Aalto University
School of Business

Collective Minding through Automation

How to build Digital High Reliability Organizations?

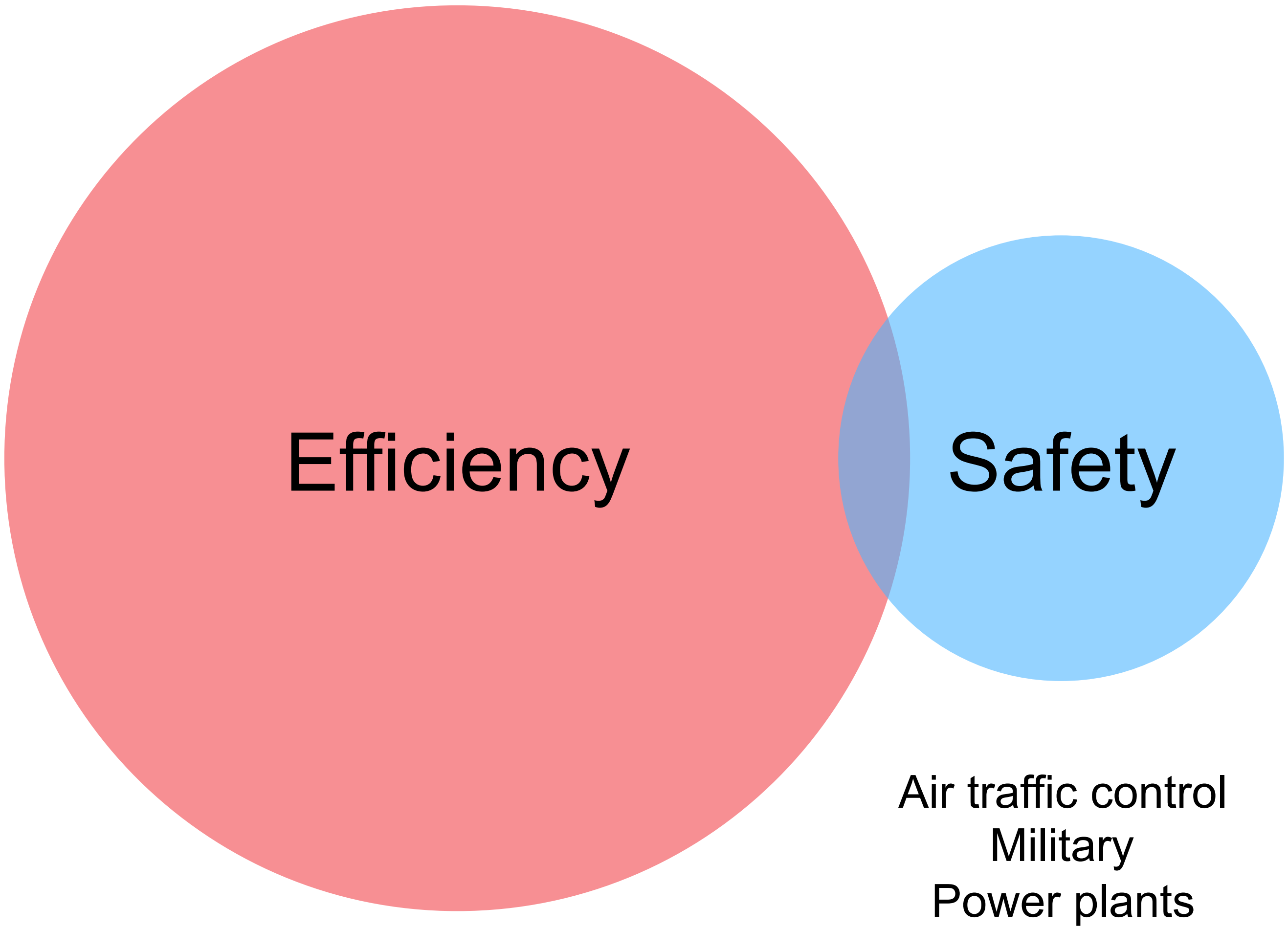
Antti Salovaara, Esko Penttinen, Kalle Lyytinen

RTE conference research track

May 27, 2016

ACADEMY
OF FINLAND

What organizations focus on



Air traffic control
Military
Power plants
Catastrophe relief

Malware protection

Efficiency

High volume

Geographical distribution

Time-criticality

Large customer base

Need to automate
the operations

Safety

Destruction of equipment

Denials of service

Thefts of confidential information

Blackmailing (“ransomware”)

But can you automate
and also ensure safety?

Approaches to ensure safety

Natural accident theory

Perrow, 1984

© Phys.org, http://phys.org/news/2010-11-airliners-cyber.html

Air traffic control room in Hong Kong

“Accidents are ultimately unavoidable”

“Complex systems with tightly coupled interdependencies always lead to errors”

E.g. when system states exceed their tolerance range

➡ Looser coupling

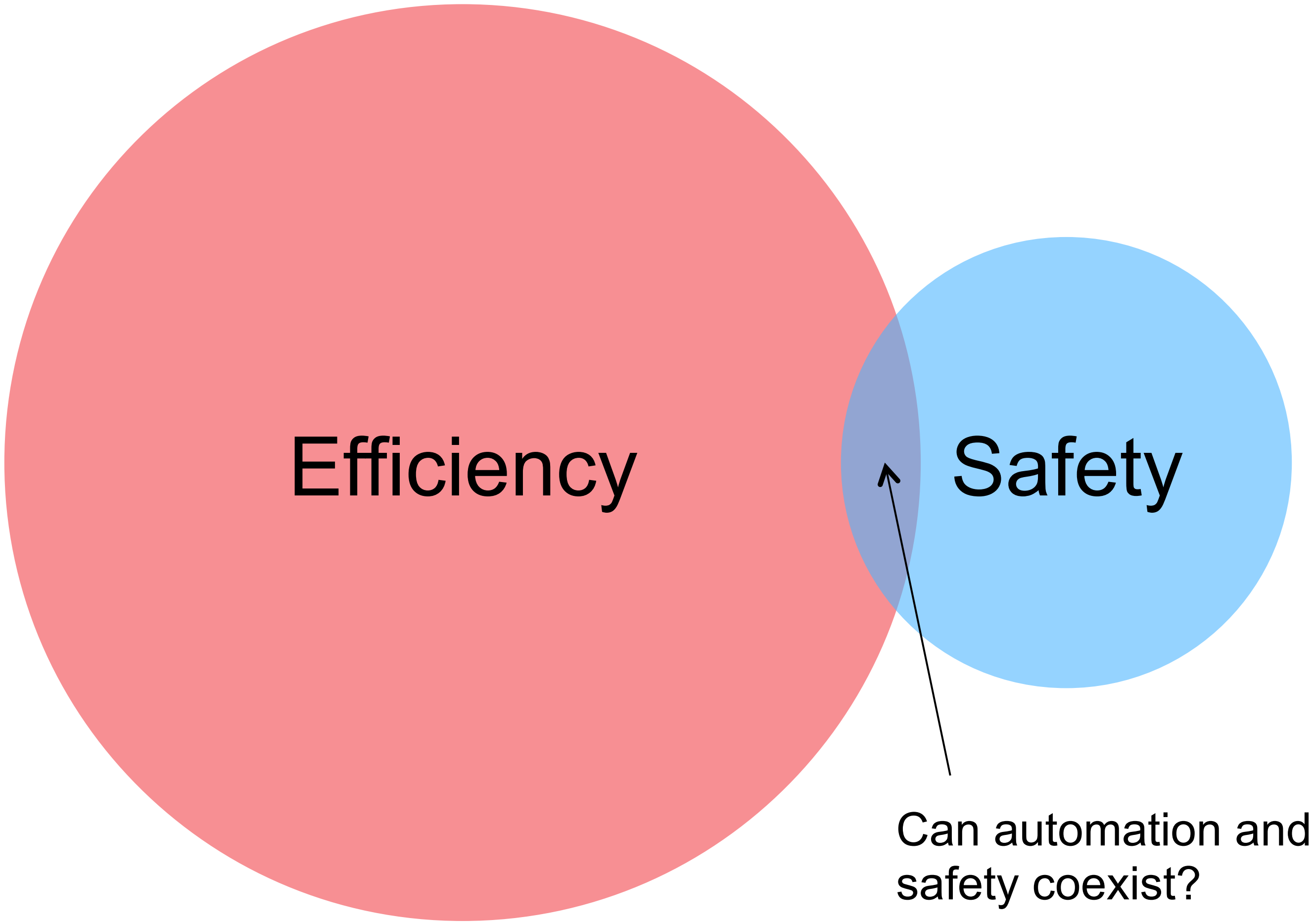
High reliability organizations (HROs)

Some organizations manage to resist the eventual failures that NAT predicts

Common characteristics:

1. Preoccupation with failure
2. Reluctance to simplify interpretations
3. Sensitivity to operations
4. Commitment to resilience
5. Underspecification of structures

“Collective mindfulness”



Efficiency

Safety

Can automation and
safety coexist?

Automation, IT
Mindlessness

Humans
Mindfulness

The problem of automation

“The boundary of my language represents the boundary of my world.”

–Wittgenstein: Tractatus Logico-Philosophicus, theorems 5.6 to 5.621

Frame problem:

≈ An artificial intelligence can never understand more than what its creators have endowed it with (McCarthy & Hayes, 1969)

➡ Automation is always limited by the frame problem *and is therefore unsafe*

Case study: F-Secure

Biggest malware protection company in Europe

~900 employees

F-Secure’s software rated best in malware protection on 4 years consecutive years

F-Secure seems to succeed in combining automation and safety (i.e., high reliability) !

Why is malware a hard case for high reliability?

Digital material is:

- Exact
- Transferable
- Editable
- Programmable

+ Malware is creatively created

= a challenge to the frame problem.

Methods

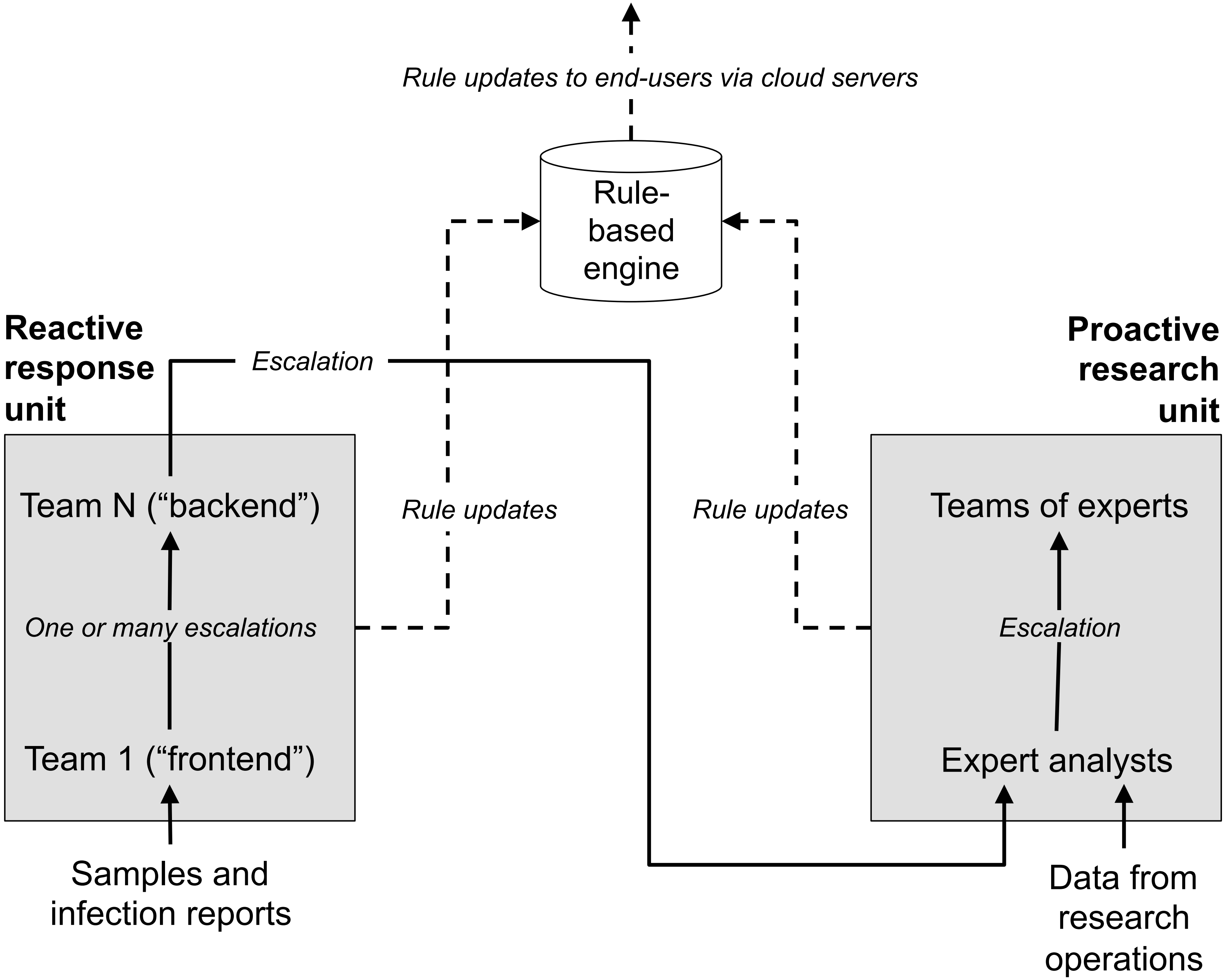
15 interviews

- malware analysts and managers
- 38–103 mins each, usually 60 mins

2 participant observations

- 4 and 6 hours

Analysis of the different malware protection operations

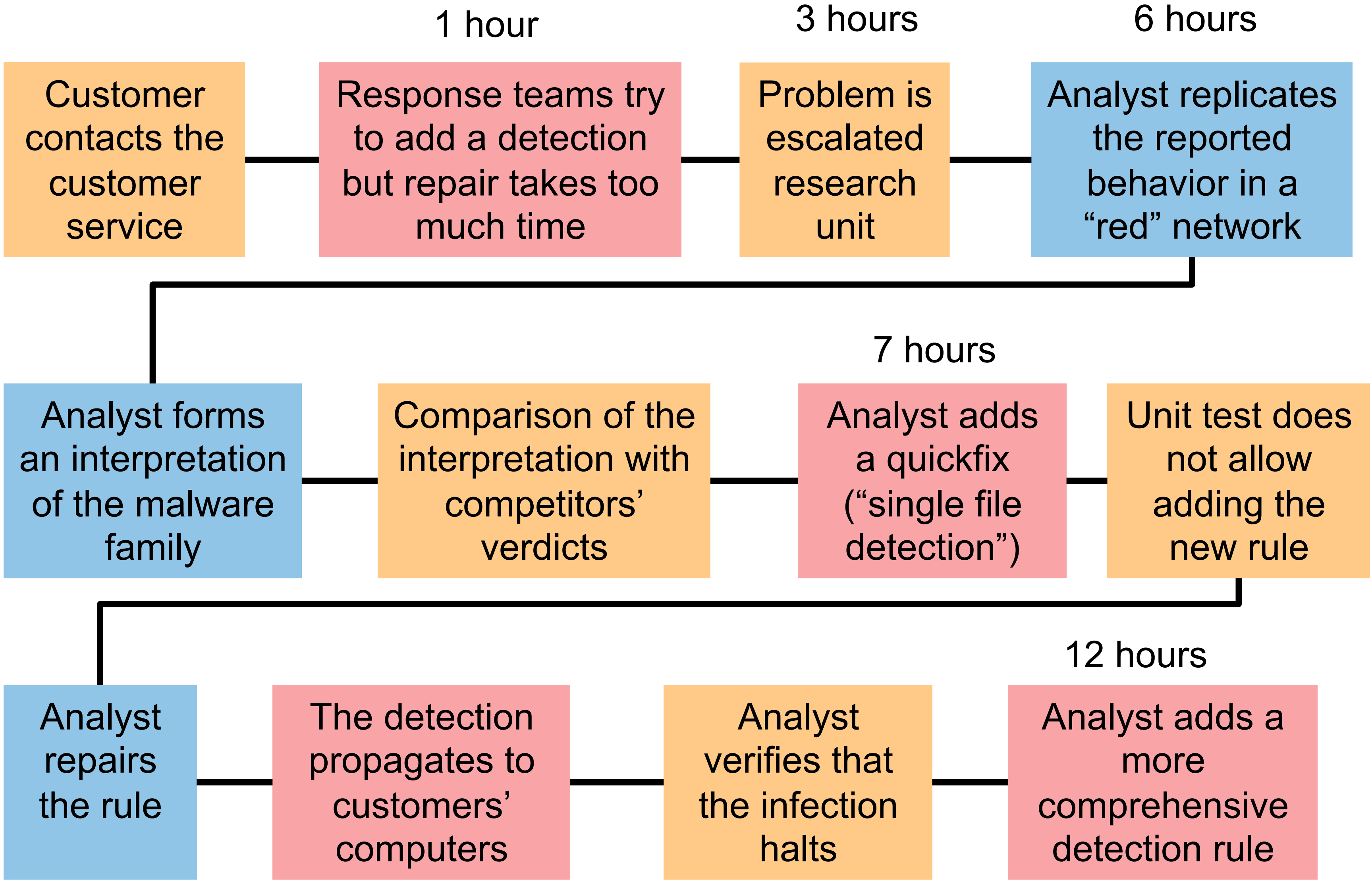

Rule updates to end-users via cloud servers
Rule-based engine
Reactive response unit
Escalation
Proactive research unit
Team N ("backend")
Rule updates
Rule updates
Teams of experts
One or many escalations
Escalation
Team 1 ("frontend")
Expert analysts
Samples and infection reports
Data from research operations

Findings

1. How F-Secure functions as a digital HRO
2. How does F-Secure succeed in automating operations in such a safety-critical domain?
3. What are the unique aspects of digital HROs

F-Secure as a digital HRO

A common story of a new malware case



Customer contacts the customer service

1 hour

Response teams try to add a detection but repair takes too much time

3 hours

Problem is escalated research unit

6 hours

Analyst replicates the reported behavior in a “red” network

Analyst forms an interpretation of the malware family

Comparison of the interpretation with competitors’ verdicts

7 hours

Analyst adds a quickfix (“single file detection”)

Unit test does not allow adding the new rule

Analyst repairs the rule

The detection propagates to customers’ computers

Analyst verifies that the infection halts

12 hours

Analyst adds a more comprehensive detection rule

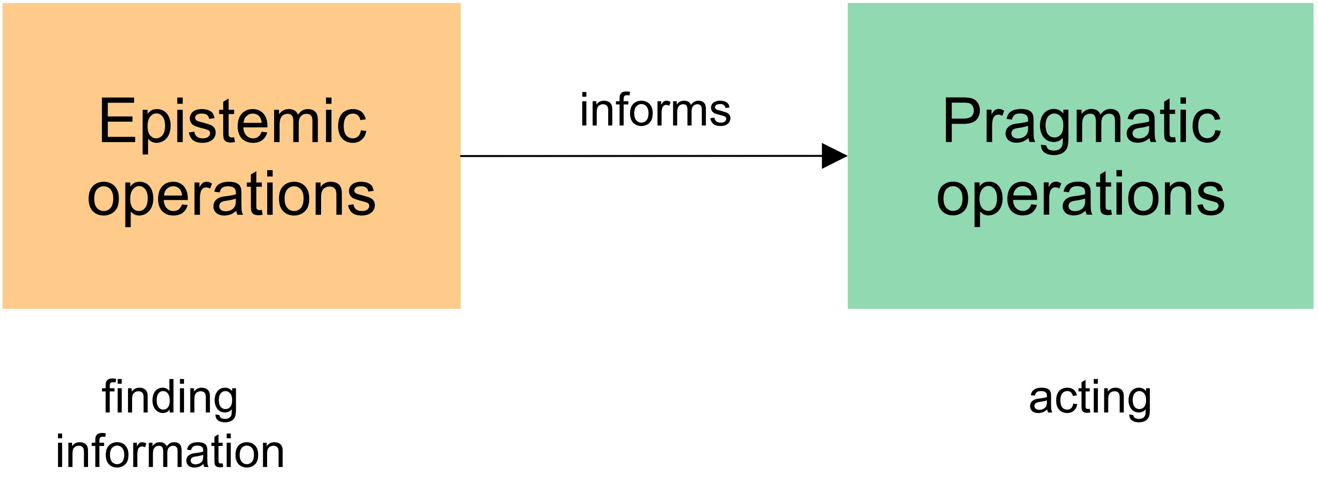
1. Preoccupation with failure	**1.1 Pre-testing new rules** 1.2 Creation of “soft” rules **1.3 Customer service** 1.4 Threat hunting 1.5 Honey pots 1.6 Algorithm analysis **1.7 Sample sharing with competitors** **1.8 Sample analysis within a closed “red” network**
2. Reluctance to simplify interpretations	**2.1. Sample hunting** **2.2. Replication of malware’s behavior** **2.3. Cross-validation of verdicts using competitors’ software** 2.4. Root cause and post mortem analyses
3. Sensitivity to operations	**3.1 Manual log monitoring** 3.2 Automatic log monitoring 3.3 Customer service (again)
4. Commitment to resilience	4.1. Ad hoc problem solving teams 4.2. Root cause and post mortem analyses (again) **4.3. Coding the lessons learned into automation**
5. Underspecification of structures	**5.1. Modification of detection rules** 5.2. False positive patching **5.3. Single file detections** **5.4. Escalation-based failure management**

How does F-Secure succeed in automating operations in such a safety-critical domain?

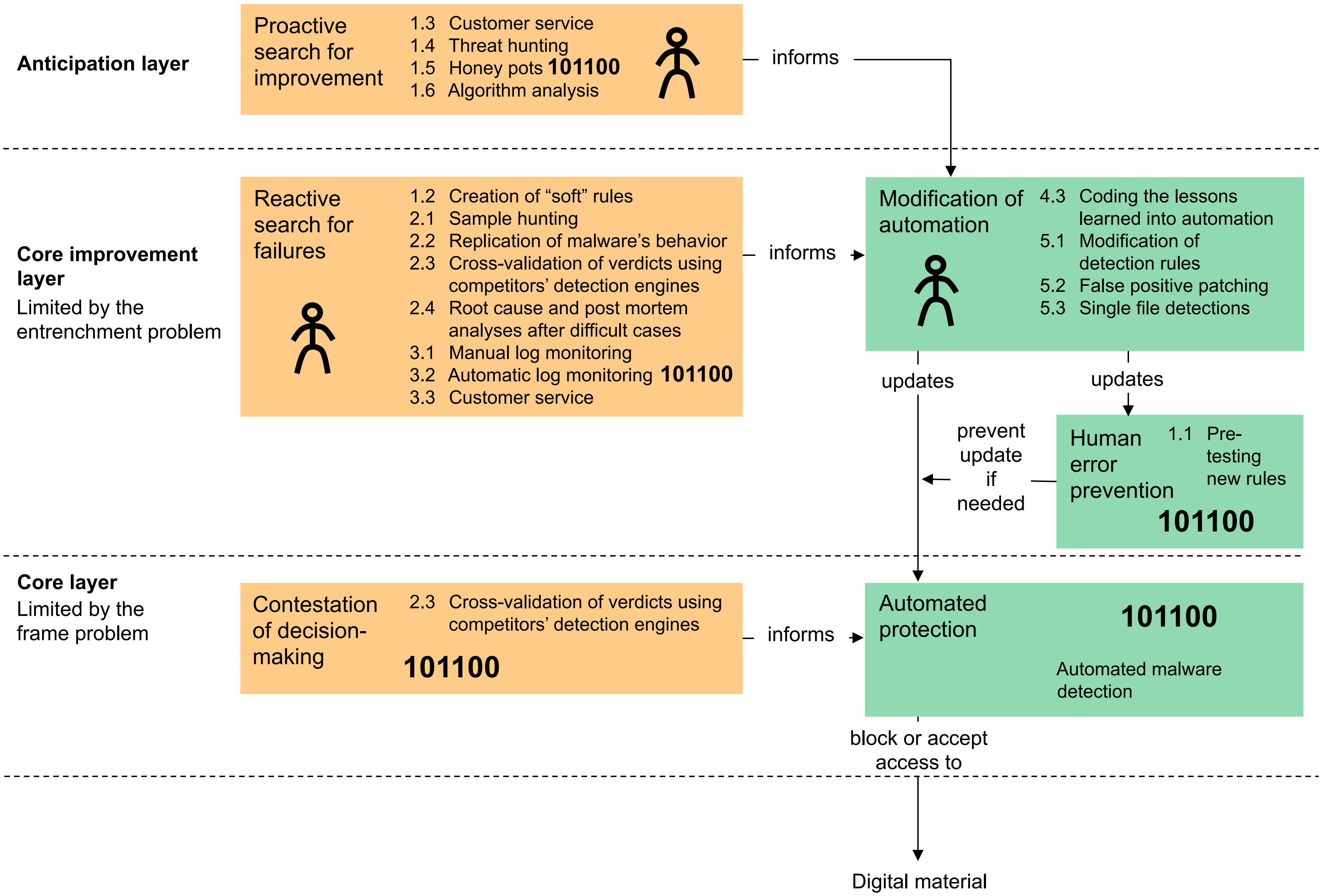
1. Preoccupation with failure	1.1 Pre-testing new rules 1.2 Creation of “soft” rules 1.3 Customer service 1.4 Threat hunting 1.5 Honey pots 1.6 Algorithm analysis 1.7 Sample sharing with competitors 1.8 Sample analysis within a closed “red” network
2. Reluctance to simplify interpretations	2.1. Sample hunting 2.2. Replication of malware’s behavior 2.3. Cross-validation of verdicts using competitors’ software 2.4. Root cause and post mortem analyses
3. Sensitivity to operations	3.1 Manual log monitoring 3.2 Automatic log monitoring 3.3 Customer service (again)
4. Commitment to resilience	4.1. Ad hoc problem solving teams 4.2. Root cause and post mortem analyses (again) 4.3. Coding the lessons learned into automation
5. Underspecification of structures	5.1. Modification of detection rules 5.2. False positive patching 5.3. Single file detections 5.4. Escalation-based failure management

Theory of actions (operations)

Kirsh and Maglio, 1994



1. Preoccupation with failure	1.1 🟩 Pre-testing new rules 1.2 🟧 Creation of “soft” rules 1.3 🟧 Customer service 1.4 🟧 Threat hunting 1.5 🟧 Honey pots 1.6 🟧 Algorithm analysis 1.7 🟧 Sample sharing with competitors 1.8 Sample analysis within a closed “red” network
2. Reluctance to simplify interpretations	2.1. 🟧 Sample hunting 2.2. 🟧 Replication of malware’s behavior 2.3. 🟧 Cross-validation of verdicts using competitors’ software 2.4. 🟧 Root cause and post mortem analyses
3. Sensitivity to operations	3.1 🟧 Manual log monitoring 3.2 🟧 Automatic log monitoring 3.3 🟧 Customer service (again)
4. Commitment to resilience	4.1. Ad hoc problem solving teams 4.2. 🟩 Root cause and post mortem analyses (again) 4.3. 🟩 Coding the lessons learned into automation
5. Underspecification of structures	5.1. 🟩 Modification of detection rules 5.2. 🟩 False positive patching 5.3. 🟩 Single file detections 5.4. Escalation-based failure management



Anticipation layer

Proactive search for improvement

1.3 Customer service
1.4 Threat hunting
1.5 Honey pots **101100**
1.6 Algorithm analysis

informs

Core improvement layer
Limited by the entrenchment problem

Reactive search for failures

1.2 Creation of “soft” rules
2.1 Sample hunting
2.2 Replication of malware’s behavior
2.3 Cross-validation of verdicts using competitors’ detection engines
2.4 Root cause and post mortem analyses after difficult cases
3.1 Manual log monitoring
3.2 Automatic log monitoring **101100**
3.3 Customer service

informs

Modification of automation

4.3 Coding the lessons learned into automation
5.1 Modification of detection rules
5.2 False positive patching
5.3 Single file detections

updates

updates

Human error prevention

1.1 Pre-testing new rules

101100

prevent update if needed

Core layer
Limited by the frame problem

Contestation of decision-making

2.3 Cross-validation of verdicts using competitors’ detection engines

101100

informs

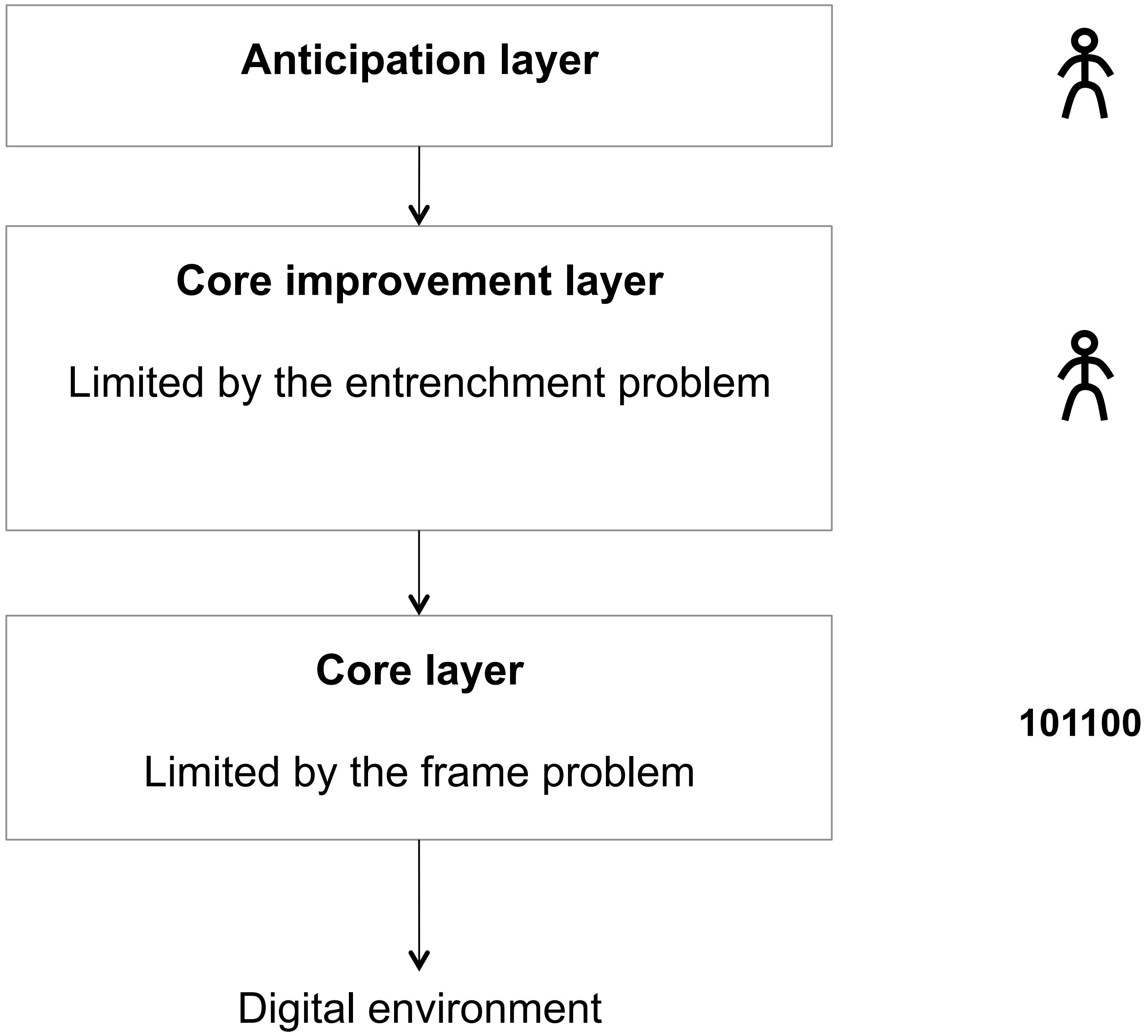
Automated protection

101100

Automated malware detection

block or accept access to

Digital material



Anticipation layer

Core improvement layer

Limited by the entrenchment problem

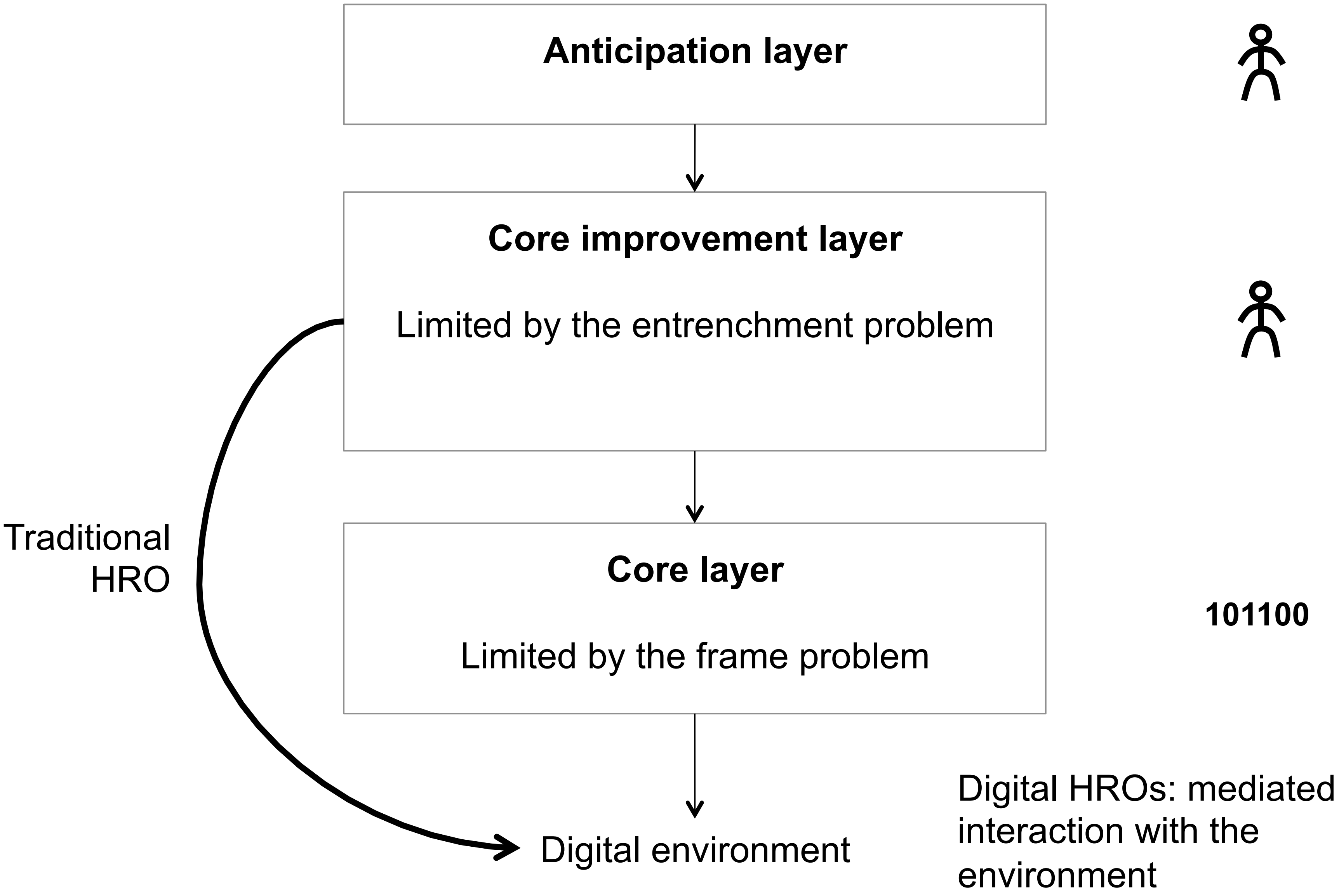
Core layer

Limited by the frame problem

101100

Digital environment

What are the unique aspects of digital HROs?



Anticipation layer

Core improvement layer

Limited by the entrenchment problem

Core layer

Limited by the frame problem

101100

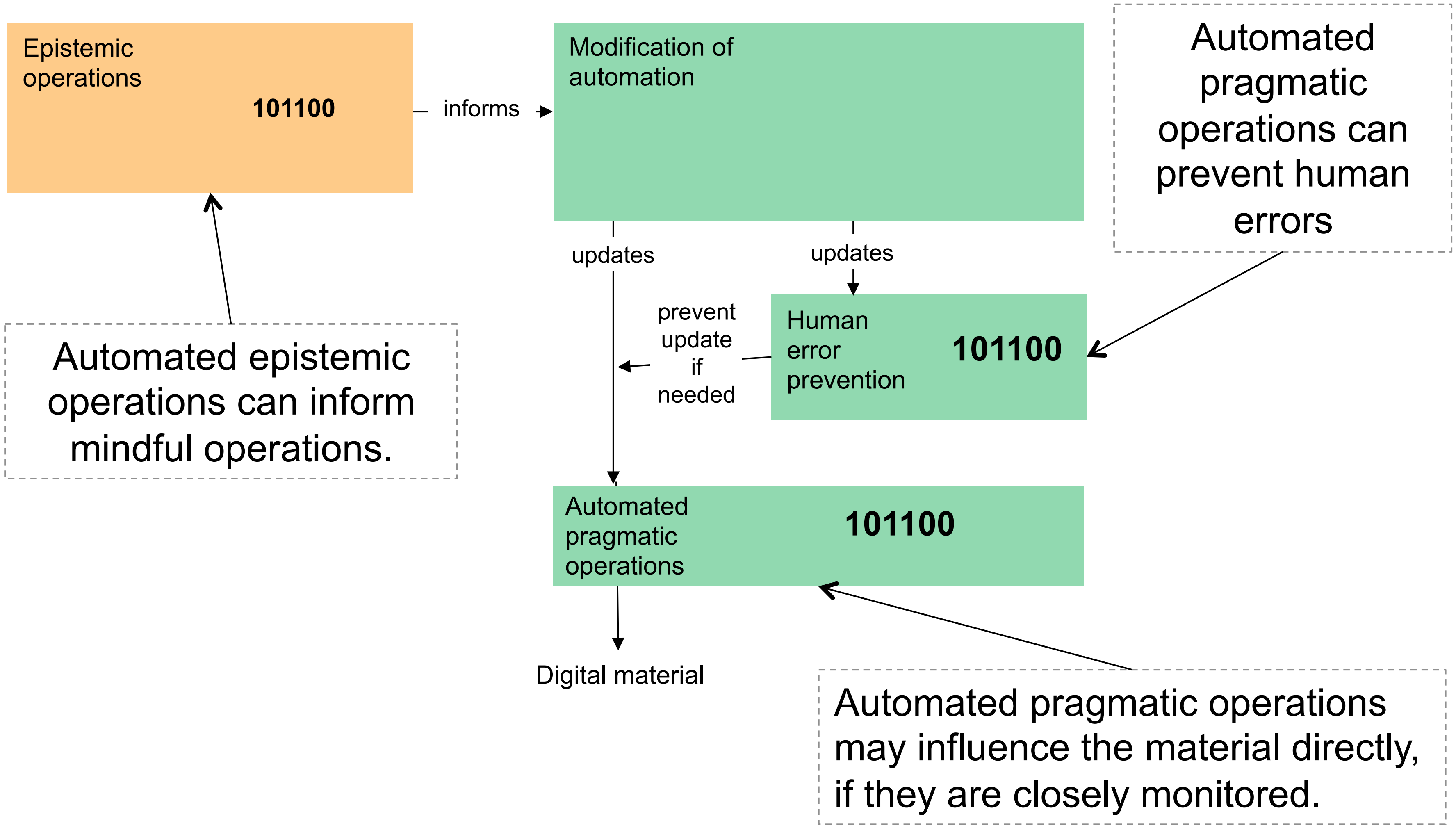
Traditional HRO

Digital environment

Digital HROs: mediated interaction with the environment

Discussion of the findings

How automation can support mindfulness and high reliability?



Epistemic operations

101100

informs

Modification of automation

updates

updates

Human error prevention

101100

prevent update if needed

Automated pragmatic operations

101100

Digital material

Automated pragmatic operations can prevent human errors

Automated epistemic operations can inform mindful operations.

Automated pragmatic operations may influence the material directly, if they are closely monitored.

Summary of contributions

Importance of the frame problem for digital organizations

First study of a digital HRO

First study on malware protection operations

First study to investigate HROs on a level of operations and show how the collective mindfulness emerges

How automation can be incorporated in an HRO

Antti Salovaara
antti.salovaara@alumni.aalto.fi

Esko Penttinen
esko.penttinen@aalto.fi

Kalle Lyytinen
kjl13@case.edu