
Federal Institute for Occupational Safety and Health


Federal Institute for Occupational Safety and Health

# Diffusion of Responsibility in Human-Robot Interaction

A PhD Project – first steps in a series of studies

# Overview

1. **Diffusion of responsibility – a definition**
2. **Assumptions from literature**
   1. Connection between intentionality and sense of agency
   2. Levels of transparency and diffusion of responsibility
3. **Goals and Hypotheses**
4. **Research methodology**
   1. Technical implementation
   2. Study design
   3. Data collection methods
5. **Current status**
6. **Conclusion and next steps**

# Diffusion of Responsibility – A Definition

**The phenomenon of diffusion of responsibility (DOR) has been extensively studied in both, sociology and psychology and describes the reduction of an individual's perceived sense of responsibility in the presence of additional people.**

- **Personally invested effort decreases because others could be taken responsible (bystander effect)**
  - Cognitive justification strategy
  - Motivational and personality factors (perceived (in)competence, self-efficacy, feelings of self-endangerment, community orientation)
  - Social influence

- **A higher degree of autonomy (e.g. self-learning systems) leads to an increase in the diffusion of responsibility**

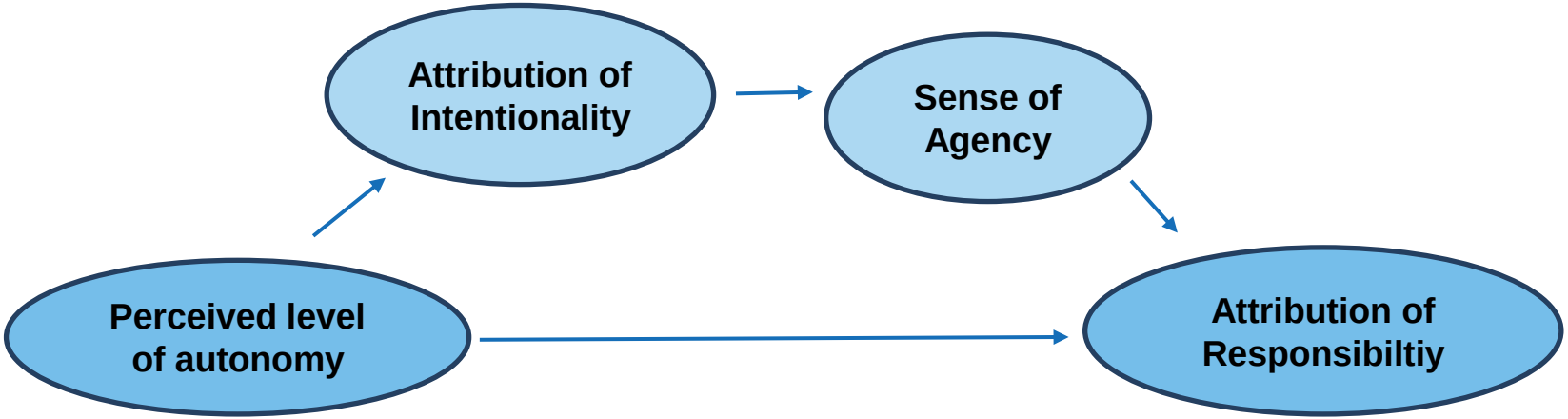
# Assumptions from Literature I

# Connection between Intentionality & Diffusion of Responsibility

**The term intentionality is used in philosophy of mind when dealing with ideas and mental acts that are directed towards objects, persons or states of affairs. It plays an important role in explaining and predicting their behavior.**

- **The “intentional stance” (Dennett, 1987) is an explanatory strategy for (human) behaviour**
  - Machines (anthropomorphic robots in particular) can trigger the "intentional stance" and thereby receive higher acceptance, trust, and empathy from humans
  - Negative outcomes are more often attributed externally (to the machine) than positive ones
  - The attribution of mental states can change over time
  - Reduces **Sense of Agency** (SoA)

# Assumptions from Literature II

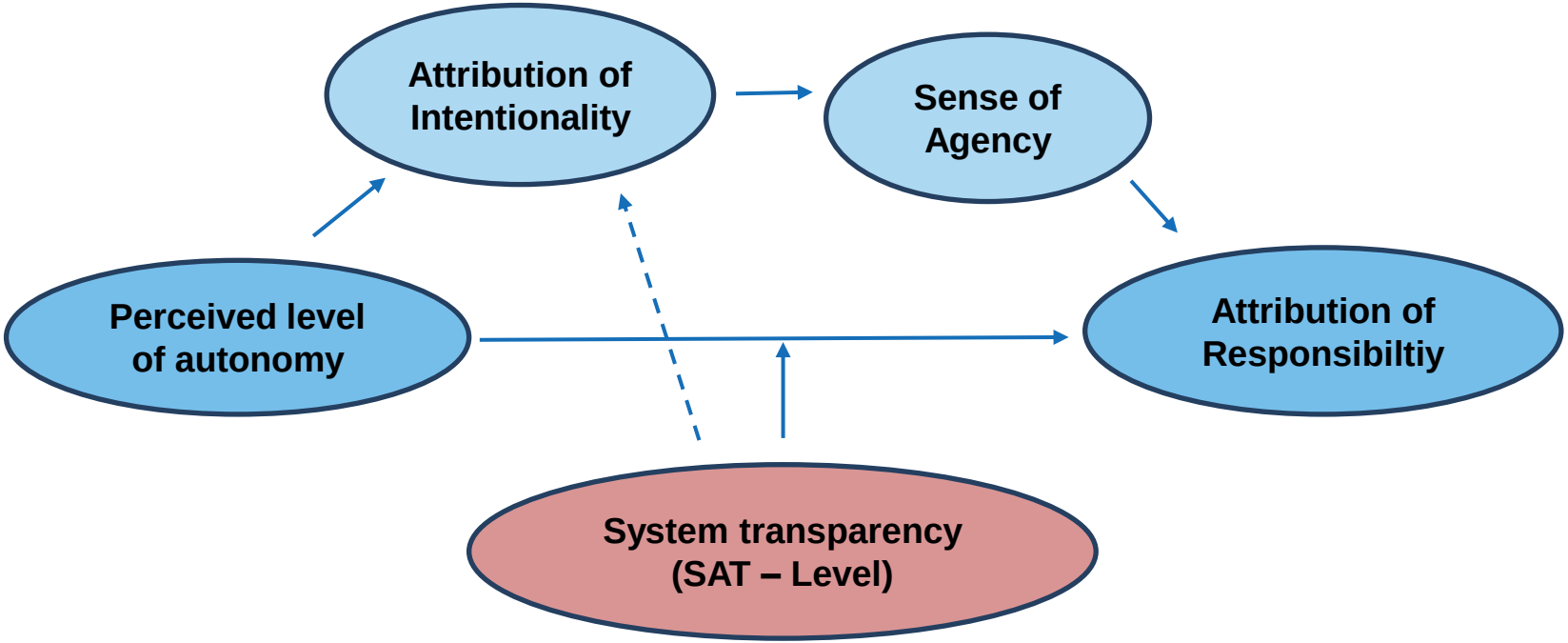


# Transparency and Attribution of Responsibilty

- **A higher degree of transparency can make it harder to avoid taking responsibility**
- **SAT model (Chen, Lakhmani et al. 2018) describes three levels of transparency**
  1. Basic information
  2. Rationale
  3. Consequences
- **Work-related MRI has been studied with regard to transparency only rarely and with conflicting results**

**→ A concept that describes how transparency should be designed in order to minimize the risk of false attribution of responsibility seems overdue and also essential to ensure effective cooperation between humans and increasingly autonomous machines**

# Assumptions from Literature III



# Goals and Hypotheses

**Overarching goal of PhD thesis:** Designing a model for system transparency in industrial human-robot interaction.

**Goal of 1st study:** Understanding the relationship between the robot's perceived autonomy and the attribution of intentionality as well as responsibility.

**Hypothesis 1:** The higher the perceived autonomy, the stronger the intentional attitude towards the robot

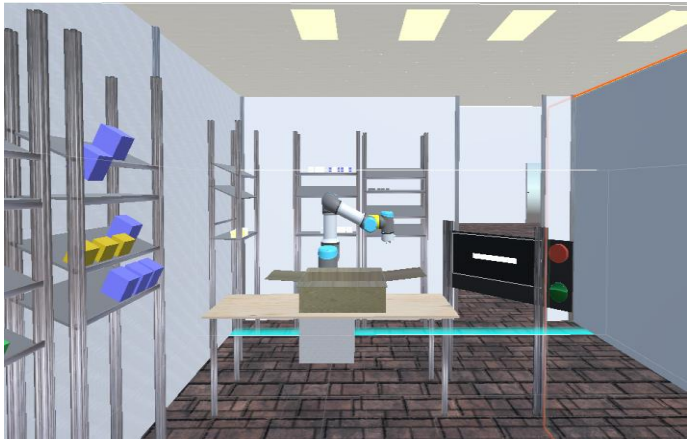
**Hypothesis 2:** People adopt a significantly stronger intentional stance when told they are working with an AI robot than when AI is not mentioned.

**Hypothesis 3:** The stronger the intentional attitude towards the robot, the higher the attribution of responsibility for errors to the robot.

**Hypothesis 4:** Humans perceive a higher level of agency (SoA) when the robot is less autonomous

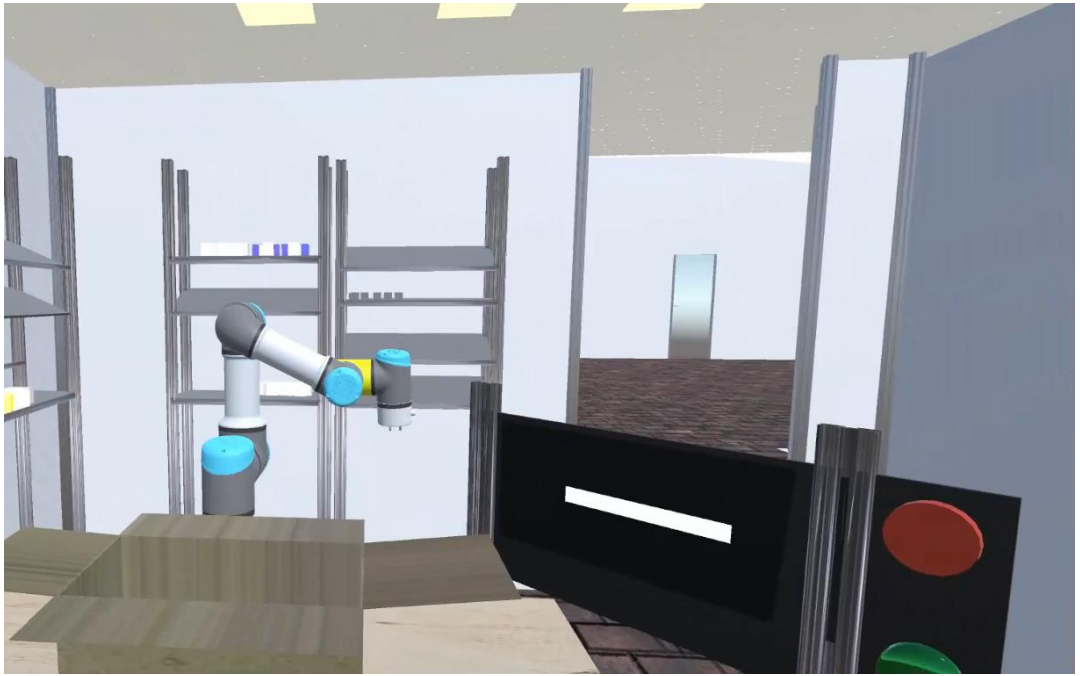
# Technical Implementation

- **Technical implementation in Virtual Reality (VR)**
  - Game engine used: Unity 2021.3
  - Program language C#
  - Design and Animations in Blender
- **Work station was designed to match the real room, robot and surroundings (2mx2m)**



**Scene in virtual reality**

# Impression of Virtual Interaction



# Study Design

- **Interaction setting and task**
  - Assembly task with a robotic arm (UR5) with the goal to achieve a specific weight without exceeding it for further processing
  - Participants are randomly assigned to either **Condition 1** (high autonomy) or **Condition 2** (low autonomy)
  - Both conditions involve the robot releasing the box in 60% of cases, with different criteria for condition-specific behaviors
- **Scoring system**
  - Submitting a box below the target weight deducts points (max. 100)
  - Points deducted are influenced by the filled weight, following compensation strategy from previous studies
- **Duration and data collection**
  - The study will take about 2h with approx. 1h in VR
  - Questions are asked before, during, and after the interaction

# Data Collection Methods

**Before the interaction:** Sociodemographic data, Experience with robotic systems, Individual differences for humanisation (IDAQ), Affinity for Technology Interaction Short Scale (ATI-S)

→ **Likert Scales**

**During the interaction:** Diffusion of Responsibility, Sense of Agency

→ **Visual Analog Scales**

**After the interaction:** Perception of autonomy, Importance of the task, Intentionality (HRIES), mental load (NASA-TLX), Usability (SUS), Immersion and Presence, Trust

→ **Likert Scales and Visual Analog Scales**

# Conclusion

- **With the increasing integration of robots into different work environments, understanding how people perceive and interact with these autonomous systems is crucial to ensuring a safe and healthy workplace.**
- **Diffusion of responsibility occurs more often when robots act (seemingly) autonomous**
- **(Seemingly) autonomous robots are more likely to induce the intentional stance**
- **Increased system transparency might reduce intentional attitude.**
- **Shift from intentional to functional attitude (tool-like interaction) might enhance appropriate attribution of responsibility**

# References


Beer, J. M., Fisk, A. D., & Rogers, W. A. (2014). Toward a framework for levels of robot autonomy in human-robot interaction. Journal of human-robot interaction, 3(2), 74–99.

Beyer, F., et al. (2017). "Beyond self-serving bias: diffusion of responsibility reduces sense of agency and outcome monitoring." 12(1): 138-145.

Bierhoff, H. W., Klein, R., & Kramp, P. (1990). Hemmschwellen zur Hilfeleistung-Untersuchung der Ursachen und Empfehlung von Maßnahmen zum Abbau.

Bierhoff, H. W., & Rohmann, E. (2017). Diffusion von Verantwortung. In Handbuch verantwortung (pp. 911-931). Springer VS, Wiesbaden

Ciardo, F., et al. (2020). "Attribution of intentional agency towards robots reduces one's own sense of agency." 194: 104109

Dennett, D. C. (1987). The intentional stance, MIT press

Karau, S. J., & Williams, K. D. (1993). Social loafing: A meta-analytic review and theoretical integration. Journal of personality and social psychology, 65(4), 681.

Kiesler, S., et al. (2008). "Anthropomorphic interactions with a robot and robot–like agent." 26(2): 169-181

Kim, T., & Hinds, P. (2006, September). Who should I blame? Effects of autonomy and transparency on attributions in human-robot interaction. In ROMAN 2006-The 15th IEEE international symposium on robot and human interactive communication (pp. 80-85). IEEE.

Latané, B., & Nida, S. (1981). Ten years of research on group size and helping. Psychological bulletin, 89(2), 308

Lemaignan, S., et al. (2014). The cognitive correlates of anthropomorphism. 2014 Human-Robot Interaction Conference, Workshop" HRI: a bridge between Robotics and Neuroscience".

Matsui, T. and A. J. S. Koike (2021). "Who is to blame? The appearance of virtual agents and the attribution of perceived responsibility." 21(8): 2646

Perez-Osorio, J. and A. J. P. P. Wykowska (2020). "Adopting the intentional stance toward natural and artificial agents." 33(3): 369-395.

Pöhler, G., Heine, T., & Deml, B. (2016). Itemanalyse und Faktorstruktur eines Fragebogens zur Messung von Vertrauen im Umgang mit automatischen Systemen. *Zeitschrift für Arbeitswissenschaft*, *3*(70), 151-160.

Spatola, N., et al. (2021). "Perception and evaluation in human–robot interaction: The Human–Robot Interaction Evaluation Scale (HRIES)—A multicomponent approach of anthropomorphism." 13(7): 1517-1539

Waytz, A., Morewedge, C. K., Epley, N., Monteleone, G., Gao, J. H., & Cacioppo, J. T. (2010). Making sense by making sentient: effectance motivation increases anthropomorphism. Journal of personality and social psychology, 99(3), 410

# Data Collection Methods

## Attribution of Responsibility

1. To what extent do you feel that the robot is responsible for the outcome of the task?

2. To what extent do you feel responsible for the outcome of the task?

## Sense of Agency

3. To what extent did you feel you had control over the completion of the boxes?

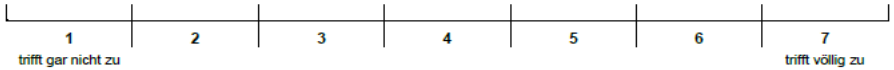
**Visual Analog Scales**

# Data Collection Methods

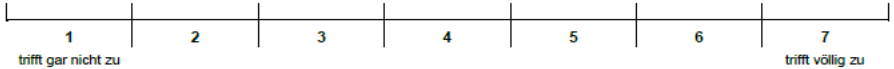
## Trust

**Classic Likert Scales**

**2. Das System verhält sich undurchsichtig.**




**3. Ich misstraue den Entscheidungen des Systems.**




## IDAQ

*By "has free will" we mean able to choose and control its own actions.*

*By "active" we mean moving frequently and quickly*.

1. To what extent is the average cat active

2. To what extent does the average mountain have free will?

0----1----2----3----4----5----6----7----8----9----10