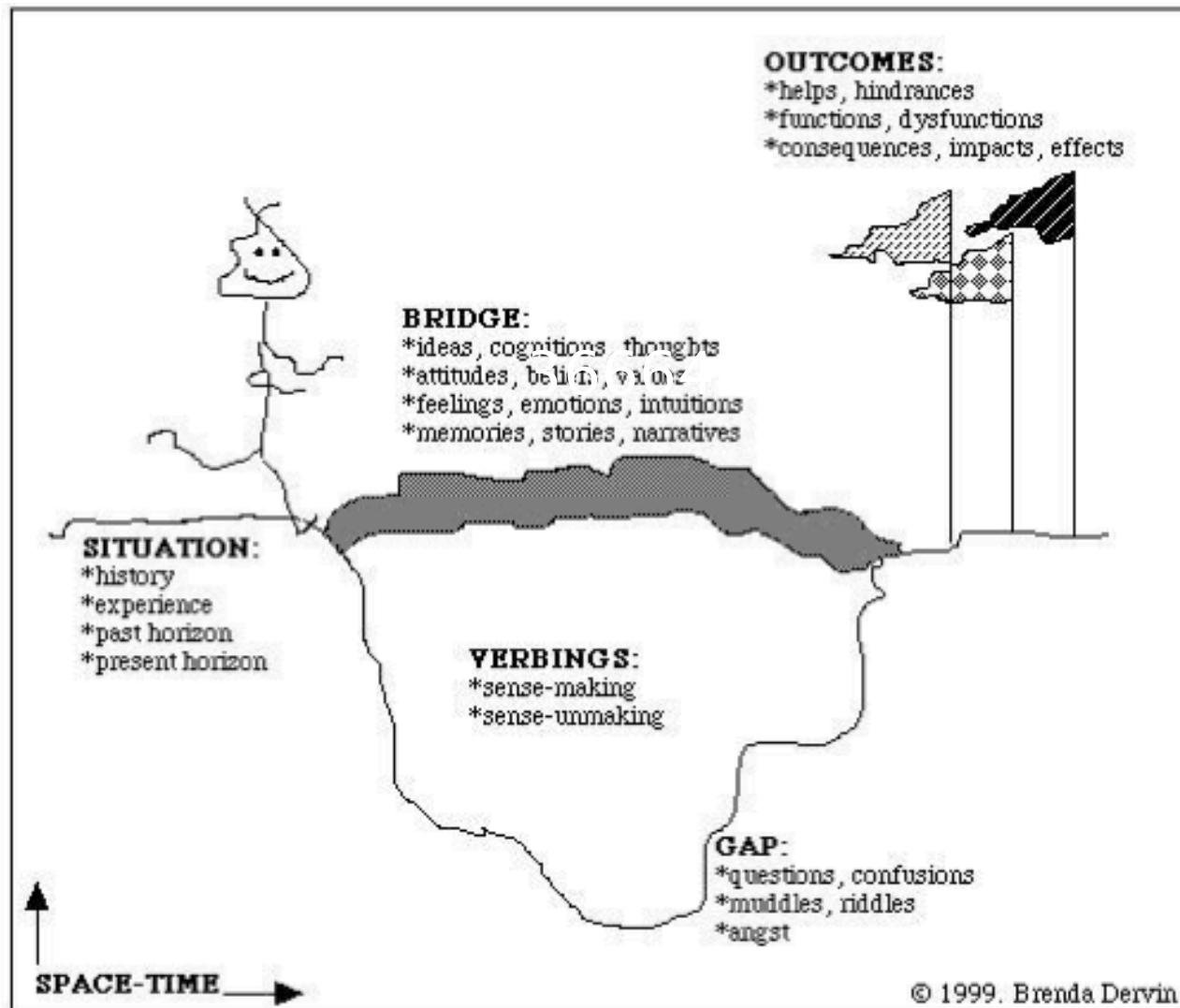
Belkin's ASK

- Information Seeking arises when an individual recognises an anomaly (gap, inconstancy) in their state of knowledge regarding a particular topic or task.
- To respond to this ASK, the individual looks for (in some way) the information required to address this.
- The ASK is then re-evaluated to determine the next step
- Comparable with the visceral need in Taylor's work

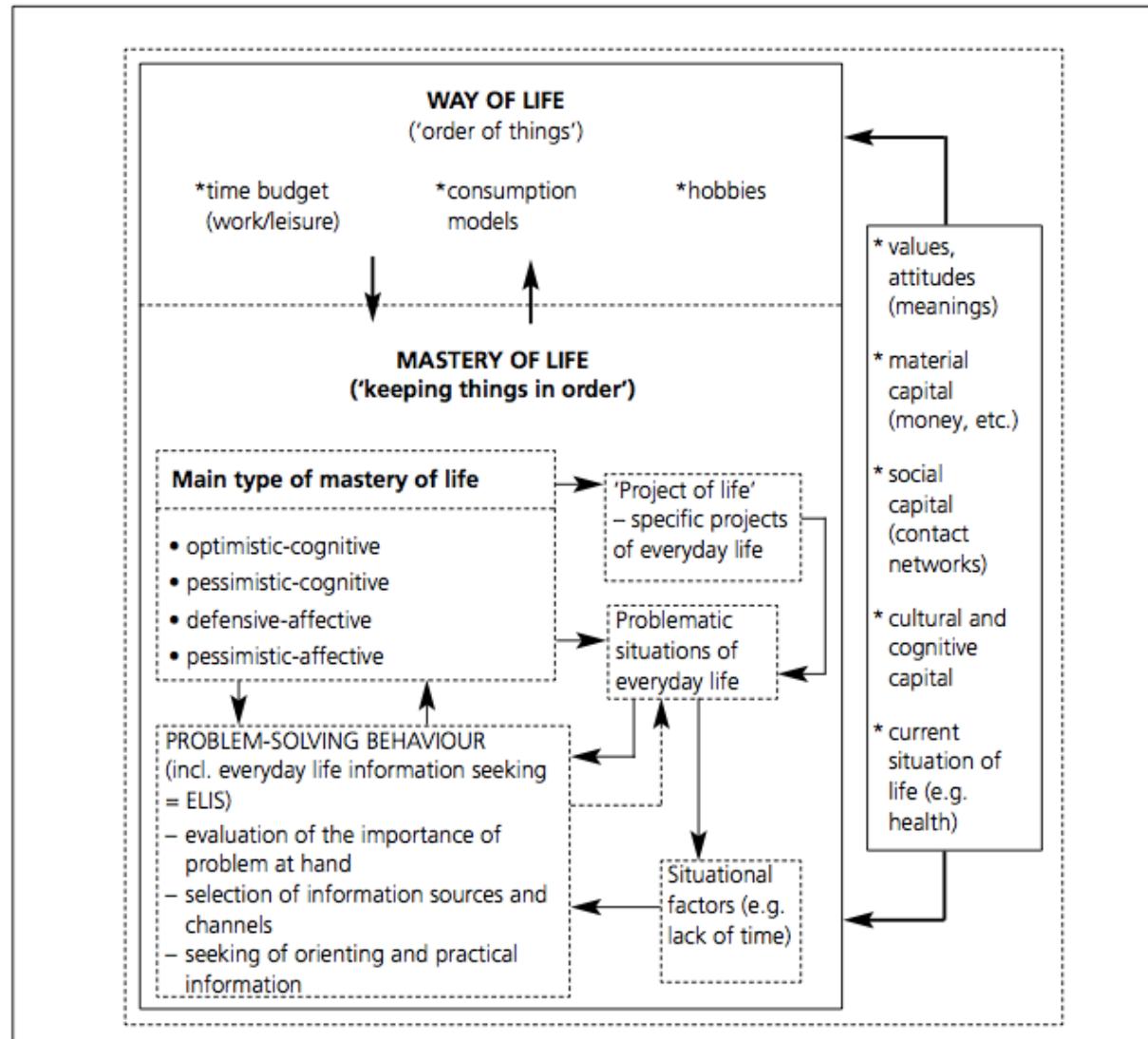
Kuhlthau's model

- Staged model (like Taylor), where an individual tries to reduce uncertainty.
- Stages associated with emotional states / affect.
- Built on psychological theories of learning
- At the beginning of a search, lots of uncertainty, feelings of anxiety
 - Can motivate work or cause people to give up
 - Different stages associated with different affects

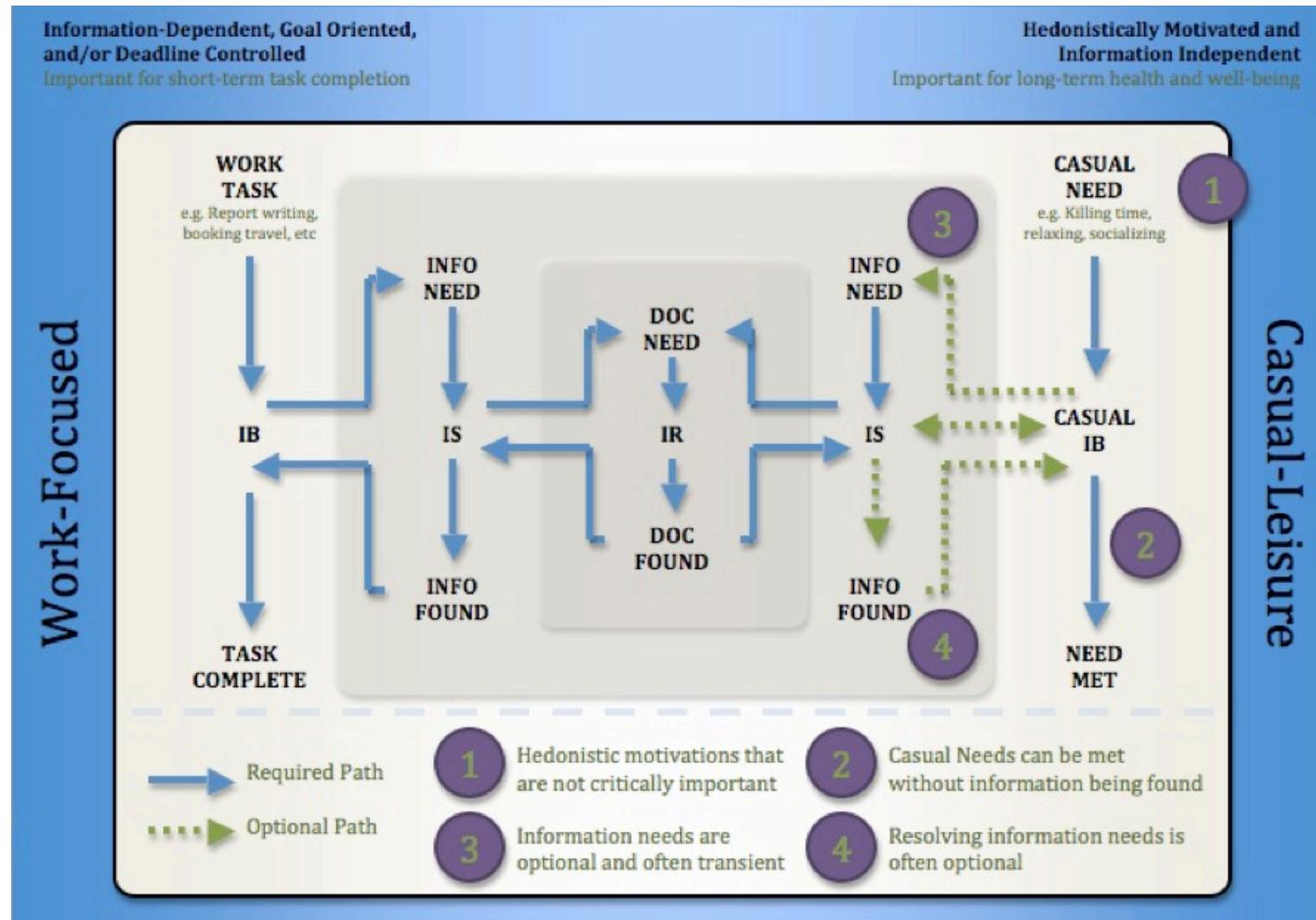
Searching to make sense



ELIS – Savolainen Model



Needs vs Wants and Desires



- Elsweiler et al. (2011)

Spectrum (Case, 2007)

- Describes a spectrum of situations motivating information seeking:
- Objective pole: concrete, fixed, well-specified needs
- Subjective pole: many (perhaps most) IS behaviour is motivated by vague feelings of unease, curiosity or anxiety.
 - Highly dynamic



How do people search?

How people seek information

- Several models have been created to model information processes
- These view the problem at different levels and from different perspectives
- We are going to look at some of the important ones in the literature

Kulthau's model

- Carol Kulthau studied information seeking behaviour of senior high-school pupils in the library.
- Pupils were on a large assignment / essay
- Models created / developed over 2 decades using various studies
- Her interest is on the movement from doubt in the early stages to confidence in the last stages
- She identifies 6 stages

Kulthau's model

- Each stage has five elements:
 - Task (what needs to be done)
 - Thoughts (in the seeker's mind)
 - Feelings (in the seeker's mind)
 - Actions (how ideas turn into practice)
 - Strategies (underlying ideas)

Kulthau's model

1. Initiation

Initiation is the phase in which one becomes aware of a need for information, an event often accompanied by uncertainty and apprehension. For instance, lets imagine that Fane Tomescu recently decided that he wanted to buy a car, prompting a need to research suitable vehicles.

2. Selection

The selection phase involves committing to constraints that narrow the information search. Fane quickly eliminated motorcycles, vans, and SUVs, deciding to look only at small family cars. Kuhlthau found that this phase tends to produce a spike in optimism once the user makes the selection.

Kulthau's model

3. Exploration

The optimism of selection usually gives way once more to confusion, uncertainty, and doubt as one realizes the many options still left to explore. Even though he had decided on small family cars, Fane still had to sift through dozens of makes and models, each of which had advantages and disadvantages. In Kuhlthau's study, about half of her students never made it past this stage.

4. Formulation

Formulation is the crucial turning point at which all the information encountered thus far is formulated into a specific, tangible requirement. Fane's car hunt reached the formulation stage when he decided that a four- to six-year-old VW Golf hatchback with 30,000 to 50,000 miles was the best fit for his needs and budget. The formulation stage is characterized by decreased anxiety and increased confidence.

Kulthau's model

5. Collection

Once the problem has been clearly articulated in the formulation phase, the next step is to evaluate the available solutions. Once Fane had a clear idea of the model he wanted, he used automotive websites to search for cars in his area that matched his criteria. Confidence continues to increase throughout the collection process.

6. Action

The final stage of the process is to act on the newly acquired knowledge. For Kuhlthau's students, this meant writing the term paper. For Fane, it meant going to look at a car, transferring money, and driving the car home.

Kulthau's model

6 Stages

	Initiation	Selection	Exploration	Formulation	Collection	Presentation
Feelings (affective)	Uncertainty	Optimism	Confusion frustration doubt	Clarity	Sense of direction/ confidence	Satisfaction or disappoint- ment
Thoughts (cognitive)	Vague		→	Focused		Increased interest
Actions (physical)	Seeking relevant information	Exploring	→	Seeking pertinent information	Documenting	

Kuhlthau's model

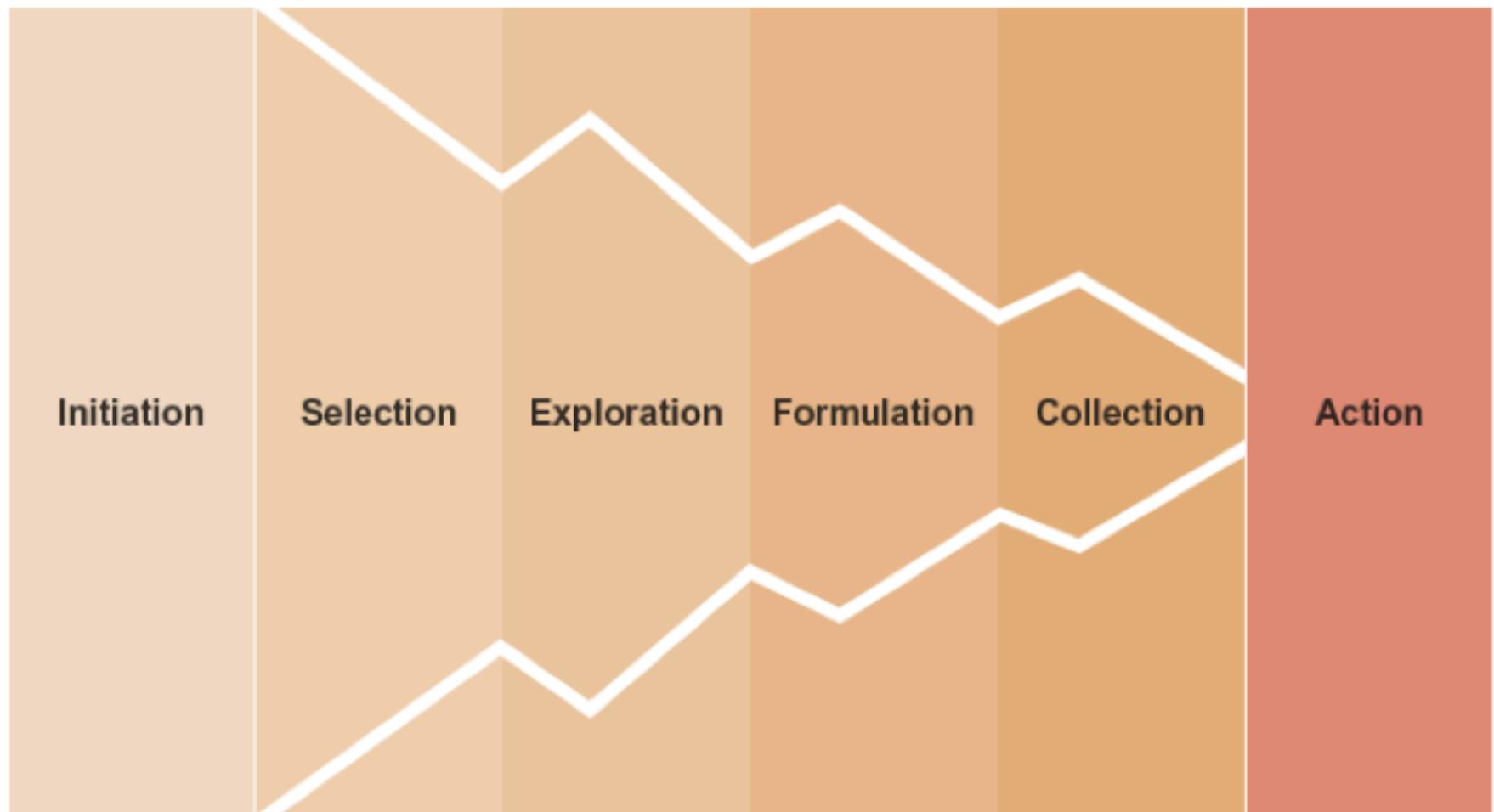


FIGURE 2.14 Kuhlthau's six stages of information seeking can be represented as a funnel that begins open-ended and ends with a resolution.

Kulthau's model

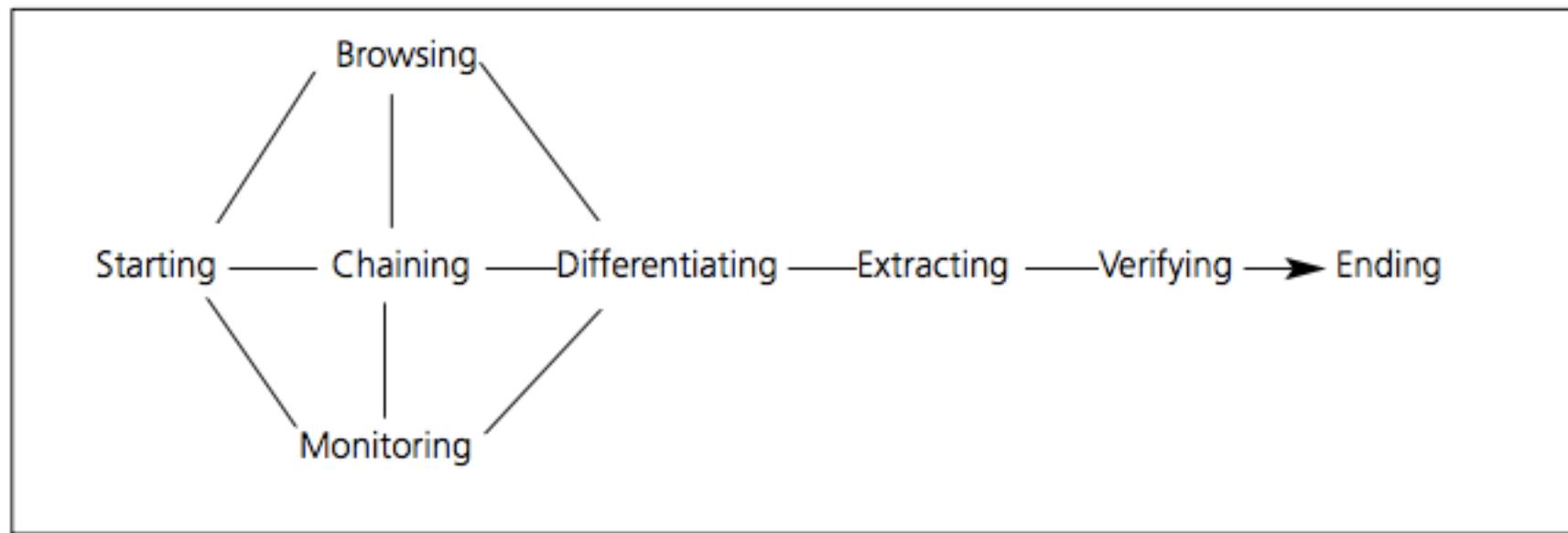
- Some properties of the model
 - Sequential
 - Gives a step-by-step listing of a series of actions!
 - Cognitive and Affective
 - Articulates the state of the user's mind (what they are thinking and how they are feeling)

Kulthau's model as for web search

Search stage and goal	Feelings	Thoughts	Actions	User need
1. Initiation—Recognize problem or gap	Uncertainty, apprehension	Vague, unclear	Identify problem and solving strategies	Reflection time
2. Selection—Choose a search engine	Optimism, trust	General, task oriented	Locate starting point; Typing URL	Reliability; accuracy
3. Search—Formulate query and submit	Anticipation, “I’m feeling lucky”	Thinking ahead to solution	Type search string, Click “GO”	Ease of use
4. Differentiation—Prioritize search results	Uncertainty, confusion, feeling overwhelmed	Unclear, mixed	Scan and prioritize results	Guidance, reduced complexity
5. Deciding—Determine most relevant results	Clarity, confidence, sense of direction	Narrowed, clearer, Increased interest	Select resource (click on link)	Help selecting
6. Extract—Read or use information	Relief or disappointment	Clearer, focused	Read, download, print	Usable, readable formats

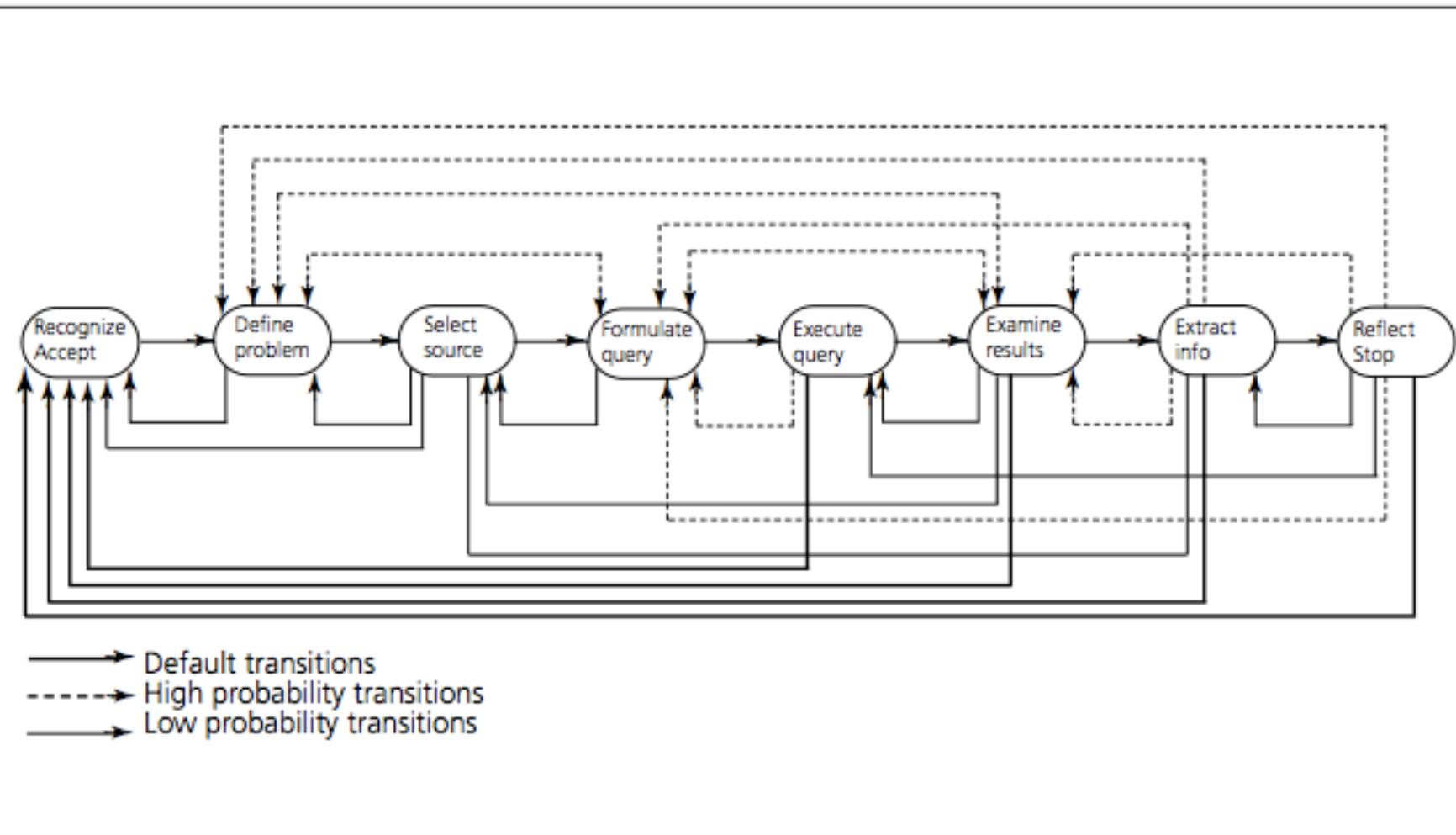
Kalbach 2006

Ellis' model



- Derived using grounded theory from series of empirical studies of academics and engineers
- Behavioural focus
- Has been extended (Meho and Tibbo, 2003; Wang et al, 2007)

Marchionini's ISP Model (1995)



Berrypicking (Bates 1989)

- A simple model that explains the meandering process of searching
- Search is not static but continuous in time in space
- Metaphor of food gathering
- People ‘pick’ berries of information
- Each berry starts another path
- Users travel from one berry to the next, collecting berries in their baskets (berries don’t come in bunches)

Berrypicking

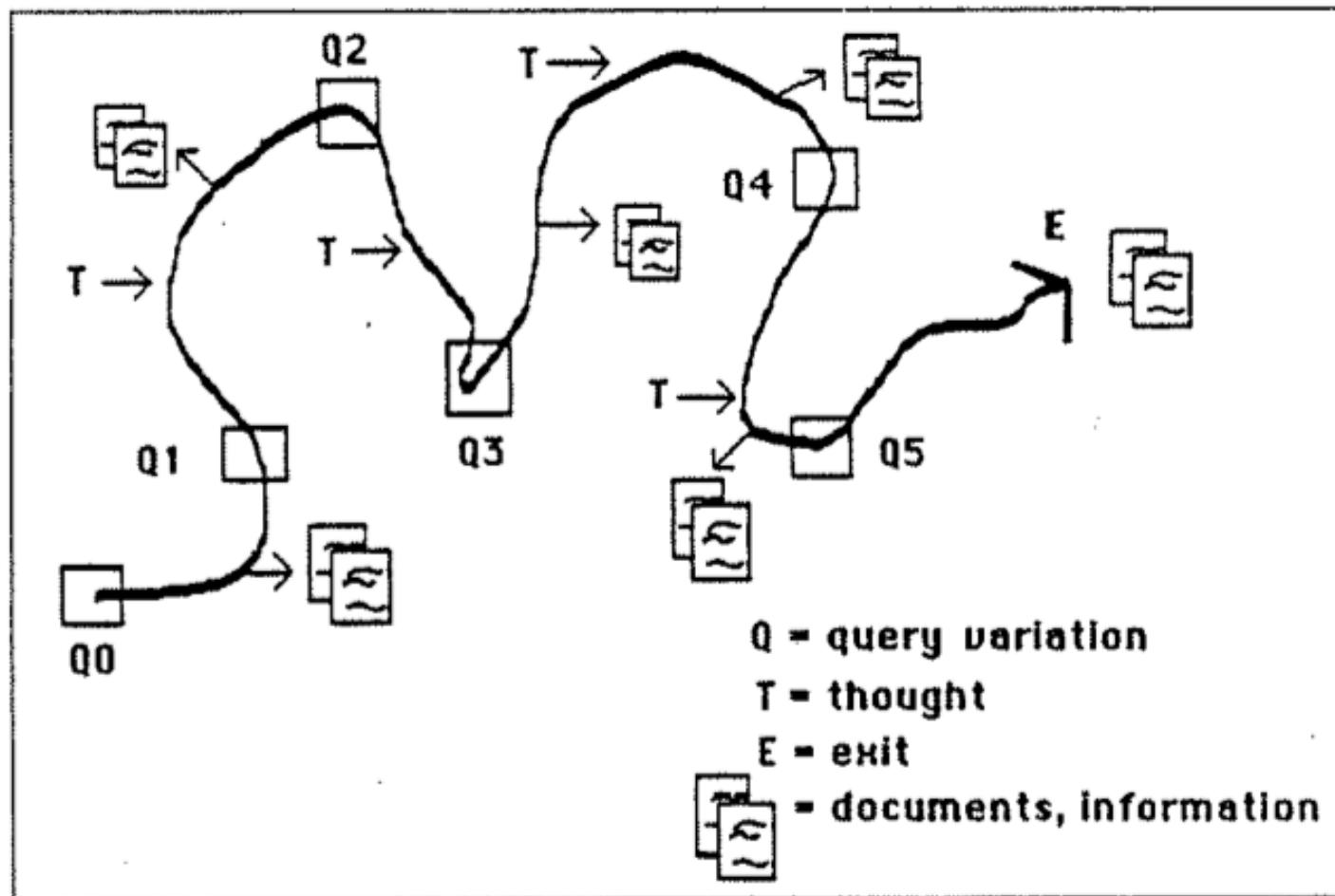


Figure 2: A berrypicking, evolving search

What does berrypicking model mean for IR?

- Typically search queries are not static – they can evolve
- People find information in bits and pieces – the need is not resolved by a set of results
 - Can learn from these
 - Can alter the need (and the usefulness of later berries)
- People use a variety of techniques and sources
 - Some people may have better knowledge of the space
 - Some people may have more motivation

A few points about these models

- These are just a few of the many models available
 - Chosen to highlight key points
- Some of them model stages (e.g. Kuhlthau), others concrete behaviours (e.g. Ellis, Marchionini) and others still important variables and dimensions (e.g. Ingwersen)
- Not always comparable as they were created in different ways (inductive / deductive) and model different scenarios (Byström – work task scenarios, Kuhlthau – school children and university students, Savolainen – everyday search scenarios)

What can we take away from these models

- Behaviour is not static
 - Change in for different actors / situations
 - Is it sensible to have 1 query per topic in trec?
- Relevance judgements
 - Not independent (relevance of a document will depend on the documents already read)
 - Will depend on the user, previous knowledge, situation etc.

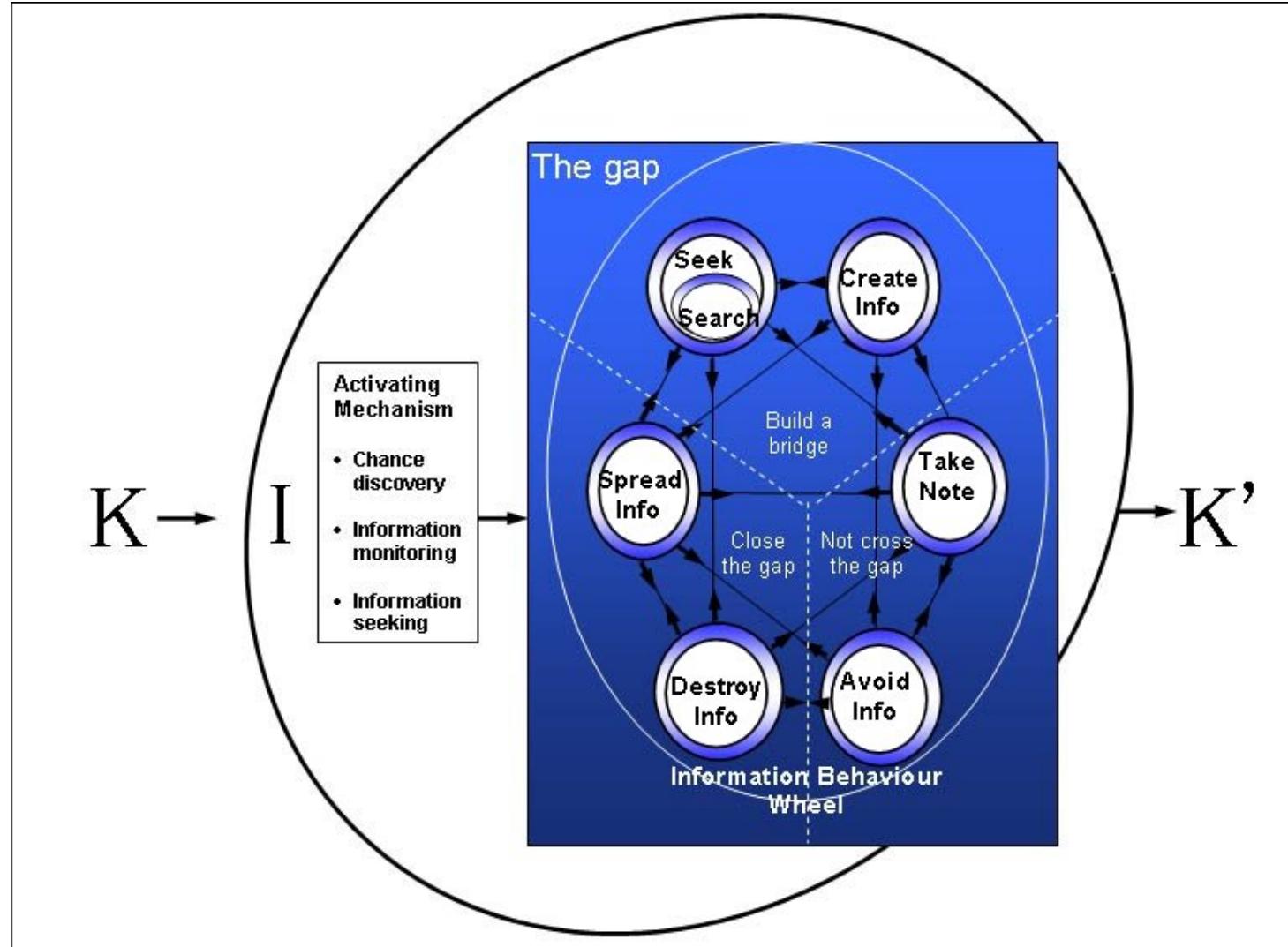
What can we take away from these models

- Are our metrics modelling things accurately?
 - Is relevance returning relevant information enough? (think searching for fun, emotion)
- Can we model interaction better?
 - Our metrics are modelling interaction in some way – think precision@k, map, recall etc.

Summary

- We have looked at IR from a different perspective
 - Human oriented
 - See that we have only been simulating a small aspect of the whole picture
 - Other aspects involved: people, feelings, emotions, interfaces, context
 - Behaviour is not static; relevance is not static
 - Hopefully this will get you thinking!!!!

Godbolt's wheel



Resources

- Ruthven and Kelly (Ch. By Wang) Models of Info Behaviour
- Ch 2. Designing the Search Experience
<http://designingthesearchexperience.com>
- Burghardt et al. (2013), Modelle des Informatinsverhaltens bei der Websuche (in GRIPS)