

# Extending IR toward CONTEXT

Kalervo Järvelin

([kalervo.jarvelin@uta.fi](mailto:kalervo.jarvelin@uta.fi))

University of Tampere

Finland

# Talk Outline

- Lab IR - traditional study designs and contributions
- Contexts of IR
- Modeling context - issues in modeling
- A cognitive approach to modeling
- Consequent research designs
- Conclusion - does it matter?



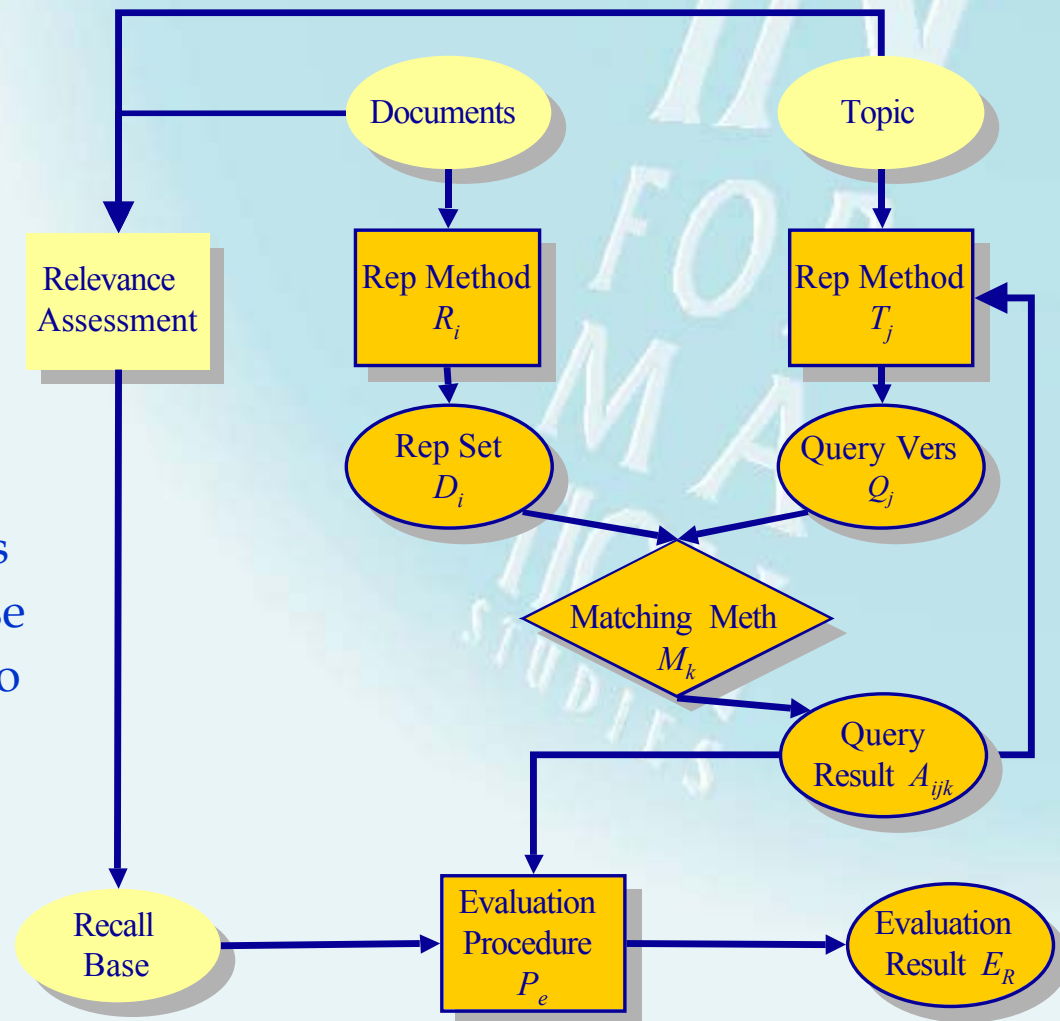
# IR Research Goals

The most popular focus - it is CS, after all, isn't it?!

- Traditional goals:
  - (a) theoretically (~formally) understanding IR,
  - (b) empirically testing IR methods, and
  - (c) designing better IR systems
- This can be seen in typical IR study designs - by populating the framework with variables

# Lab IR Study Types / Designs

- Study types:
  - experimental
  - methodological
  - constructive
  - theoretical
- Experimental designs – IR evaluation
  - focus areas with variables
  - designs operate with these
  - fringe areas controlled (no analytical variables)
  - dependent variables:  $E_R$
  - many hidden variables



# LabIR: Hypos, Laws, Theories

- The dependent variables - typically recall and precision - using various metrics
- The independent variables typically are the use or non-use of various techniques
- The controlled variables - collection, requests
- Thus Lab IR is about the *explanation of the variation of recall and precision through IR techniques applied*

# LabIR: Contributions

- IR technology
- Theoretical growth
  - theory expansion, e.g., due to enrichment of indexing theory
  - greater analytical power through formal model building
  - improved empirical support through testing over several test collections
  - new research questions – within the given set of variables
- Theoretical growth has been limited by not bringing in radically new concepts or relations
- *Cannot possibly analyze real world applicability*

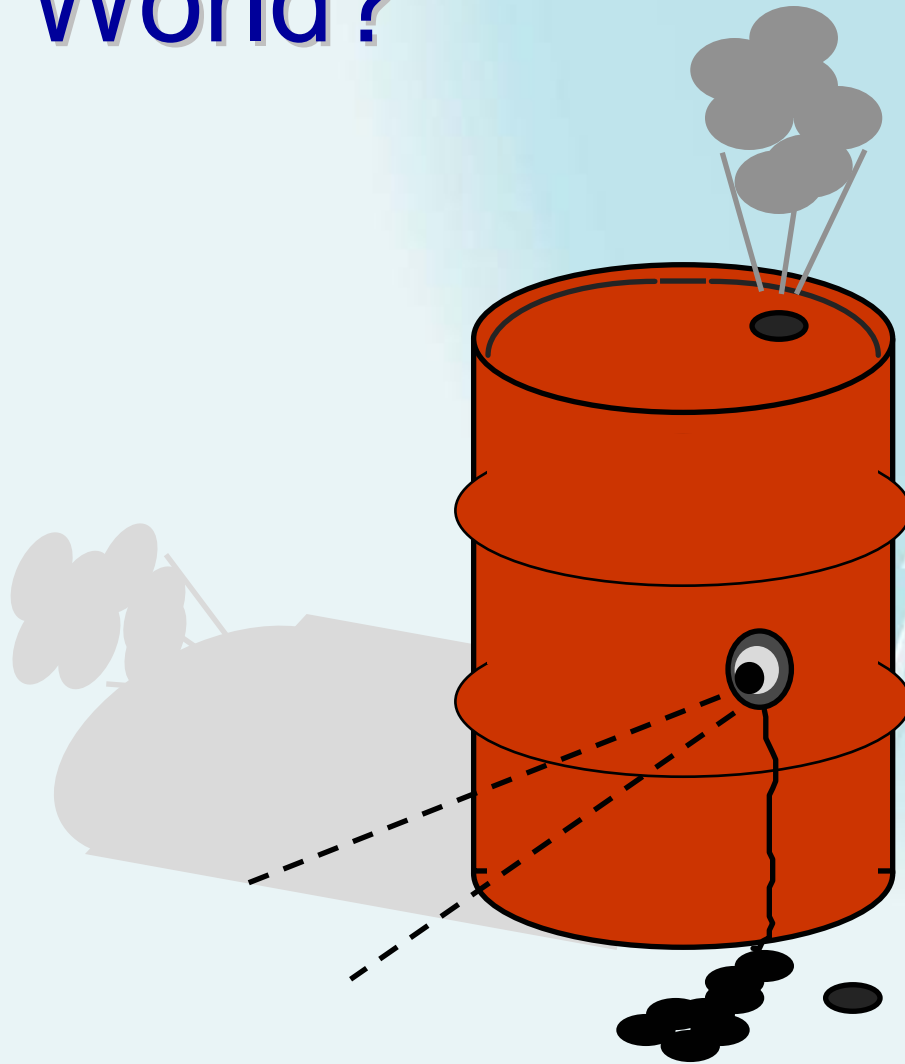
# Lab IR Tests

➤ IR research typically considers only *retrieval tasks* which most often are:

- (a) purely topical
- (b) content-only
- (c) well-defined
- (d) static, and
- (e) exhaustive

This is like saying that no matter what your situation is, your needs **always are** purely topical, content-only, well-defined, static, and ...  
... True, isn't it?

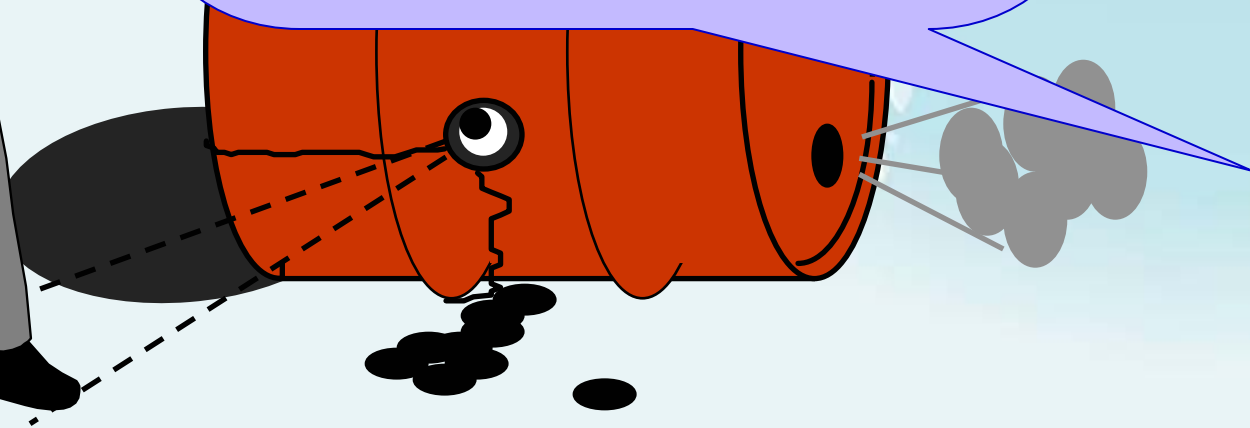
# Small World?



Alternatives?



Psssst ...  
... IR in  
Context



**One possible context:  
Task  
Performance  
Augmentation  
through IR**

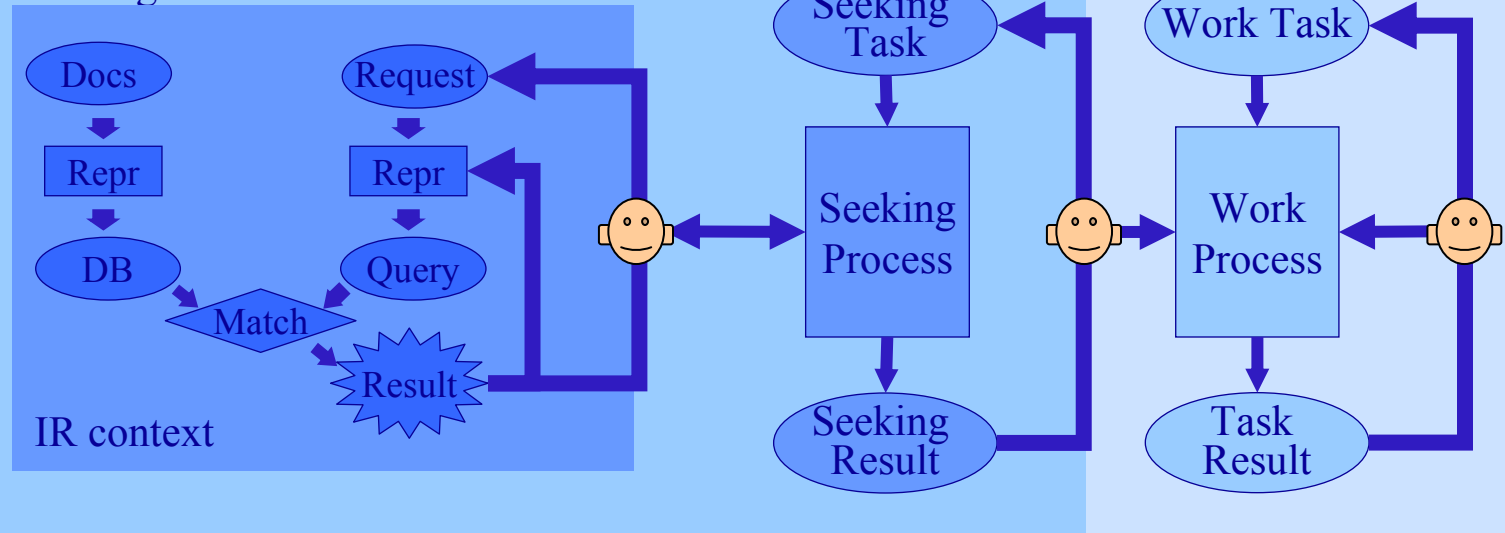


# Nested Context Frameworks

Socio-organizational & cultural context

Work task context

Seeking context



- How much should one cover?
- How to model what is covered?

# What to cover?

- Clearly, current IR test designs are not representative of real world IR
- In principle,
  - *whatever has important relationships with the application of IR, should be covered*
  - *however, what has, is not well known*
  - *answering this requires theoretical modeling, operationalization and empirical testing*
  - *so, what is context?*

# Context? What?!

## ➤ Dervin (1997):

- there is no term that is more often used, less often defined, and when defined defined so variously as context - it has become almost a ritualistic invocation
  - for some, context has the potential of being virtually anything that is not defined as the focus
  - for others, it is inextricable surround that denies all generalizations
  - there are endless lists of contextual factors
- This is why we hate it - it is foreign to CS!

# Some Contextual Themes

- Reality is discontinuous, knowledge partial
- Context is not useful as an independent entity
- Context requires a focus on process and relationships between product and process
- Need to focus on multiple interdependencies
- Context is a necessary source of meaning

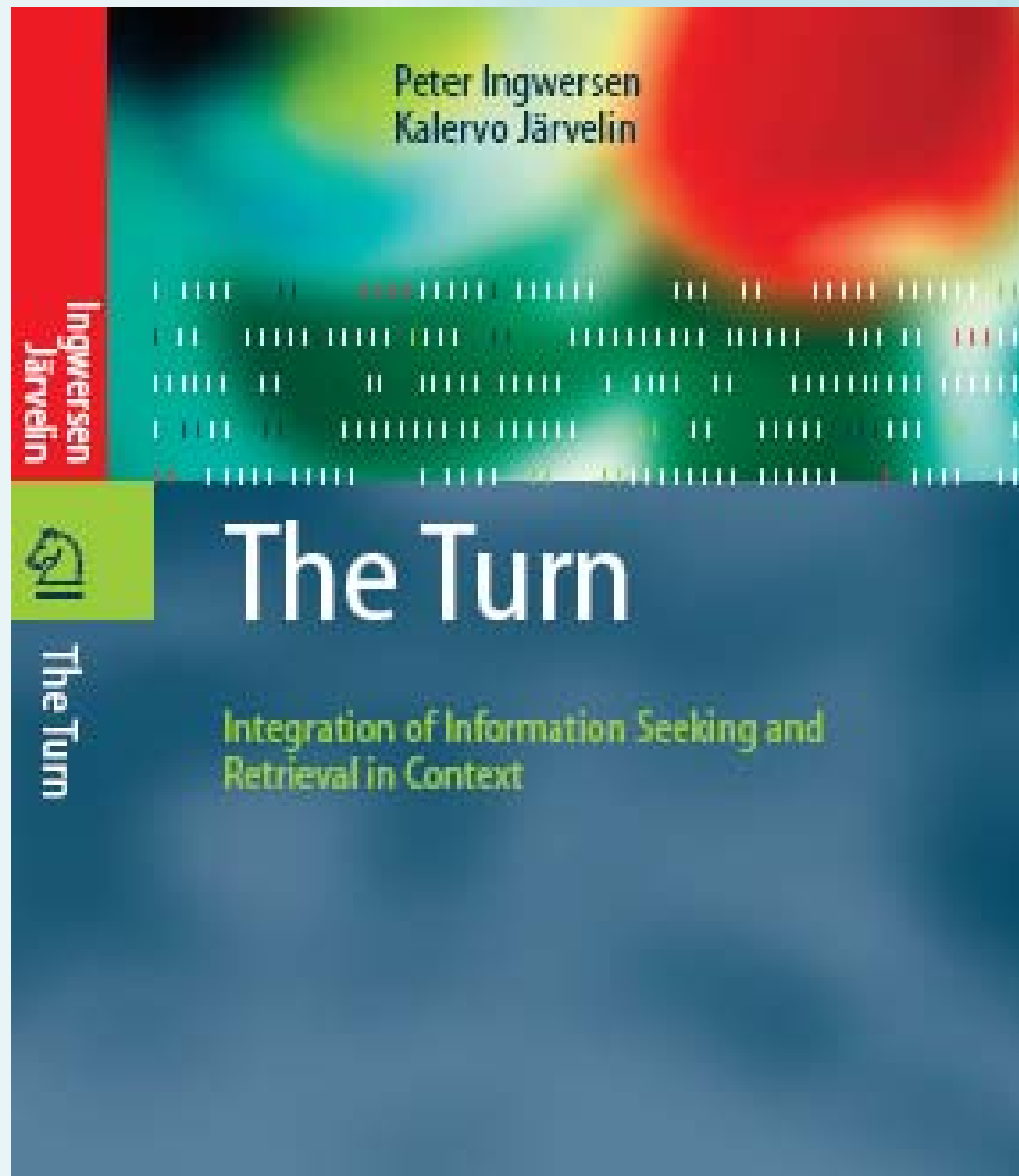
**(Dervin 1997)**

# How to model? - IR Context Dimensions

1. Work task dimension
2. Search task dimension
3. Actor dimension
4. Perceived work task dimension
5. Perceived search task
6. Document dimension
7. Algorithmic search engine dimension
8. Algorithmic interface dimension
9. Access and interaction dimension

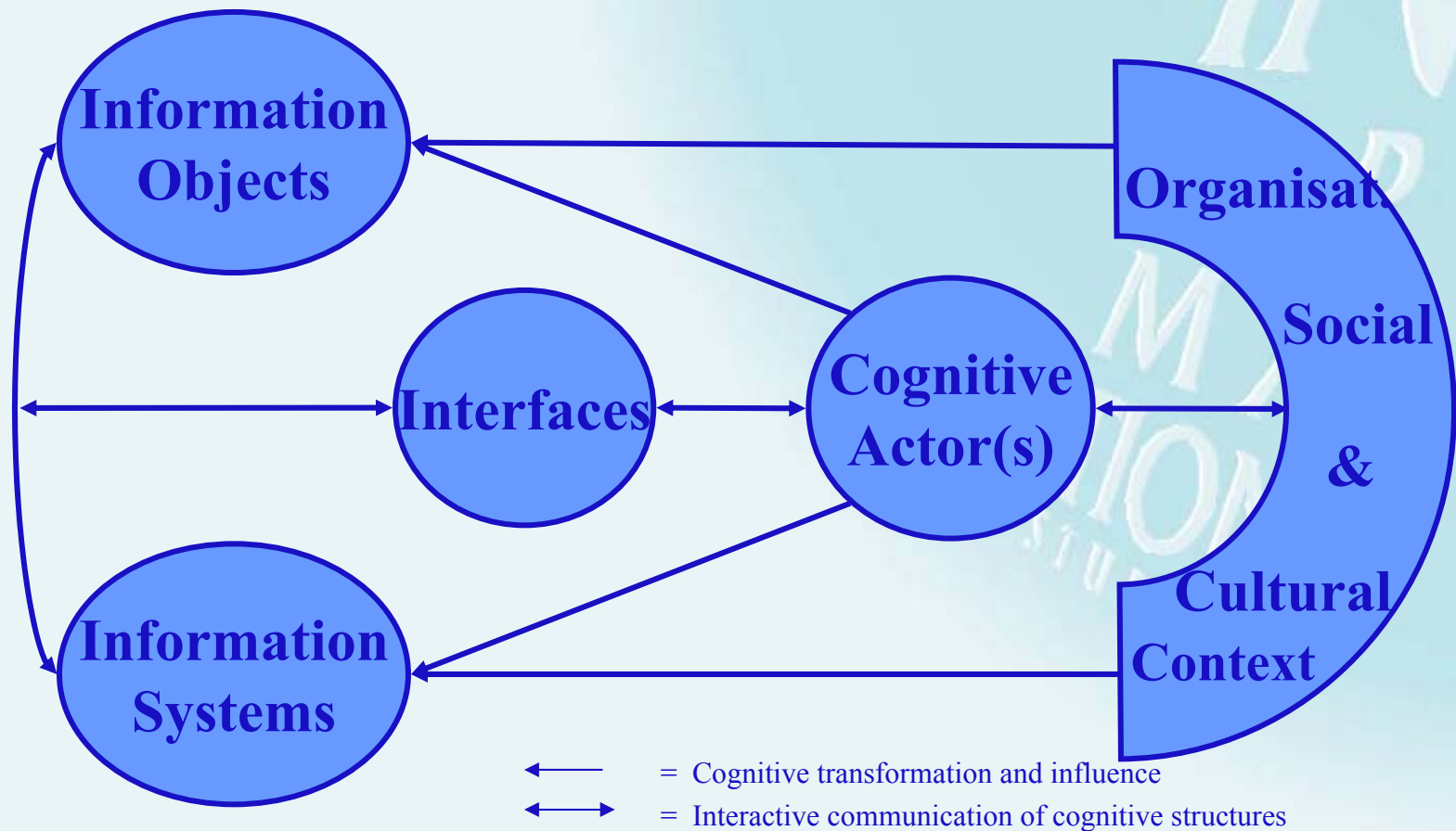
Each dimension containing multiple variables

Much more detail in the forthcoming book by Peter&Kal

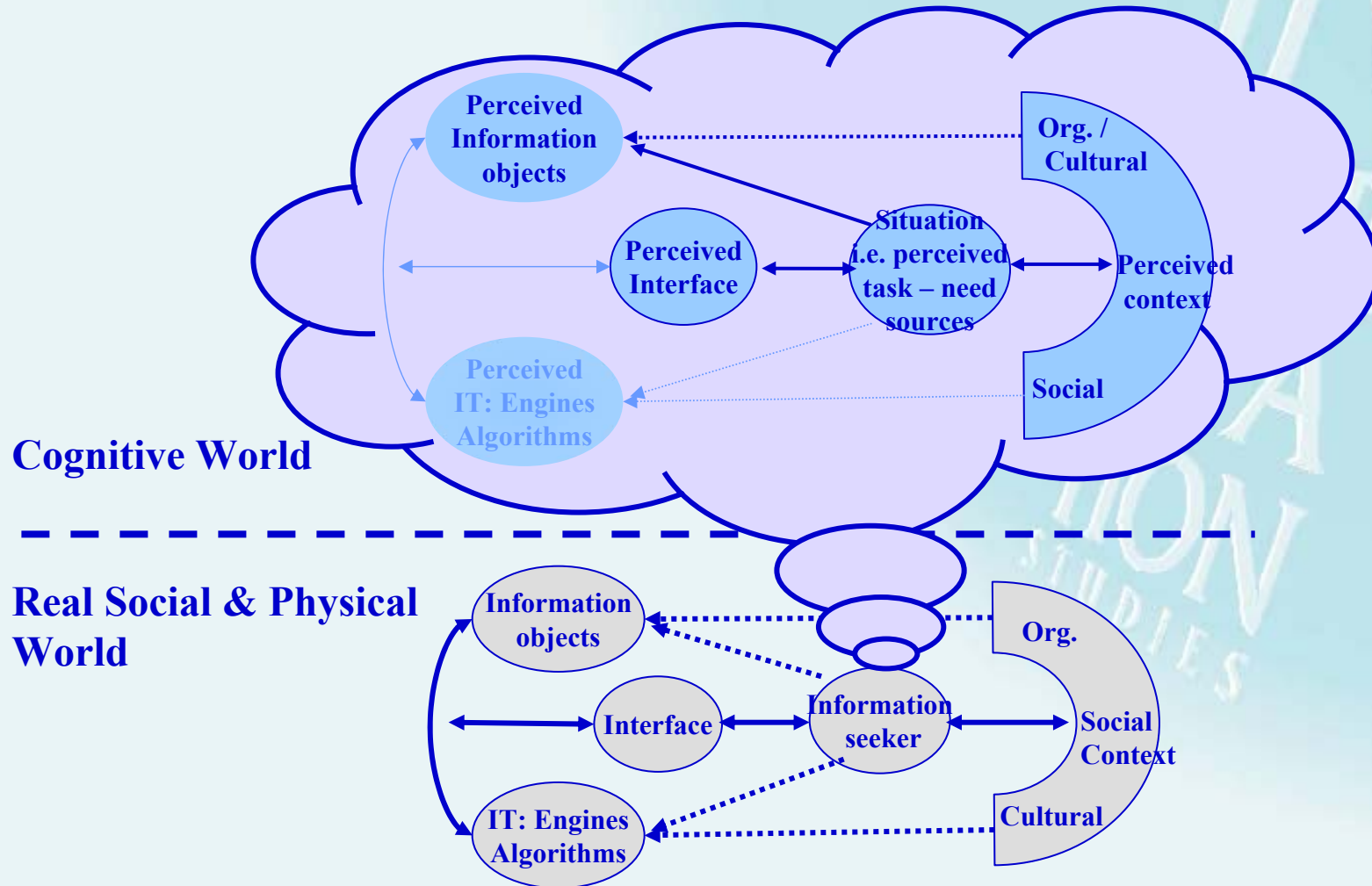


<http://www.springeronline.com/>

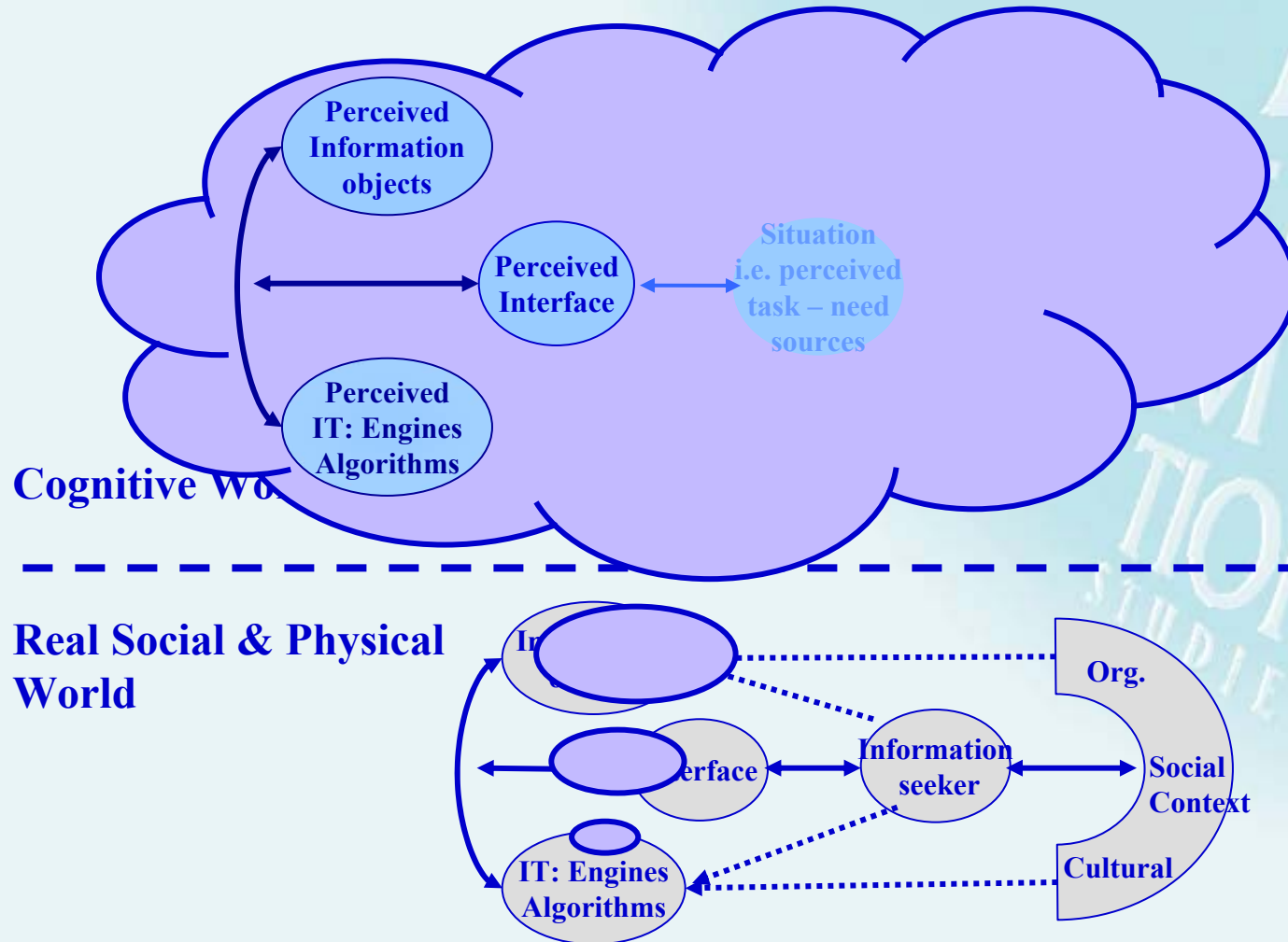
# A Cognitive Model of IS&R



# Context around and within



# Context around and within



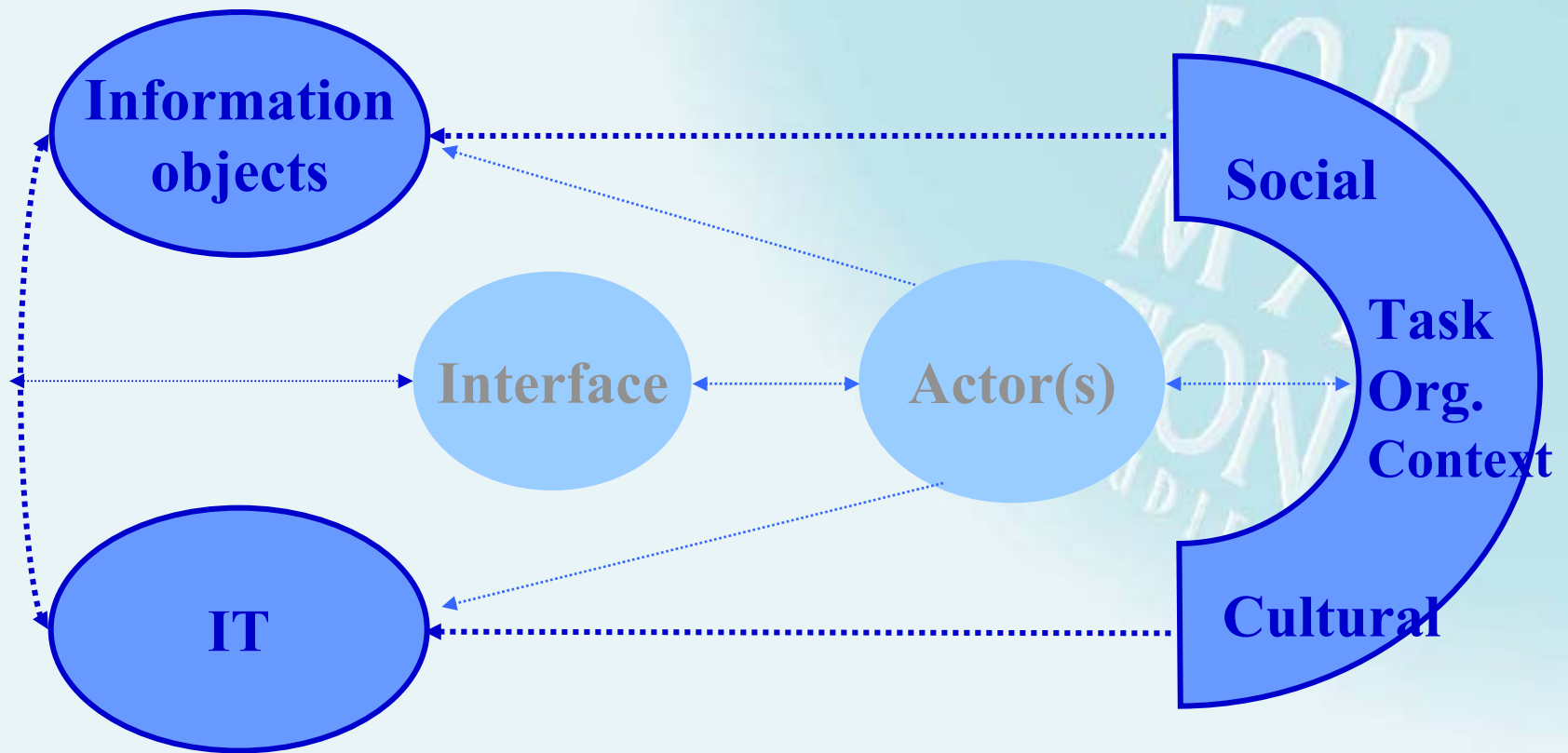
# Augmented IR Research Goals

- Theoretically understanding the application of IR in / to context; not just engines but also their use in varying situations
- Empirically describing / testing the application of IR in / to various contexts
- Supporting the design of information systems, services and information management
- This does not downgrade IR into Social Science but rather means more / better engineering

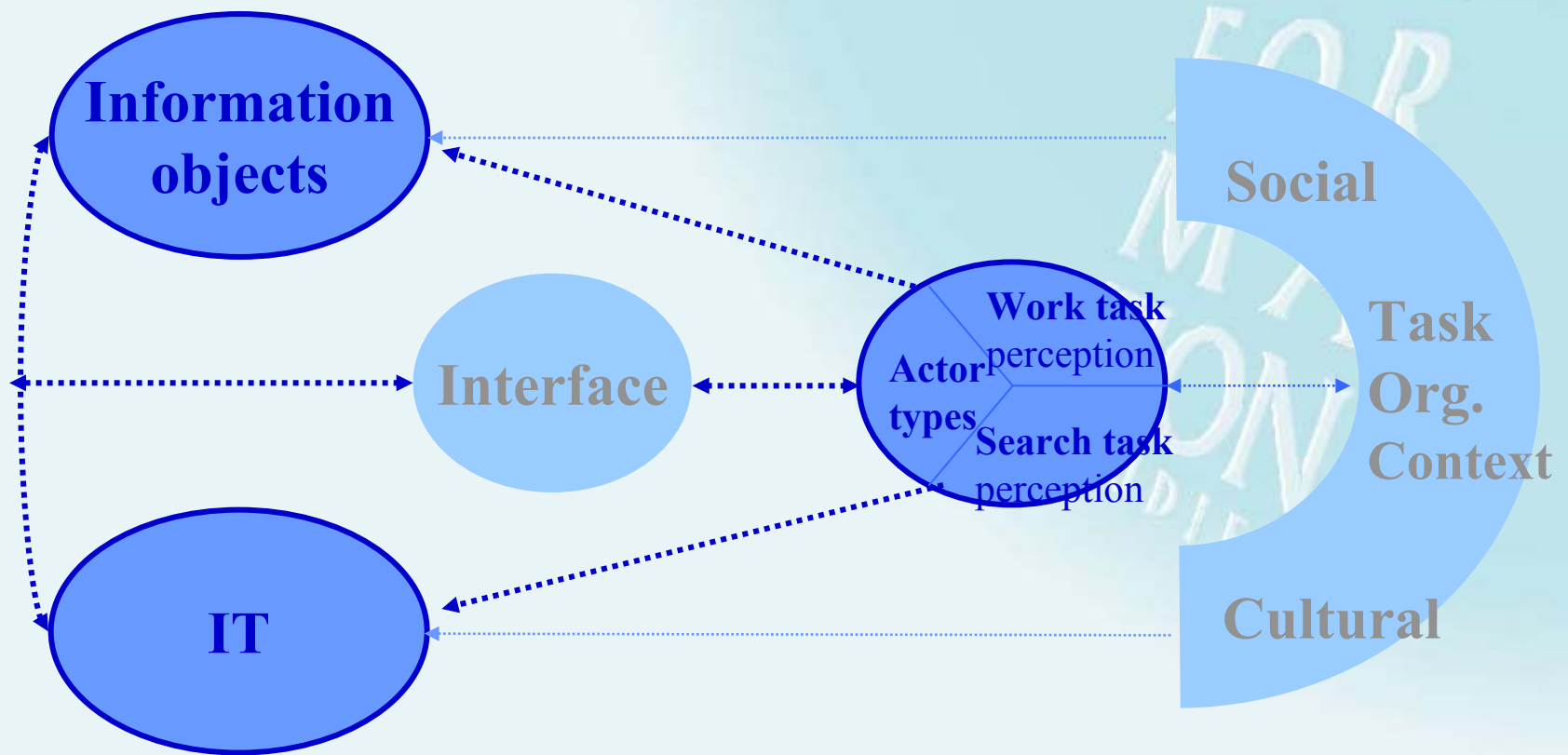
# New Research Settings

- Extended laboratory experiments
  - What kinds of work tasks are typical in Domain X?
  - What kinds of search tasks do these generate?
  - What kinds of documents tend to be relevant?
- IIR experiments
  - Lab approach but including the seeking actor
- Research on interfaces / actors in context
  - interaction of actors and information system interfaces in diverse socio-organizational contexts

# Lab IR settings focused on **domain/task** variables and dimensions of **IT** and **information objects**



# Lab IR settings focusing on actor variables and dimensions of IT and information objects



# Conclusion

- Context is endless and pervasive
  - its management requires perspective, goals and interests: what do we want to achieve??
  - we may see multiple, nested contexts
- Thus, what is IR about? As technology ...
  - developing systems
    - ✓ based on a technological imperative?
    - ✓ based on people's communication / access needs?
- As a discipline ... [vs. a technology]
  - understanding ... perhaps ... but what? C-n-c-l?

# Does context really matter?

- We do not know really (if, or how much)
- Some certainly wish / believe it does not
- I don't believe them until shown that
  - no matter what one's situation is, one's information needs are of just one type
  - no matter of what type information needs are, they are always best served in one already known way

Any comments  
or questions?