Theory of mind and information relevance in human centric human robot cooperation

Vom Fachbereich
Elektrotechnik und Informationstechnik
der Technischen Universität Darmstadt
zur Erlangung des akademischen Grades
eines Doktor-Ingenieurs (Dr.-Ing.)
genehmigte Dissertation

von

M. Sc. Moritz Clemens Bühler

geboren am 2. Februar 1991 in Heidelberg

Referent:	Prof. Dr.-Ing. J. Adamy
Korreferent:	Prof. Dr. rer. nat. B. Sendhoff
Tag der Einreichung:	10. Januar 2022
Tag der mündlichen Prüfung:	10. Mai 2022

D17
Darmstadt 2022

Bühler, Moritz : Theory of mind and information relevance in human centric human robot cooperation
Darmstadt, Technische Universität Darmstadt,
Fachgebiet Regelungsmethoden und Robotik
Jahr der Veröffentlichung der Dissertation auf TUprints: 2022
URN: urn:nbn:de:tuda-tuprints-214324
URI: https://tuprints.ulb.tu-darmstadt.de/id/eprint/21432
Tag der mündlichen Prüfung: 10. Mai 2022

Danksagung

Diese Dissertation enstand in meiner Zeit am Fachgebiet für Regelungsmethoden und Robotik der TU Darmstadt in Forschungskooperation mit dem Honda Research Institut Europe GmbH. Dazu möchte ich mich bedanken bei Professor Adamy, für die Freiheit und das gewährte Vertrauen, sowie die sehr gute Arbeitsatmosphäre am Institut. Mein Dank gilt weiterhin Professor Sendhoff für die Unterstützung der Forschung und Übernahme des Korreferats.

Danken möchte ich Thomas Weisswange für die zahlreichen Diskussionen und eine gute Zusammenarbeit. Danke an Valentina für die visuelle Beratung und Gestaltung der Grafiken. Danke an Moritz Knaust und Thomas Weisswange für das Korrekturlesen und die wertvollen Rückmeldungen. Gerne denke ich zurück an die Gespräche und Kaffeepausen mit den Kollegen und den intensiven Austausch mit Fabian, Sebastian, Niko, Dirk und Moritz. Zuletzt möchte ich mich bedanken bei allen Personen, die mich begleitet, unterstützt und geprägt haben, insbesondere auch meinen Eltern und meiner Frau.

"It is the obvious which is so difficult to see most of the time. People say 'It's as plain as the nose on your face.' But how much of the nose on your face can you see, unless someone holds a mirror up to you?"

— Isaac Asimov, I, Robot

Contents

Abbreviations and symbols **VII**

Abstract **X**

Kurzfassung **XII**

1 Introduction **1**
1.1 Overview and contributions 5

2 Interacting with humans in uncertain environments **8**
2.1 Human factors and human interaction 8
2.1.1 Human factors and technology 8
2.1.2 Human–human interaction 13
2.2 Coping with uncertainty 17
2.2.1 Bayesian inference in probabilistic models 17
2.2.2 Acting under uncertainty 24
2.2.3 Uncertainty by other agents and strategic interaction 31
2.3 Modeling human behavior for cognitive human–robot interaction 35
2.3.1 Complex rational human models 35
2.3.2 Human models in human–robot interaction 38
2.4 Communication strategies with humans and other agents . 41
2.5 Summary and conclusion 45

3 Artificial Theory of Mind as online inference of human belief **47**
3.1 Belief definition and representation 49
3.1.1 Grid world illustrative example 50
3.1.2 Factorization of human belief 52
3.2 Human model 53
3.2.1 Human belief transition 54
3.2.2 Human perception 54
3.2.3 Human decision making 55

3.3 Belief inference 56
3.3.1 Human belief particle filter 57
3.3.2 Linearization-based inference 58
3.4 Quantitative formulation of situation awareness 62
3.5 Illustration 63
3.5.1 Belief inference 64
3.5.2 Estimation of human situation awareness 64
3.5.3 Inference methods and computation times 65
3.5.4 Rationality parameter 70
3.6 Summary and conclusion 71

4 Theory of mind based assistive communication 72
4.1 Effects and efforts of communication 73
4.2 Communication planning for information sharing decisions . 76
4.3 Illustration 77
4.3.1 Alternative communication concepts 77
4.3.2 What to communicate 78
4.3.3 When to (not) communicate 79
4.4 Summary and conclusion 86

5 Human centric human–robot cooperation 88
5.1 Belief inference and situation awareness estimation in a sequential human–robot cooperative task 90
5.1.1 Sequential cooperative task 90
5.1.2 Application of belief inference 93
5.1.3 User study 94
5.2 Assistive communication in a complex human robot cooperative sushi task 98
5.2.1 Sushi task description 99
5.2.2 Human model and robot behavior 102
5.2.3 User study 107
5.2.4 Performance improvement by human centric assistance 111
5.2.5 Comparison of information sharing concepts 116
5.3 Summary and conclusion 119

6 Conclusion 121

A Questionnaire and results 123

Bibliography 128

Abbreviations and symbols

Abbreviations

H	Human
IRL	Inverse Reinforcement Learning
MDP	Markov Decision Process
POMDP	Partially Observable Markov Decision Process
R	Robot
SA	Situation Awareness
ToM	Theory of Mind
ToM-Com	Theory of Mind based Communication

Symbols

Latin uper cases

A	Action Space
E	Expectation of a random variable
J	Accumulated reward
K	Number of particles / samples
$\mathcal{N}$	Normal distribution
N	Number of discrete states respectively dimension of continuous beli
O	Observation function
Q	Action value function
R	Reward function
S	State space
T	State transition function
$\mathbf{T}$	State transition matrix
V	State value function

Latin lower cases

a	Action
b	Belief
$\mathbf{b}$	Belief vector
d	Distance function
f	Nominal continuous transition function
h	Nominal continuous measurement function
k	Discrete time step
l	Likelihood function
n	Noise
o	Observation
p	Probability distribution / probability density function
s	System state
w	Weight
$\mathbf{z}$	Transformed belief

Greek letters

α	Parameters of Dirichlet distribution
δ	Situation awareness threshold
γ	Discount factor
Ω	Observation space
π	Policy
μ	Mean
θ	Model parameters
τ	Rationality parameter, weight of softmax function
Σ	Covariance

Abstract

In the interaction with others, besides consideration of environment and task requirements, it is crucial to account for and develop an understanding for the interaction partner and her state of mind. An understanding of other's state of knowledge and plans is important to support efficient interaction activities including information sharing, or distribution of subtasks.

A robot cooperating with and supporting a human partner might decide to communicate information that it has collected. However, sharing every piece of information is not feasible, as not all information is both, currently relevant and new for the human partner, but instead will annoy and distract her from other important activities. An understanding for the human state of mind will enable the robot to balance communication according to the needs of the human partner and the efforts of communication for both.

An artificial theory of mind is proposed as Bayesian inference of human beliefs during interaction. It relies on a general model for human information perception and decision making. To cope with the complexity of second order inference – estimating what the human inferred of her environment – an efficient linearization based filtering approach is introduced. The inferred human belief, as understanding of her mental state, is used to estimate her situation awareness. When this is missing, e.g. the human is unaware of some important piece of information, the robot provides supportive communication. It therefore evaluates relevance and novelty of information compared to communication efforts following a systematic information sharing concept. The robot decides about whether, when and what type of information it should provide in a current situation to support the human efficiently without annoying. The decision is derived by planning under uncertainty while considering the inferred human belief in relation to the task requirements. Systematic properties and benefits of the derived concepts are discussed in illustrative example situations.

Two human robot collaborative tasks and corresponding user studies were designed and investigated, applying artificial theory of mind as belief inference and assistive communication in the interaction with humans.

Equipped with the artificial theory of mind, the robot is able to infer interpretable information about the human's mental state and can detect a lack of human awareness. Supported by adaptive human centric information sharing, participants could recover much earlier from unawareness. A comparison to state-of-the-art communication strategies demonstrates the efficiency, as the new concept explicitly balances benefits and costs of communication to support while avoiding unnecessary interruptions. By sharing information according to human needs and environmental urgency, the robot does not take over nor instruct the human, but enables her to make good decisions herself.

Kurzfassung

Bei der Interaktion mit Kooperationspartnern ist es wichtig, nicht nur die aktuelle Situation und Aufgaben zu berücksichtigen, sondern auch ein Verständnis für den Partner zu entwickeln. Solch ein Verständnis für die Pläne und den Kenntnisstand der anderen ermöglicht erst erfolgreiche kooperative Tätigkeiten, wie z.B. das Teilen von relevanten Informationen oder eine sinnvolle Arbeitsteilung. Mit den gestiegenen technischen Möglichkeiten und interaktiven Anwendungsfeldern wird es auch für Roboter oder andere komplexe technische Systeme relevant, Interaktionsaspekte zu berücksichtigen. Ein Roboter oder Assistenzsystem kann z.B. einem Menschen gewisse Informationen mitteilen. Dabei ist es jedoch nicht sinnvoll und auch nicht möglich, jede Einzelheit zu kommunizieren, da dies zu viel Zeit sowie kognitive Resourcen erfordern würde. Gleichzeitig ist auch nicht jede Information neu und relevant und würde den Partner nur stören oder ablenken. Stattdessen ist es notwendig, die Erfordernisse des Menschen in der gegebenen Situation zu berücksichtigen um zu erkennen, ob zusätzliche Informationen oder anderweitige Unterstützung benötigt wird.

In dieser Arbeit wird ein Verständnis für Interaktionspartner, eine "Theory of Mind" eingeführt als Bayssche Inferenz des menschlichen Belief bzw. Kenntnisstands. Es basiert auf dem beobachteten Verhalten während der Interaktion und einem allgemeinen Modell für Informationsaufnahme und Entscheidungsfindung des Menschen in Hinblick auf eine gegebene Aufgabe. Eine besondere Herausforderung besteht dabei in der Komplexität der hierarchischen probabilistischen Inferenz, der Inferenz der menschlichen Inferenz. Dazu werden verschiedene Approximationen eingeführt, unter anderem ein linearisierter Filterentwurf. Mit der Schätzung des menschlichen Beliefs wird es möglich, das Situationsverständnis des Menschen zu bewerten, bzw. zu erkennen, welche wichtigen Informationen ihm fehlen. Darauf aufbauend wird ein intelligentes Kommunikationskonzept entworfen, welches relevante Informationen gemäß der geschätzten Notwendigkeit zur Verfügung stellt. Dies beinhaltet das Abwägen des kommunikativen Aufwandes gegenüber dem erwarteten Nutzen. Durch probabilistisches Planen unter Unsicherheit gelangt der Roboter zu der Ent-

scheidung, ob, wann und welche Informationen geteilt werden sollen. Das grundsätzliche Kommunikationsverhalten und konzeptuelle Vorteile der neuen Konzepte werden an illustrativen Beispielen präsentiert und diskutiert.

Um diese Konzepte und Methoden zu testen, wurden zwei kooperative Mensch-Roboter-Studien aufgesetzt und durchgeführt. Mittels Theory of Mind ist der Roboter in der Lage, das Verhalten des Menschen zu interpretieren und das Fehlen von situationsrelevanten Informationen zu erkennen. Durch das Teilen relevanter Inhalte wurden die Versuchsteilnehmer der zweiten Studie unterstützt und waren in der Lage, Fehler schneller zu erkennen und zu korrigieren. Im Gegensatz zu anderen Kommunikationskonzepten wird die Relevanz von Informationen explizit berücksichtigt, um zielgerichtet und effizient zu helfen, während unnötige Unterbrechungen vermieden werden. Durch eine solche frühzeitige, aktive Intervention ist es oft nicht nötig, direkt in die Aktionen des Menschen einzugreifen oder zu instruieren. Stattdessen ermöglicht die intelligente bedarfsgerechte Kommunikation dem Partner das Treffen von eigenen, kompetenten Entscheidungen.

1 Introduction

Technology plays an important role in our daily and working life, with increasing functionality, complexity, and degree of automation. Examples include vacuum cleaning robots, recommendation systems or advanced driver assistance systems. These new technologies facilitate many tasks and support us in many situations. With more complex functionality however, interacting with it becomes more complex and new problems arise such as a loss of transparency and unclear responsibilities (e.g., missing coordination of which agent needs to account for which subtask). While classical industry robots operate in dedicated areas (spatial separation), they can work on predefined tasks with high precision and low uncertainty about environment and task. Despite their technical complexity, they are used as tools as they fulfill a single specific, delegated subtask (functional separation). Likewise, vacuum cleaning robots work on a clear task, however, they have to interact with humans in a basic sense as they are physically collocated. Their environment can change dynamically by human behavior, introducing uncertainty regarding the robot's plan execution requiring robust and adaptive behaviors.

Although its complexity increased, technology is often used like simple tools to address one specific use case. We initiate a process (e.g. vacuum cleaning) and hold a clear expectation of its outcome. Failures are usually understandable, as the task is clear and transparent.

However, the simple interaction structure of demanding and executing actions implies limits on the efficiency and achievable support. The tool-like interaction mode especially becomes inadequate for complex support systems that do not only execute orders but further influence human perception and planning on different abstraction levels. Examples are complex driver assistance systems for partly autonomous driving (e.g. automated lane change), autopilot functions in aviation, or cooperative robots, where the responsibility is shared between both agents. In such situations, the human might not always understand what the system is doing, inducing interaction problems of transparency, misunderstandings, and conflicts. For efficient cooperation, it is necessary to extend good task functionality by considering the perspective of a human partner to allow flexible interaction modes [103].

Approaching such a human centric direction, first applications take the human behavior into account (also called “human in the loop”), as e.g. in fatigue detection systems in automotive. However, these are limited to narrow use cases and do not account for interaction effects nor human perspectives.

For more sophisticated interaction of robots or other intelligent technical systems, the representation of current task needs to be extended with an understanding of the human partner, like a theory of mind that humans develop of others cognitive states, representing a key component of inter human interactions [57]. Theory of mind is the capability to explain the behavior of others by the inference of mental states, such as beliefs, desires, and intentions [119]. It is a central capability for human interaction, co-operation, and communication [40]. Since mental states of others are not accessible, theory of mind is based on interpretation of and conclusions from observable behavior. Based on a perception model (e.g. respecting the differing perspectives) and a model of decision making, one can infer latent human cognitive states that could have caused the observed actions or information gathering. When interacting with others, theory of mind allows to detect common goals as well as others' false beliefs which can be addressed in cooperative interaction. A similar capability, an artificial theory of mind, will help a robot to flexibly cooperate with and support a human partner. When a robot is aware of what its partner might know, be uncertain about, or miss to know, it could warn her, or share appropriate missing information. However, developing a theory of mind, respectively the inference of others' beliefs, desires, and intentions is computationally challenging, since it requires a second order inference of what others inferred about their environment. Additionally, often multiple mental configurations are valid explanations of the same observed behavior, making the inference problem ambiguous.

Current assistance systems can take over subtasks in narrowly defined conditions (e.g. lane keeping) or instruct the human what to do, independently of her behavior (e.g. navigation system). A human centric assistance approach opens new opportunities with benefits for interaction efficiency, transparency, and support. With an artificial theory of mind, a technical system could respect human situation awareness and specifically support a human partner according to her needs in a transparent way.

This can be achieved by intelligent information sharing, using a theory of mind to tell her about relevant information she might have missed. Combining inference results of human belief regarding her knowledge together with relevance of information for the current task, a robot can support a human partner by communicating related information aspects. State of the art communication concepts may contain a human model, but rather instruct her what to do, either at salient situations (navigation system) or when a deviation from expected behavior is detected (fatigue detection). This might prevent immediate human errors, but will not help her gaining situation awareness and taking future decisions, as it does not account for reasons of human behavior. Instead, when understanding why she made an error, e.g. due to a misunderstanding or false beliefs, a robot could directly address such mental causes by sharing important related information, to not instruct her what to do, but to enable her making good decisions herself.

As basic example, a blind spot assistant in a car implicitly contains and evaluates a human model for driver support (Figure 1.1). It relies on a static perception model that the driver cannot perceive traffic participants located in the car's blind spot area. If the driver nevertheless believes the left lane to be free, she might decide to indicate and initiate lane change. The system's build in understanding of human perception and decision making allows to intervene in such situations and communicate the presence of an obstacle by visual and auditory signals enabling the driver to correct her bad decisions and prevent severe outcomes.

Such warnings or communication interventions should not occur too frequently, since they require human attention, distract, or delay other important tasks which can result in annoyance, decreased performance and acceptance. Instead, decisions, when to interrupt and what type of information to share, should be based on its relevance for the current situation to respect such costs of communication. As long as the driver keeps her lane, the presence of a car in the left is not relevant for her task. This changes as soon as a lane change is intended. Still, a blind spot assistant is designed for one specific use case for which human behavior and relevance evaluation are statically encoded. When assisting a human in more general settings, it will not be possible to account for all occurring situations in advance and define adequate behaviors. Instead, these problems can be addressed by creating an explicit understanding for task and human mental states, a Theory of Mind, to flexibly evaluate human knowledge and needs according to task relevance. The importance of different types of information can then be balanced against costs of communication. Such an approach can be used to create an intelligent information sharing strategy deciding whether, when, and what type of information to communicate to optimally support the human while minimizing distraction and annoyance.

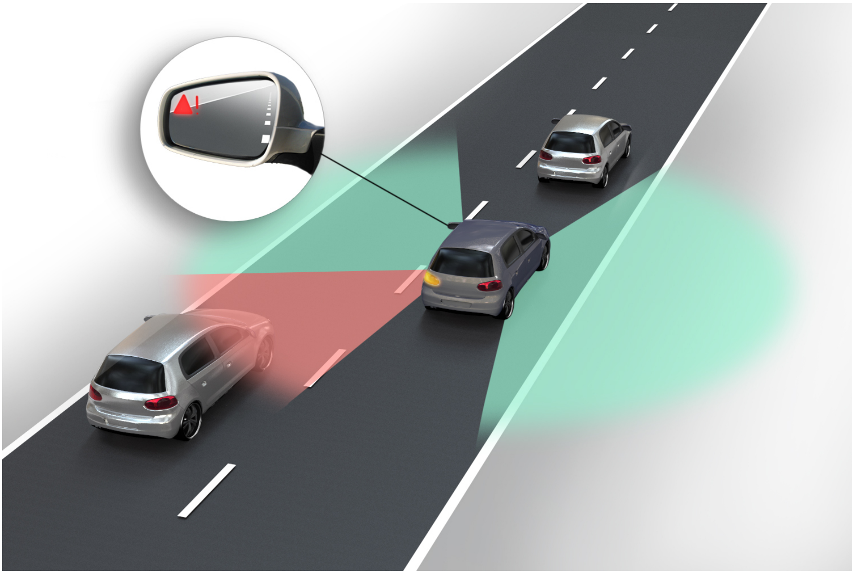

Figure 1.1: Blind spot assistant for a car as example for a basic human centric interaction concept.

Such assistance concepts based on an understanding of human behavior and needs allows for earlier and less intrusive intervention, compared to emergency systems such as emergency braking. Often, it is possible to detect and support problems in human situation awareness before a situation becomes dangerous. In this early stage, information sharing can help the human regaining awareness of the situation and solve it herself. Consequently, sharing important information that she probably missed, can efficiently support her, while she stays in control, instead of ignoring and taking over or instructing the human what to do. At the same time, technological failures are less critical, as the decision is still made by the human. Besides performance improvements, such interactive and transparent behavior – communicating information only when necessary to support her awareness – may lead to a high level of acceptance.

The general idea of a human centric information sharing concept proposed in this thesis is illustrated in a schematic example in Figure 1.2, in which a robot supports a human to mount a shelf. The human is not aware about the current plank being reversed sided (respectively that the plank is not symmetric). By observing her information gathering (the human does not look at plank sides) and actions (the human prepare for mounting the plank), the robot can infer a false human belief of the plank's side. It can evaluate the relevance by evaluating possible consequences or outcomes (need for partial disassemble or aesthetic discrepancy), comparing benefits and costs of communication. Consequently, it decides when and what type of information it communicates to the human.

The concept is motivated to use the rising capabilities of modern technical systems by sophisticated interaction modalities to employ efficient cooperative interaction. Human centric communication will allow for support, that does not instruct but provide necessary information and enable well-founded human decision making.

1.1 Overview and contributions

To recognize and understand occurring challenges when intelligent agents interact with humans, concepts from human factors and human human interaction are presented in section 2.1. These are complemented by methods to cope with uncertainty in inference and decision making, which arise from complex tasks, uncertain dynamics, and other agents. Regarding understanding and communication with other agents, human model and communication policies are discussed afterwards. The methods are relevant for understanding humans and plan communication behavior and different approaches are discussed afterwards.

The first challenge towards human centric interaction consists in the development of an adequate human understanding, an artificial theory of mind. Available information of human perception and decision making need to be combined to estimate latent human states online during interaction. In chapter 3, a new approach of handling this problem of second order inference, inferring what the human inferred about her environment, is presented. The inferred human belief can be evaluated using a decision model providing a new quantitative method to evaluate her situation awareness (section 3.4). In difference to classical situation awareness measurement approaches (see e.g. [32]), it can be evaluated online in a non-intrusive way, without a need for interruptions, post trial ratings, or additional expert evaluations.

The second main question consist in how to use an understanding of human behavior to support her and create human centric interaction strategies. A new concept of sharing information that is relevant and unknown to the human is presented in chapter 4. It formalizes the general problem of when and what type of information to communicate to support a human partner while minimizing interruptions. Consequently, a robot policy is developed which plans with expected benefits, uncertainties, and costs of communication.

Humans can show a large variety of different behaviors requiring robust interaction strategies. For tests and evaluation, it is important to consider real interaction in user studies. Consequently, in chapter 5, task design and results of two user studies are presented to investigate human understanding and communication planning and its effects on performance and acceptance.

Figure 1.2: Illustrating the structure of human centric assistance and information sharing. Robot infers what human partner inferred over state and evaluates future consequences. In this situation, she might not be aware of the plank being reversed side front. Hinting about it might help.

2 Interacting with humans in uncertain environments

2.1 Human factors and human interaction

Human robot cooperation is part of the field of human technology interaction, subject of consideration in different research fields, from human factors and interface design, psychology, and cognitive science. Interaction concepts and cognitive models become relevant for the design of robots since its manipulative capabilities start to allow for interactive use cases and the setup of human robot teams. Extending the experience of human machine interaction, also theories from human human interaction provide interesting insights and inspirations for the design of complex and efficient interaction modalities. In the following, concepts from human factors in technology interaction and inter human interaction are presented, which provide theoretical foundations for an intelligent robotic assistant.

2.1.1 Human factors and technology

Human factors research analyzes human technology interaction from a human perspective [20]. It considers interaction and interface design as well as new challenges that arise from interaction, transparency, and a change of the human role in a human automation team. There is long research history of human technology interaction in complex environments with applications such as aircraft operation or plant supervision. In these domains, a human operator is supported by assistance systems that are able to take over different sub tasks with differing levels of complexity and abstraction. Optimally, these systems are designed to support human situation awareness, by reducing her workload and presenting relevant information, and the human should build an adequate level of trust.

Situation awareness

Interactive support of a human will especially be helpful in complex dynamic settings with high uncertainty, such as complicated driving scenarios, operating an aircraft in special situations, or emergency rescue coordination. These situations are characterized by a large amount of information, that need to be prioritized and perceived by a human operator (e.g. objects and other traffic participants). She further needs to relate the pieces of information to each other to evaluate the current situation and potential future evolution (detect possible conflicts or collision paths). These steps are considered as gaining awareness of the current situation which is required for good planning and decision making [31].

The concept of Situation Awareness (SA) is often used to evaluate and train human performance in such demanding scenarios [31], [1]. It describes cognitive processes involved in developing a situation understanding as basis for informed decision making. It consists of information perception, bringing them in relation to each other and anticipating future evolution, which are known as the three levels of situation awareness [31]. Short decision cycles and multi-tasking can lead to a high cognitive and perceptual load, making it necessary to pre-select possible situation relevant sources from which information is gathered and processed (e.g. concentrating on traffic participants with potentially conflicting paths). Especially for novice users it is challenging to develop and maintain situation awareness in such scenarios, and it can be necessary to train adequate strategies and procedures beforehand.

To support a human operator in these domains, automation functions were introduced. These can take over subtasks to free the human operator's resources for other activities. However, these automated technical systems do not only help, but also bring in new complexity into the whole system. The human needs to know the interface, how to activate and configure automation. Further, she needs to be aware of state of operation, capabilities and limits. Otherwise, there might evolve a problem of under relying or missing trust, where the operator does not use the system or supervises it in detail which prevents the intended cognitive savings. Or the opposite problem of over reliance might occur, where she totally relies on the system without considering its limits or reliability [46]. Further, unclear role attributions and interdependence of agents provide dangers such as a fight for control, when human and automated system work against each other (as e.g. happened in the Boeing 737 Max 8 accidents, where a faulty autopilot regularly overwrote the pilots' control inputs who neither achieved to disable the function [51]).

In such interactive and interdependent settings, it is important to consider the whole system including interaction processes and information exchange, rather than only concentrating on the human and her situation awareness in isolation. If a relevant type of information is not accessible, this limited interface prevents her from gaining awareness and appropriate decisions. Consequently, the concept of distributed situation awareness takes the whole system and interactions between agents into account [98]. Situation awareness is seen as property of the system rather than that of individuals and blaming human operators for failure generally falls too short. Rather, it is important to analyze interaction and information exchange between humans and automated functions, which often plays an important role [98]. For a system to hold situation awareness, it is crucial that relevant information is shared between the agents or entities that are involved in the current perception and decision making processes. By analyzing information flow graphs, weaknesses and central elements of the system can be detected and improved [98].

Since situation awareness depends on cognitive processes involved in forming a situation understanding, it is difficult to measure it (for an individual's awareness as well as from a system's perspective). Despite its importance for task performance, it is not possible to use task performance as a proxy for SA as even unaware humans can select good actions by chance [32]. Different approaches were developed to measure an operator's situation awareness. The Situation Awareness Global Assement Test uses a simulator for the situation of interest, and simulation is interrupted for testing [32]. While the simulator screens are blanked, the operator is asked questions about situation relevant aspects (e.g. depending on flight situations, this can be own speed or state of other aircrafts). The aspects of relevance respectively the questions for each situation need to be defined in advance by domain experts. Although this method achieves good and objective SA measurements [32], it comes with a high effort for preparing questions and its intrinsic limitation to simulator studies.

Simulation interruptions can be avoided by using subjective retrospective ratings such as the Situation Awareness Rating Technique [111] or expert assessments based on observed behaviors [109]. These SA measurement techniques are used for training and evaluation of different interfaces. However, they are not suited for online analysis and intervention to support a human directly in the situation. Similar to expert assessments an automatic system may analyze observed behaviors to collect evidence regarding cognitive processes related to SA [1]. In this thesis, such a concept is developed and formalized (section 3.4), providing an online evaluation of human situation awareness.

Transparency and delegation

One important factor influencing situation awareness and interaction performance is transparency. It is important to enable the human developing an adequate understanding of capabilities and limits of automation systems she interacts with to develop a good level of trust. This is necessary to decide, which subtasks to do herself and which to delegate to automation and further the degree of detail she needs to supervise the results of automation.

Transparency however does not mean, that the operator should track and supervise every low-level action of the system, which would require similar cognitive engagement and load as if she does the subtask herself [65]. As in a human team, the operator should trust the system to execute the delegated subtask, which in turn might report back the result or problems, respectively abstracted states of progress to acertain transparency on a higher level. Such delegation-based interaction can significantly reduce the operator's cognitive load (which is then free for other tasks), while she stays aware of the system's activity [65]. Missing transparency can lead to the mentioned problems of over reliance or distrust and potentially severe outcomes [46]. The need for transparency leads to the question, which information (regarding environment perception, task progress, or possible problems) are relevant for a human operator and should be communicated, which is important to achieve distributed situation awareness. It can be further useful to communicate the degree of uncertainty of the system, especially when reliability may change between situations [92], [52].

Human role

Besides the need for transparent and efficient interfaces and communication processes, the rising performance of autonomous systems will also change role interpretations and responsibility attribution. Currently, a human operator usually is responsible for task achievement with options of using technology like tools, delegate subtasks (lane keeping) or receiving warnings (low attention warning). With the promise of highly autonomous systems such as automated driving, the system may overtake an operating role including responsibility. It will further need to interact cooperatively with other traffic participants where no strict role assignment is employed, requiring coordination and flexible negotiation of initiative.

Although technology has shown enormous progress, it is still far from full autonomy in many application settings (while in others, full autonomy is never desirable [103]). Consequently, it seems reasonable, that robot and human capabilities are used to complement each other, which should benefit cooperative operation. Capabilities typically differ, as human reasoning is very flexible and allows for highly abstracted reasoning processes. She is good at coping with unknown situations and finding intuitive and creative solutions. On the other hand, a robot or automated system may show benefits regarding processing and memorizing multiple information sources, search, and precision in computation, planning and execution. When the capabilities of agents are known or can be estimated, these can be used to optimize for joint performance by exploiting individual agents' strengths [118]. In general, an adequate task and role distribution will depend on application domain and target. This can range from a supervisory role to flexible, situation-based role assignments.

Performance, trust, and alignment of values

A crucial aspect for trust and the relation of humans and automated systems is the alignment of values, respectively if both entities share the same goals. When optimizing for joint task performance in a common task, it is important how this is specified and understood by the agents to avoid misunderstandings. Specifying a target function of an autonomous agent by hand can be hard and may lead to unintended effects, known as the value alignment problem [39].

Transparent operation and action execution is important that a human can trust an artificial agent's capabilities. However, for cooperative operation it is further relevant that she trusts its intention, respectively that the systems' goal is to support her [53]. Instead, when goals are not fully aligned, an agent or robot could exploit human trust and let the human contribute to its own targets [60]. Consequently, in cooperative settings, it can be important to explain robot's (controversial) decisions to a human to avoid confusions and improve acceptance and trust.

Considering the value alignment problem from a more general perspective, it might not always be desirable to optimize for objective criteria such as task success rate or speed, as this might not correlate with subjective human satisfaction. Alternative objectives include empowerment, where a robot aims to support a human by providing her the ability to achieve any goal, even if it is unknown to the robot [30]. Other approaches directly address human psychological needs, that a robot agent could support, which are subsumed by the concept of user experience [41].

2.1.2 Human–human interaction

Due to the high complexity of autonomous technical systems like robots, classical role understanding and interface design are not sufficient any more, imposing problems of transparency, trust, and coordination, as discussed in the previous section. Instead, technology use should be considered from an interaction perspective, for which human human cooperation may provide useful inspirations. Humans are able of complex forms of interaction as they build effective teams and large societies. For efficient interaction with partners, it is necessary to understand them, as well as to behave transparently to be understandable [26], [19]. Humans develop a theory of mind, an understanding of others cognitive states. This is necessary to evaluate the situation, including relevance of external and internal information for a cooperation partner (e.g. making own behavior more transparent by communication) and to coordinate behaviors according to a joint strategy.

Theory of mind

A theory of mind describes a representation of others' mental states as beliefs, intents, or desires. It is an important cognitive capability, as it allows to understand and predict others' behaviors, base interaction on these expected mental states, as to initiate and understand communication [119], [57].

A theory of mind is based on observed behaviors of others, that are rationalized, meaning that hypotheses about cognitive configurations (containing intentions and beliefs) are rated regarding how well they explain the observed behaviors. Observation may include information gathering such as eye gaze as well as movements or manipulation actions.

To understand and interpret observed behaviors it is important to respect the perspective of othersIn the literal sense, perspective taking is important to account for what is occluded from others' perspectives respectively what action opportunities are available for others in their current situation. Further, it is important to account for differing mental configurations, e.g. others' goals, priorities, or emotional states, to understand behaviors and developing empathy. When holding and tracking a mental representation of others, it is possible to attribute false beliefs, that is to detect that others' belief deviates from a known environmental state. This discrepancy in mental representations also affect the own reasoning and can slow down responsiveness [35].

In interactive situations, others' mental states may include their representation of oneself. Consequently, inferring their mental states adds a hierarchy level of reasoning (of what I believe others believe I believe) which can be continued recursively, yielding higher orders of theory of mind. Higher levels of theory of mind are important for strategic interaction (e.g. what do others want me to believe), in competitive settings e.g. poker. But also in cooperative situations it is relevant, where knowing intentions of others helps us in understanding their utterances and even supports general language understandings of small talk and irony [40].

Although humans are capable in creating a sophisticated theory of mind, it requires cognitive processing capacities. When other tasks need to be done (that are not interactive, respectively not related to ToM), quality of mental inferences is reduced [4]. The correlation of task solving and theory of mind is not restricted to explicit reasoning, as also implicit processes of theory of mind affect our cognitive performance, (and a pure presence of a person with false or uncertain belief can slow down a response) [35].

A theory of mind is not achieved by exact emulation of cognitive processes nor does it contain a complete set of others' beliefs. It rather focuses on abstracted, relevant aspects with an important influence in explaining observed behaviors and interactions, while neglecting commonly known or irrelevant aspects [48]. This is important for efficiency as it limits hypothesis space, but may also lead to approximation errors, (e.g. when neglecting impairments of blind people, as it easily happens when being stressed).

Communication, information sharing, and relevance

From a functional perspective, communication has the purpose of changing others' mental states, especially their beliefs [82]. This makes theory of mind an important capability for communication, as it is important to consider, what they already know or what they falsely believe, as it avoids unnecessary, and inefficient repetitions. When deciding what information to provide, even young children do not only focus on external relevance (how interesting or salient something is) but also account for novelty of information for the receiver [82]. Theory of mind is important when deciding for communication actions, but also for understanding communicated utterances, as communication depends on intentions and a common frame of reference (intent behind communication also lets us concentrate on utterances in contrast to background noise [40]).

Efficiency is a factor in communication, because it requires efforts in cognitive processing of both, sender and receiver, which can delay or distract from other tasks. These costs of communications need to be considered when taking decisions on when or what to communicate.

It is a basic principle in relevance theory, defining relevance of information as combination of positive cognitive effects and costs required to transmit information [106]. Consequently, communication decisions are made and interpreted according to the relevance of contained information. In chapter 4, relevance of communication is quantitatively formalized for a human robot communication setting.

Explicit communication is at first a cooperative action to provide useful information, as otherwise the receiver can simply ignore it [47]. Communication can contain external information with the purpose of information sharing as well as internal mental states such as intentions or proposed plans to coordinate agents' behaviors.

Communication has many facets, can happen via rich natural language but also implicitly via demonstrative actions. For successful communication, it is important, that there is a common ground respectively a frame of reference of sender and receiver [25], [35]. This can include the communication protocol (e.g. the language) but also common knowledge, context, and conventions. When a robot wants to efficiently communicate with a human, it is necessary to consider these aspects to adapt to and understand human communication modes. This allows both to create a common frame of communication for efficient information exchange.

Coordination and acting together

Besides sharing external information, communication can enable the coordination human behaviors. Coordination is necessary in different situations and facets. Often, there are multiple ways to solve a task, but each way will only work, if all agents agree on choosing it. Coordination ranges from synchronization (e.g. coordinated pedaling at tandem bicycles, or speed synchronization for platooning) to higher level goal negotiations or role distributions in teams. Other examples include collision avoidance e.g. for pedestrians regarding the decision on which side to pass, how to divide work or which task to work on first.

Coordination is also related to initiative taking. Many human robot settings consider a fixed initiative setting, where either the human takes initiative and the robot adapts to her or the robot takes initiative while literally ignoring human actions with the premise that she will adapt. In (more natural and flexible) mixed initiative settings instead, it can be necessary to coordinate both behaviors, e.g. to agree on a common plan.

The process of achieving coordination differs depending on time constraints and level of abstraction. In some situations, such as pedestrians coordinating to pass left or right, there is time to observe other's behavior and react multiple times until convergence to a coordinated solution. Such agreements can be achieved via randomizing low-level actions or negotiation of higher level sub goals and goals.

In contrast, in time critical situations, or if others' behaviors cannot be observed, the problem can be represented by a one step coordination game [76]. Other mechanisms are required to solve coordination, such as virtual bargaining [66], common knowledge reasoning [112] or by introducing a communication channel (which changes the problem, more in the direction of multi step coordination, or bargaining). Coordination then can be achieved by proposing and negotiating equilibria solutions of the coordination problem at hand [26].

Solving human robot coordination problems goes beyond the scope of this thesis. Still, the developed formulation for belief inference and information sharing are useful and can be similarly applied when coordination states are introduced in task description, as e.g. proposed in [13]. Such theory of mind with coordination states will enable convergence of behaviors, to solve human robot coordination problems.

2.2 Coping with uncertainty

In a complex and dynamically changing world, humans and robots permanently have to cope with uncertainty. They need to gather relevant information of the environment and generate robust plans for their future behavior.

In this chapter, different probabilistic methods are introduced, regarding inference of uncertain states and the generation of policies to effectively behave in uncertain environments. These methods serve two purposes, modeling and understanding a human's behavior in uncertain domains as well as to plan interaction with her, where she represents a new source of uncertainty. The goal of modeling does not consist in reproducing human reasoning exactly, but rather to obtain an abstracted understanding and explanation of her behavior.

Uncertainty might be introduced by different factors, affecting the initial state of the environment and stochastic state transitions [49]. Regarding current state, limits in perception can restrict information gathering as e.g. visual occlusions or measurement noise. Even when every piece of information could be gathered, there might be uncertainty due to perceptual overload, meaning that there is not enough time to perceive and process all available information until decisions need to be made, which is e.g. considered in complex situations within the concept of situation awareness. New uncertainty can be introduced by state transition. Not every state aspect is controllable by an agent's action, but may depend on stochastic events or external influences. Consequently, an agent will have to track the actual state changes and react accordingly. When interacting with other agents, their behaviors are not known beforehand which introduces more uncertainty as they will also change the environment. Especially humans can show diverse and stochastic behaviors which are difficult to predict precisely.

2.2.1 Bayesian inference in probabilistic models

Bayesian inference is a method to quantify uncertainty in the form of probability distributions, based on available information respectively observations. The method is shortly introduced and considered for the temporal filtering problem. For a detailed introduction and formalization of Bayesian inference, graphical probabilistic models, and approximation approaches, the reader is referred to the book of Bishop [10].

Bayesian inference considers uncertain variables of interest, which can be continuous or discrete, such as unknown system states s or process parameters. Despite being unknown, there might be assumptions regarding their values, which can be expressed as prior probability distribution $p(s)$.

For example, in the blind spot scenario from chapter 1, there is uncertainty about the state of the left lane, where another vehicle could be, which can be represented as binary state, $s \in \{\text{free}, \text{occupied}\}$ and a prior probability $p(s = \text{free}) = p_0$.

To refine the prior guess, a human can gather information (e.g. collecting visual stimuli) called observation o that depend on the underlying state s. However, due to uncertainty in perception (the left lane is in the blind spot of the driver), she might not receive true information about the state, perception is not reliable. The perception model considers the generation of an observation, given state s, formalized as probability distribution $p(o \mid s)$.

Using the evidence from the perceived observation o under the perception model $p(o \mid s)$, the posterior distribution is calculated according to Bayes theorem,

$$p(s|o) = \frac{p(s)p(o|s)}{p(o)}. \tag{2.1}$$

Since o is directly observed and exactly known, $p(o \mid s)$ represents the likelihood of observing o in state s, a function of uncertain state s. The marginal likelihood $p(o) = \sum_{s'} p(o \mid s')$ is independent of the current state s. It normalizes the product of prior and likelihood to yield a proper posterior distribution.

Through this structure, Bayesian inference can combine system knowledge expressed as prior distribution with collected data respectively experience to achieve better probabilistic models of the environment. It is a rich tool for many different problems and domains, from filtering, state estimation and sensor fusion to classification and regression problems.

Bayes theorem eq. (2.1) describes Bayesian inference in a very general form. Often, an inference problem can be further structured by considering causal dependencies of involved variables, as not all variables depend on all others. The structure can be illustrate as directed graphical model that represents the causal interrelations of related variables. In the temporal filter problem (considered in the next paragraph), each observation only depends on current state and each state only on the previous state, as is visualized in Figure 2.1.

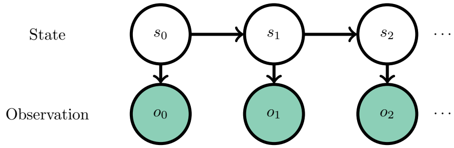

Figure 2.1: Directed graphical model for the filtering problem. Nodes represent different variables (e.g. state and observations at different time steps), observed variables are represented by filled nodes. Arrows indicate causal dependence, e.g. each state only depends on the last state and effects current observation and next state.

Filtering

In the blind spot scenario, a human might collect multiple observations at different points in time, where each can be used to adapt her state estimation. The environmental state may change over time, leading to the iterative filtering problem. The agent may receive repeated observations or measurements o_k, one at each discrete time step k that carry information of the current environmental state s_k. In the filtering problem, state s_k is chosen such that it fully describes the environment, meaning that the next state s_{k+1} only depends on current state but not on past states. These dependencies are summarized in the graphical model Figure 2.1.

At each time step, the agent can use the received observation o_k to refine its state estimate $p(s_k)$ via Bayes theorem, which is then used to predict the next state probabilities. Hence Bayesian filtering consists of these two steps, state update and state prediction.

Therefore, models for state transition and generation of observations are needed. For example, a continuous state system is often modeled using Gaussian uncertainty around nominal transition and measurement functions f, h, leading to transition probabilities

$$\begin{aligned} s_{k+1} &= f(s_k) + n_s \text{ with } n_s \sim \mathcal{N}(n_s \mid 0, \Sigma_s) \\ p(s_{k+1} \mid s_k) &= \mathcal{N}(s_{k+1} \mid f(s_k), \Sigma_s), \end{aligned} \tag{2.2}$$

and observation probabilities

$$\begin{aligned} o_k &= h(s_k) + n_o \text{ with } n_o \sim \mathcal{N}(n_o \mid 0, \Sigma_o) \\ p(o_k \mid s_k) &= \mathcal{N}(o_k \mid h(s_k), \Sigma_o) \end{aligned} \tag{2.3}$$

Instead of inferring all hidden states $s_0 \cdots s_k$, one usually focuses on the current state s_k and its conditional probability $p(s_k \mid o_0, \cdots o_k)$ (as it will be the basis for decision making). It is visible from the graphical model, s_k is not directly linked to previous observations, but depend on the last state s_{k-1}, reducing the inference to

$$p(s_k \mid o_0, \cdots, o_k) = p(s_k \mid o_k, s_{k-1}) p(s_{k-1} \mid o_0, \cdots, o_{k-1}).$$

The second term represents the inference result from the last time step leading to an iterative inference approach. Starting with prior belief $p(s_0)$, the update step is

$$p(s_k \mid o_0, \cdots, o_k) = \overbrace{\frac{p(o_k \mid s_k)}{p(o_k)}}^{\text{observation model}} p(s_k \mid o_0, \cdots, o_{k-1}) \tag{2.4}$$

and prediction step

$$p(s_{k+1} \mid o_0, \cdots, o_k) = \sum_{s_k} \underbrace{p(s_{k+1} \mid s_k)}_{\text{transition model}} p(s_k \mid o_0, \cdots, o_k). \tag{2.5}$$

For continuous variables, the sum is replaced by integration over possible last states s_k.

Filtering uses all information available in a current situation, which might be used for decision making. From a retrospective view, also future observations can be respected in inference yielding a more accurate estimate of past states, $p(s_0, \cdots s_T \mid o_0, \cdots o_T)$. This state smoothing can be achieved by a second reverse iteration starting with the last state estimation of the filtering process.

Approximate Inference

Exact Bayesian inference is possible in finite, discrete state spaces as well as for linear models combined with Gaussian uncertainties, as is the case for the linear Kalman filter. For continuous states subject to nonlinear dynamics or non Gaussian uncertainties, exact inference becomes intractable as it requires the integration for marginalization in eq. (2.5) and normalization in eq. (2.4). Instead, approximations are necessary to derive useful estimates [10].

Many approximation schemes have been proposed in the literature, which either approximate nonlinear functions as in the Extended Kalman filter, or probability distributions (UKF, variational inference, sampling methods). In the following, linearization based Extended Kalman filter and sampling based particle filter are shortly presented as they are important for the next chapters, as they will be applied respectively provide orientation for belief inference in chapter 3.

Extended Kalman Filter

The Extended Kalman filter (EKF) follows the derivation of the linear Kalman filter while it considers nonlinear functions with Gaussian uncertainties [69]. In each filtering step, the nonlinear transition and observation functions f and h are approximated by a first order Taylor series around the current state expectation. Consequently for each time step, a local linear model is used and the resulting approximate distributions remain Gaussian, as for the linear Kalman filter, and can be reused for the next filter iteration.

The approximate prediction and inference steps of the EKF are illustrated for a one dimensional example in Figure 2.2 for prediction and Figure 2.3 for the observation update. Both steps rely on a single evaluation of the related function and its derivative to construct the linear approximation.

Sampling based inference

As alternative to deterministic approximations of inference, sampling based approaches such as the particle filter can be used as stochastic approximation methods of the distribution [70].

In contrast to the EKF, a particle filter uses the full nonlinear model, while the distribution $p(s_k)$ is approximated by a set of samples or particles. Correspondingly, expectations are approximated with a weighted sum replacing the integration of a continuous distribution. The number of particles is a crucial parameter here, as it trades of approximation quality and computational load.

At the beginning of inference, an initial set of K particles is sampled from the prior distribution, $s^i \sim p(s_0), i \in (1, \cdots K)$, with a corresponding initial weight $w^i = 1$. During filtering, each particle s^i is transformed according to the non linear transition model (e.g. eq. (2.2)) and its likelihood is computed using the observation model, eq. (2.3), to update the particle weight $w^i_{k+1} = p(o \mid s^i_{k+1})w^i_k$.

Expectations are computed according to the weighted particle sum, e.g. state expectation is evaluated to $E[s] = \frac{\sum_i s^i w^i}{\sum_i w^i}$. During the filtering

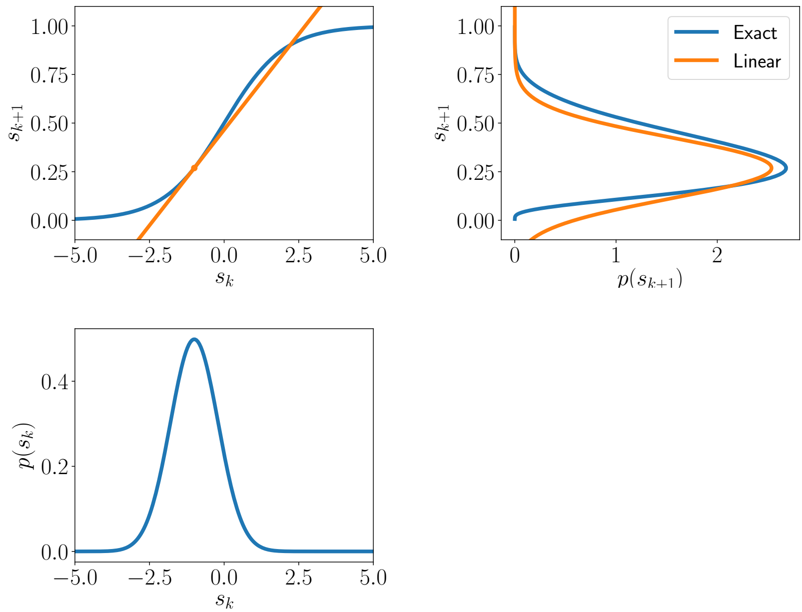

Figure 2.2: Illustration of prediction step of extended Kalman filter. Nonlinear, sigmoid transition function is linearized around last expectation (orange, top left). Starting with a Gaussian distribution $p(s_k)$ (bottom left), it is predicted to yield $p(s_{k+1})$ (top right) according to nonlinear function (blue) respectively linear approximation (red). As long as the last covariance is small compared to the nonlinearity (narrow distribution), the approximation quality is good.

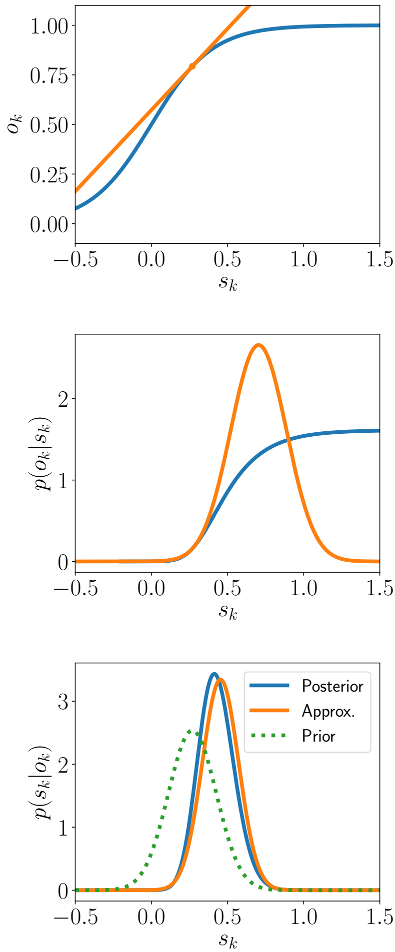

Figure 2.3: Illustration of update step of extended Kalman filter. Top: Nonlinear observation function (sigmoid, blue) is linearized around the mean (orange). This leads to a Gaussian approximation of likelihood function (middle). The resulting posterior Gaussian is shown together with the true posterior (bottom) and the prior (dotted).

process, the weights of some of the particles will decrease over time approaching zero. Since particles with low weights do not contribute to the state estimation, this depletion effect reduces the effective sample size. One solution is provided by resampling, where new particles are sampled according to particle weights when particle weights become too small.

Since there is no restriction to the probability distribution, a particle filter is able to cope with multi modal distributions, making it more flexible than the EKF. However it requires more function evaluations (equal to the number of particles), leading to higher computational load.

2.2.2 Acting under uncertainty

Bayesian inference and filtering provide methods to interpret observations and understand processes and corresponding uncertainties of the environment. However, robots and humans are not passive entities or observers, but can choose actions to change the uncertain environment towards reaching a goal. The decision making problem can be formalized as Markov Decision Process, respectively Partially Observable Markov Decision Process, which provide a basis for reinforcement learning and planning under uncertainty. This section provides a short introduction of the representations and methods which will be needed in later chapters. For a detailed consideration of reinforcement learning in stochastic environments, the reader is referred to the book of Sutton and Barto [110]. Regarding partial observability and perception uncertainties, the work of Kaelbling et al. [50] provides a good introduction.

Markov Decision Process

A Markov Decision Process (MDP) defines the problem of an agent repeatedly interacting with an stochastic environment to optimize long term rewards. An MDP is defined as fivetuple (S, A, T, R, γ), of state space S, action space A, stochastic transition function $T : S \times A \times S \to [0,1], T(s,a,s') = p(s' \mid s, a)$, reward function $R : S \times A \times S \to \mathbb{R}$ and a discount factor $\gamma \in (0,1]$. At each discrete time step k (with the current environmental state s_k), the agent needs to choose an action $a \in A$ with the target to maximize expected, accumulated reward,

$$J(s_k) = E_{s_i}\left[\sum_{\tau=0}^{T} \gamma^{\tau} R(s_{k+\tau}, a_{k+\tau}, s_{k+\tau+1})\right] \tag{2.6}$$

An action a effects the state according to the transition function T, influencing future rewards. State transition is uncertain and it is therefore not possible to plan one action sequence in advance and execute the plan. Instead, depending on the actual transition and new state s_{k+1}, it will be necessary to adapt or replan the sequence. An important aspect of an MDP is the Markov property, meaning that the transition to the next state s_{k+1} only depends on current state and action, but not on the history of states or actions before, $p(s_{k+1}|s_{0,\cdots k}) = p(s_{k+1}|s_k) = T(s_k, a_k)$. The behavior of an MDP agent can be formalized as policy $\pi : S \times A \to [0, 1], \pi(s_k, a_k) = p(a_k|s_k)$, that assigns a probability of taking action a_k in state s_k.

The resulting behavior of an agent strongly depends on the reward function. Depending on the domain, it can be difficult to specify a good reward function that reflects the designer's expectations [39]. Besides its central influence on the resulting optimal behavior, the reward function also affects the complexity of problem solving, respectively learning speed in reinforcement learning.

For domains, where it is difficult to specify a reward function by hand, Inverse Reinforcement Learning (IRL) can be an alternative [72]. In IRL, the inverse problem is considered of finding a reward function that explains an observed optimal policy in an MDP. Thus it is possible to demonstrate trajectories of desired behavior and generate a reward function with IRL. This can be used to define an MDP for the robot, employing the expected behavior, while it typically provides more flexibility than direct imitation learning.

IRL is also used for understanding human behavior from a rational perspective [6], [94]. Assuming that her actions follow a good policy in an MDP, the observed behavior can be used to estimate an underlying reward function. It is not necessary, that the human herself is aware of an explicit reward function, respectively a weighting of different goals. This is used for apprenticeship learning, where a robot solves a similar task as demonstrated by a human expert [72].

Reinforcement Learning

Reinforcement learning describes the process of learning an optimal policy to solve an MDP. By interacting with its environment, observing state transitions and rewards, a policy can be improved iteratively. If an environment model (the MDP) is known or estimated from experience, it can be used to directly optimize for the accumulated expected reward, eq.

(2.6) via dynamic programming [110].

Otherwise, the agent needs to explore its environment to find desirable states and refine its policy. This requires a trade off between exploration, to learn about new state transitions and rewards, and exploitation, optimizing rewards by following the best known policy. Therefore, it is useful to choose a stochastic policy, where the probability of expected suboptimal actions (exploration) is relatively high at the beginning and reduced during learning process (to favor exploitation).

Given a policy π of the agent, a corresponding value function V_π is defined as mapping from state s_k to the expected accumulated future reward,

$$V_\pi(s_k) = E_{s_i,a_i}\left[\sum_\tau \gamma^\tau R(s_{k+\tau}, a_{k+\tau}, s_{k+\tau+1})\right]$$

With the value of the expected next state $V_pi(s_{k+1})$, the value function can be formulated recursively as

$$\begin{aligned} V_\pi(s_k) &= E_{s_{k+1},a_k}\left[R(s_k, a_k, s_{k+1}) + \gamma V_\pi(s_{k+1})\right] \\ &= \sum_{s_{k+1}} \sum_{a_k} \pi(s_k, a_k) T(s_k, a_k, s_{k+1}) \left(R(s_k, a_k, s_{k+1}) + \gamma V_\pi(s_{k+1})\right). \end{aligned}$$

While the reward function provides feedback only regarding current state or transition, the value function considers future and long term effects of actions.

When the value function is available, the policy π can be improved by choosing actions according to the highest value. The alternating process of evaluating value function of the current policy and optimizing the policy is known as policy iteration yielding the optimal value function V^* and policy π^* [110]. The optimal value function fulfills the Bellman equation [110]

$$V^*(s_k) = \sum_{s_{k+1}} \max_{a_k} T(s_k, a_k, s_{k+1}) \left(R(s_k, a_k, s_{k+1}) + \gamma V^*(s_{k+1})\right)$$

As alternative to policy iteration, value iteration directly evaluates the Bellman equation to iteratively compute the optimal value function (without explicit consideration of policies).

Besides assigning a value to each state, it is often useful to also consider the influence of an action. Consequently, the action value function maps current state and action to an expected future reward. The optimal action value function $Q^*(s,a)$ is defined as

$$
\begin{aligned}
Q^*(s_k, a_k) &= \sum_{s_{k+1}} R(s_k, a_k, s_{k+1}) + \gamma V^*(s_{k+1}) \\
&= \sum_{s_{k+1}} R(s_k, a_k, s_{k+1}) + \gamma \max_{a_{k+1}} Q^*(s_{k+1}, a_{k+1}).
\end{aligned}
$$

For medium size discrete state MDPs, optimal value and action value functions can be computed and represented as tensors. For large discrete or continuous state spaces however, approximations are required, as not every state can be fully evaluated. Powerful approximations can be obtained by introducing artificial neural networks for value and action value function, which are optimized by Reinforcement learning [61].

Partially Observable Markov Decision Process

An MDP models the decision problem of an agent interacting with an environment with stochastic state transitions. Although the transition is uncertain, the agent has direct access to current state and reward signal. In complex environments, it is often not possible to gather all information sources and to keep all state information in mind. If the state itself is uncertain, respectively not fully accesible, the problem can be formulated as Partially Observable Markov Decision Process (POMDP) [50]. Instead of a direct access to the current state s_k as in MDPs, an agent in a POMDP only receives a state dependent observation o_k according to an observation function $O(s_k, o_k) = p(o_k \mid s_k)$. The stochastic observation reveals partial information about the true underlying state. The structure of an agent interacting in a POMDP is shown in Figure 2.4. In each time step, the agent can select an action a_k as in the MDP, influencing the state transition. It receives the observation o_k generated according to the observation function together with the current reward r_k. A POMDP is formally defined as 7 tuple $(S, A, T, \Omega, O, R, \gamma)$, where state set, action set, transition function, reward function, and discount factor, S, A, T, R, γ, are defined as for the MDP. These are extended by an observation set Ω and the observation function $O : S \times \Omega \to [0,1], O(s_k, o_k) = p(o_k \mid s_k)$.

A POMDP provides a flexible framework to model uncertain and complex situations. A POMDP agent can develop interesting behaviors including information gathering, respectively active perception [96]. It means, that the agent is choosing specific actions that do not progress the task

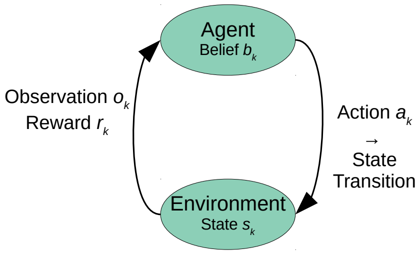

Figure 2.4: Structure of a POMDP. An agent interacts with its environment. Via its actions, it can influence the environmental state transition. Instead full access to state s_k, it only receives partial information via the observation o_k together with the scalar reward signal r_k.

itself, but provide her more information regarding environmental state. With this additionally gathered information, it can make better decisions in the future. Examples for everyday information gathering actions are asking others for information or reading a manual. Typically, the agent needs to trade off information gathering (which is typically costly, as it requires time or effort) and task progress using the knowledge obtained so far.

State transitions, like in an MDP, only depend on the current state and action and not on previous states. However, the Markov property does not hold for the available observations. When planning with state transitions and potential observations, the agent needs to respect the history of past observations and actions. Therefore, it will need some form of memory. Instead of storing the action and observation history, it can explicitly represent an uncertain estimate of underlying state, a belief, or use an implicit representation as realized for example in recursive neural networks [67].

Agent belief

An agent in a POMDP does not know the true environmental state but does only receive partial information via observation o_k. Based on the past observations and its own actions, it can infer a probability of current state s_k via Bayesian inference, respectively Bayesian filtering (section 2.2.1), to construct its belief $b_k = p(s_k)$. Since the state itself fulfills the Markov property, the belief contains all relevant information of the history.

In each time step, it needs to update the last belief estimate by predict-

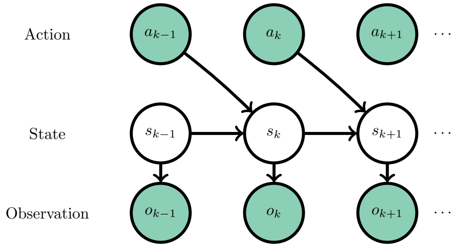

Figure 2.5: Directed graphical model for inference in a POMDP (nodes of observed variables are colored). With observation o_k and agent action a_{k-1}, it can update its belief over hidden environmental state $s_{k-1} \to s_k$. The current state only depends on last state and last action.

ing state transition and updating with the observation. In addition to the filtering structure (Figure 2.1) the state transition respectively prediction is influenced by the known action a_k, as represented in the graphical model in Figure 2.5. Consequently, the belief update is computed according to

$$b_{k+1} = p(s_{k+1} \mid a_k, o_{k+1}) = \underbrace{\frac{p(o_{k+1} \mid s_{k+1})}{p(o_{k+1})}}_{\substack{\text{information gain /} \\ \text{filter update}}} \underbrace{\sum_{s_k \in S} T(s_k, a_k, s_{k+1}) p(s_k)}_{\substack{\text{state transition/} \\ \text{prediction}}} \tag{2.7}$$

Starting with prior belief $b_0 = p(s_0)$, the agent can maintain a belief over current state s_k by iteratively updating the distribution of current state.

Solving a POMDP is much harder than solving the underlying MDP due to the state uncertainty, that is expressed in an increasing history of observations, respectively a probability distribution of state. A simple approach to act in a POMDP is to apply an optimal MDP policy of the base MDP by considering only one state from the belief distribution (maximum or expectation). However, this approach ignores the state uncertainty and expected information gains and therefore can not yield information gathering policies [93], [96].

Based on the agent belief b, a POMDP can also be interpreted as an equivalent, specially structured MDP over continuous belief space, with transitions according to belief updates (eq. (2.7)) and expected rewards. However, the state space of a belief MDP is continuous and increases exponentially in the number of discrete states of the POMDP, which usually renders the application of MDP solution methods infeasible.

Solving POMDPs

One approach for solving POMDPs consists in value iteration by iteratively computing a value function over beliefs, $V(b)$ [50]. However, it is only tractable for small state spaces or short horizons, as the complexity increases exponentially in the number of states and observations. This exponential increase is also described as curse of dimensionality and curse of history [104]. As approximate method, point based value iteration restricts value updates of value iteration to a limited subset of the continious belief space [83].

As alternative to such offline methods that aim to compute a value function for the full belief space in advance, there are online approaches that search for a good action in each time step by planning into the future [93]. They can handle large belief spaces, as they search locally for a good action. In a concrete situation, only a small subset of possible beliefs will be relevant, reducing the necessary evaluations [93]. As a drawback, each time step requires replanning and the computation of the local action values, which challenges real time applications. Online POMDP algorithms create a planning tree over possible future actions and observations, starting from the current belief, as is visualized in Figure 2.6. The planning tree can be a full tree up to a fixed planning horizon, respectively tree depth, or the actions and observations can be sampled or heuristically selected to focus search on interesting trajectories. Still, the planning tree will expand very fast (with size of action and observation sets, $|A| \cdot |\Omega|$ per planning step) and it will usually not be possible to plan until the end of an episode. Instead, tree expansion needs to be limited and values of leaf nodes are estimated e.g. by using coarse approximations of value iteration or MDP based values. For the base MDP, value functions for optimal policy as well as random uniform policies can be computed and used as rough proxies for the leaf values [93]. For tree evaluation, values are propagated through the tree, providing a good estimate for the current action values and the expected best action can be selected.

Through leaf value estimates, online tree search is combined with approximate offline evaluations. Accuracy and computation effort depend on planning depth and the quality of offline value estimates. This makes it possible to solve much larger POMDPs, as demonstrated by specific implementations such as POMCP [104] or DESPOT [105]. Planning approaches are also more flexible than offline methods, as they can cope with changes in their environmental dynamics by adapting tree generation (while offline methods require a full re-evaluation).

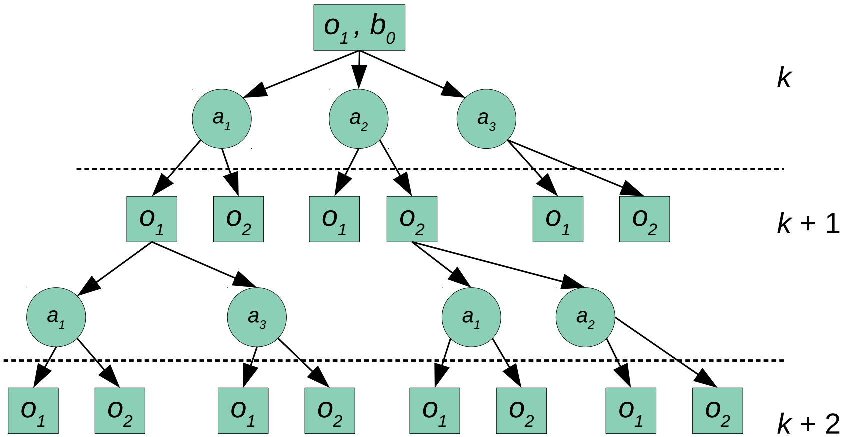

Figure 2.6: Planning tree for POMDP online methods. Starting with current observation and belief, future actions and observation options are evaluated to choose a good next action.

2.2.3 Uncertainty by other agents and strategic interaction

A POMDP provides a formalization of an agent interacting with its dynamic uncertain environment. If multiple agents are present, interaction effects between these become relevant. Besides environmental uncertainties, the behavior of other agents represents a new source of uncertainty with effects on the environmental state. In a simple approach, other agents are often considered as part of the stochastic environment instead of introducing an explicit agent representation. In this case, the problem can be again formulated as a POMDP. However, a transition function including other agents' behaviors does not need to be stationary. This introduces new challenges to solve respectively learn good policies, especially regarding the convergence of policy and value functions.

For the setting of a multi agent team cooperating to reach a joint goal, the concept of decentralized POMDP was proposed to explicitely account for other agents in uncertain domains with limited information exchange [71]. Although any individual agent's policy must rely on local observations, the policy is globally optimized, being part of the joint policy, to reach a common goal. This approach can be used to train a multi agent team for uncertain scenarios, but cannot cope with heterogeneous settings with independent policies, like human robot interaction.

When policies are not synchronized or rewards of agents are not fully aligned, the framework of interactive POMDP provides another extension to the single agent POMDP [36]. It includes a hierarchical recursion of modeling others (agent A believes that B believes that A believes . . .), which allows powerful logical reasoning. It comes at the expense of a high complexity and its limitation to very small domains.

An alternative to potentially infinite recursive reasoning as in interactive POMDPs is provided by the concept of equilibria in game theory. Game theory investigates interaction and interaction effects of rational agents. A Nash equilibrium describes a set of policies, where no single agent can benefit from changing its policy, while the other agents do not. For an introduction to game theory, the reader is referred to the book of Osborne, [76]. Here, two phenomena of strategic interaction are introduced, as they closely relate to human robot cooperation: coordination and signaling as strategic information exchange.

Coordination

When interacting with others, one often faces the problem of coordination in different forms. From a game theoretical perspective, the coordination problem is characterized by the presence of multiple equilibria, which represent multiple ways to solve some task. Accordingly, the agents need to agree to choose one of these, meaning to coordinate their behaviors. A pure coordination problem is introduced with the payoff matrix shown in Table 2.1. There are two Nash equilibria, where the agents choose the same actions and in each of them, both agents receive the same reward of 1. Such symmetric coordination problems occur for example, when pedestrians moving on collision paths coordinate on passing each other on left or right side.

Another coordination problem is present in the stag hunt game, represented by the normal form payoff matrix Table 2.2. The agents can decide to jointly hunt for a stag, an equilibrium with high rewards for both, or individually go for a hare with a smaller reward. Best outcome for both agents is achieved, when they coordinate on hunting the stag (payoff dominance). However, if one agent is uncertain about the other's behavior, it is less risky to choose the hare (risk dominance). Here, the uncertainty of other's behavior is central for the choice of equilibria.

These normal form games investigate one step coordination, which is

Table 2.1: Payoff matrix for a symmetric coordination problem (entries of the matrix are the payoffs respectively rewards of the agents, when the corresponding actions are chosen). There are two Nash equilibria (a_1, a_1) and (a_2, a_2) which requires coordination of the agents.

		Agent 2: a_1	Agent 2: a_2
Agent 1	a_1	1, 1	0, 0
	a_2	0, 0	1, 1

Table 2.2: Payoff matrix for stag hunt game. The equilibrium with actions (a_1, a_1) is payoff dominant, while (a_2, a_2) comes with a low risk.

		Agent 2: a_1	Agent 2: a_2
Agent 1	a_1	3, 3	0, 2
	a_2	2, 0	2, 2

subject to dominance criteria or salience. In daily situations, coordination often evolves over time, allowing the agents to negotiate and find common solution in an extended time sequence, allowing to refine initial decisions. During the process of coordination, uncertainty regarding others' behaviors is reduced until coordinated behaviors are reached. This process is investigated in repeated games, as well as in general sequential games [58], [13], [55]. The process of coordination is supported by communication as well as transparent policies [26]. Agent coordination is not in the main focus of this thesis, although the contributions regarding understanding other agents and selecting communication actions should support new, efficient solutions of human robot coordination problems.

Signaling and information exchange

Strategic effects of information exchange are considered in signaling games [76]. A signaling game is based on information asymmetry. One agent (the sender) has private access to some piece of information, also considered as the sender's type. It can decide to share (parts) of it with the other agent (receiver) by selecting different signaling actions (messages). In a second step, the receiver can choose an action (depending on the sender's message) deciding over the final outcomes.

In such uncertain games, the agents need to develop a belief over uncertain states as basis for decision making. With a model for the sender's policy, the receiver can use the message to refine its initial belief via Bayesian inference. The actual amount of information exchange (equilibrium strategy of the sender) depends on the alignment of interests and the costs of misleading communication.

2.3 Modeling human behavior for cognitive human–robot interaction

For a robot interacting cooperatively with a human in a complex environment, an understanding for its partner is central to support her according to her needs. In the following, an overview is presented regarding research on cognitive human models in the application to human–robot interaction. Besides testing cognitive theories and hypotheses, a human model can serve for predicting human actions, to detect problems in her situation awareness and to support interaction and coordination. Such insights into human cognitive states are useful when planning interventions such as sharing relevant information to her (as in the blind spot scenario), or adapting actions according to the human plan.

Modeling other agents is not only relevant in human–robot interaction, but also investigated in multi agent settings, e.g. multi robot scenarios. Depending on the domain, the complexity of other agents, and the type of interaction, different models with varying levels of complexity are proposed in literature. On overview of agent modeling approaches is given in the survey of Albrecht et al. [3]. Agent models differ regarding whether and which internal states are considered, the available information sources for inference, whether they consider time variant behaviors, and intended use cases. In contrast to multi agent models, humans show complex, divers and time varying behaviors, which makes it difficult to adequately explain these by static or type based models.

2.3.1 Complex rational human models

Subintentional models, such as rule-based models, finite state machines or black box neural networks, can be used to predict an agent's behavior [78], [3], [42]. When trying to explain, why an action is chosen or which information might be helpful to the actor, it is however necessary to reason about the cognitive states including intentions and beliefs that could have caused these decisions. An important underlying assumption considers goal directed human behavior, which implies that she decides for actions to maximize her desires and long-term goals.

Inverse Reinforcement learning

One approach to estimate cognitive states explaining human behaviors consists in collecting observed sequences of actions and interpreting them retrospectively to find good explanations. This is done in inverse reinforcement learning (IRL, see also section 2.2.2), where an MDP model for the human is inversed to compute a reward function that best explains the observed behavior [72]. Since the problem of IRL is ambiguous, as several reward functions lead to the same optimal behavior, different criteria are proposed to derive useful explanations [126], [56], [6]. These approaches also relax the assumption of rational observations, which is important as humans do not always behave optimal. It is assumed commonly, that a human decision making can be explained by a softmax strategy that assigns higher probabilities to better actions (e.g. [88], [6]). When actions do not represent independent distinct options, the similarity of choices should also be considered in the human action selection model [12]. Besides decision noise, which is modeled by a softmax policy, there can also be a human decision bias caused by a different mental representation, e.g. regarding the transition function of an MDP [89]. Consequently, Reddy et al. include static biases into the human model and estimate them together with the reward function [89].

A reward function alone might not be sufficient to explain human behavior. In environments with uncertain environmental states that can be formalized as POMDP, not only human goals represented as reward function, but also her beliefs as representation of her knowledge will influence her behavior. There are extensions for IRL regarding uncertain environments with POMDP models. In a configuration, where only a robot observer is uncertain, while the human demonstrations are formed under full state knowledge, the robot can employ strategic perception, and additionally use observed decisions to reduce state uncertainty [62], [43]. Choi and Kim [24] consider the situation, that the demonstrations itself occur under uncertain state knowledge, based on the demonstrator's belief. Still, the approach considers the belief to be known to both agents, which can be reasonable in synchronized multi agent settings, but is not practicable for human–robot interaction.

Bayesian theory of mind

Modeling the human as fully independent POMDP agent is proposed in Bayesian theory of mind respectively Meta-Bayesian reasoning to understand observed human behavior under uncertainty [84], [5], [7], [27]. Inspired by human theory of mind (the capability of inferring other's cognitive states including desires and beliefs, see section 2.1.2), Bayesian inference is applied to estimate hidden human beliefs and desires that best explain her behavior. This was so far used to investigate human theory of mind, comparing inference result with human observer attributions, respectively to estimate prior human beliefs, but not to improve interaction with humans.

A POMDP formulation including human beliefs requires the specification of human perception to yield the observation function, complementing transition and reward function used in the noisy rational decision model in IRL. Interpreting human perception requires perspective taking to respect her point of view [68], and further needs to account for potential differences in information processing between human and robot (e.g. differences between human visual perception and robot's cameras). The resulting POMDP model for human behavior can be used to infer hidden mental states. The inference is still challenging, as the continuous human belief is already the result of human state inference, leading to a second order problem of what she inferred over state. Commonly, it is only applied to small underlying state spaces and retrospective evaluation [5], [7], [27].

Similar to human theory of mind, artificial theory of mind should concentrate on relevant aspects of human belief to keep inference tractable [48]. Correspondingly, Pöppel et al. propose to reduce complexity (e.g. the uncertain state space to consider) by switching between differently detailed models according to the needs and available computational resources [85].

When interacting with a human, a POMDP model for her, a model of her acting in isolation with her environment, does not account for her theory of mind, her representation of the robot she interacts with [80]. Full modeling of human theory of mind would lead to recursive construction of higher levels of theory of mind. Starting with an individual agent model and corresponding beliefs as level zero, higher levels of beliefs respectively behaviors can be generated recursively. Such construction of higher orders of theory of mind is proposed in context of multi agent settings to coordinate synchronized behaviors or solve logic puzzles [123], [121]. Avoiding infinite recursion, one could estimate another agent's recursion level from observed interaction behavior. Such higher-level ToM can support a coordination process of the agents, since one agent can expect others to reason about own plans [121].

All these approaches are computationally demanding due to the iterative construction of hierarchies. Illustrations use small multi agent settings, that either do not contain state uncertainty or rely on synchronized joint policies. When interacting with humans, higher orders of ToM become increasingly sensitive to human action noise and can easily produce misunderstandings.

Use in repeated interaction

The human models presented so far can explain complex and goal directed human behaviors. In repeated episodic interaction or similar static settings, estimated reward functions can be used to predict human behavior in future interactions, as it is e.g. proposed for driving scenarios [95], [96]. This is based on the assumption that the human reward function does not change between episodes. Similarly, inference results for static human belief configurations can be used to improve robot behavior in future episodes [89], [33]. But in practice, human belief will often change dynamically according to her perception, and also her goals might change.

2.3.2 Human models in human–robot interaction

The complex intentional human models presented so far followed the purpose to understand observed human trajectories retrospectively. It is used for research of human mentalizing, as in cognitive science and psychology, for training humans, or to transfer desired behaviors to a non-interactive robot.

The setting changes when entering direct interactions with humans. During interactions with others, a mental understanding is central to predict others' actions, to coordinate and to assist reaching goals. A human model to support human–robot interaction will need to be evaluated online during interaction to enable the robot to adapt its behavior and account for her in the current situation. In human–robot interaction research, simpler models are commonly used, concentrating on few aspects that may distinguish behaviors in concrete, investigated scenario. Such aspects include eye gaze behavior, trust, or intended movement goals.

Modeling human attention and perception

Measurement of human eye gaze can provide useful insights into human information gathering and attention. For visual perception of their environments, humans direct their gaze in the corresponding direction to fixate objects of interest. This can be measured via eye tracking glasses (equipped with eye cameras and a world camera) or by stationary eye tracking cameras.

Gaze information can help to predict a human's next movement. When reaching for an object, a human often directs gaze towards it in advance, which can be used to support early prediction of movement targets [97]. Similarly, a robot can interpret human gaze as attention signal, when assisting her in an assembly task by providing corresponding pieces [97], [77]. A gaze-based model of human attention can further be used to minimize disturbance by robot path planning, which avoids interfering with her line of sight [79].

Visual perception is important to gather information about the surrounding environment. This is necessary to gain situation awareness as perception is considered as first level of situation awareness, see section 3.4. Awareness regarding important aspect of the current situation is crucial for decision making and especially challenging in complex dynamic domains. To analyse human information gathering, it is necessary to combine gaze information with objects respectively information sources of the environment, while respecting the human perspective and potential occlusions. This is used to detect which of the objects and types of information might be visually perceived [101]. Such analysis can be used to detect problems in human situation awareness that are related to information perception [28]. Higher levels of human situation awareness, the relation and interpretation of different information sources and the prediction of future evolution, cannot be detected by only considering human gaze.

Evaluation of human perceptual behavior can further be used for assistance or warning systems. In critical situations she can be warned about missed objects, or a warning can be triggered, whenever her gaze deviates from required scanning patterns [102].

Aspects influencing human decision making

Besides models for human gaze and visual attention, also models for human decision making based on single hidden states, like robot trust or movement goal, are applied to human–robot cooperation. The decision model can be used to infer the latent variable of interest which can further be respected in the robot policy.

When a robot needs to coordinate its action with a human, it is generally useful to know the current human plan for her next actions. Anticipating her next subgoal can be used to avoid collisions and to decide on which subtasks to work on. Estimation of subgoals or movement targets is also often called intention inference and used in different domains, such as cooperative assembly, mobile navigation around pedestrians, automated driving, and virtual task planning [81], [73], [45], [114], [59], [38], [117], [87]. Methods include model free approaches, such as time series classification or neural networks, or simple action models, where the human policy model directly depends on the intended subgoal.

The problem of estimating human goals or intentions resembles a problem typically faced in spoken dialogue systems. Here, the system needs to understand the hidden intention of the user that led to the engagement with the system [122]. Uncertainty and difficulties arise from the diversity of human utterances and noisy speech processing.

Similar to human intention or movement goals, there can be other mental or other hidden aspects with significant influence on human behavior. In human–robot interaction, there are human models considering human trust in the robot, attentiveness respectively sleepiness, driving style, or trustworthiness [21], [96], [54], [115]. Using a model for their influence on human decision making, these static aspects can be inferred and used for human action prediction to improve the robot behavior.

Limits of human modeling

Models of human behavior can provide useful insights into causes and problems in reasoning. However, when using complex human models to build complex robot policies, one should be aware of possible new problems, e.g. due to missing transparency [26]. For interaction, it is not only important to understand (and adapt to) a human partner, but further that the human partner can understand the robot, hence that it behaves transparently. This is important as also the human will build a model of the robot and change her behavior accordingly [19], [80]. Hence neither assuming the human to behave as pure leader (requiring the robot to adapt) nor as pure follower (robot does not need to account for the human, as she totally adapts to it) will be adequate. To achieve efficient and flexible cooperation modalities, instead the robot will need to find a balance of understanding human goals and beliefs and transparency in its own behavior.

2.4 Communication strategies with humans and other agents

Communication is central for successful interaction as it enables exchange of information and coordination of behaviors (see also section 2.1.2). In this thesis, it is considered how communication can be used to efficiently support a human partner. Basic support systems may interrupt and instruct (as in car navigation systems) or take over control (as in emergency braking). In contrast, a less intrusive and human centric way consists in the communication of important information. It would support the human according to her needs and enable an informed human decision for the current situation.

When considering information sharing communication, it is central to account for the novelty for the human receiver (does she already know it) and the relevance of information (is this type of information important for the current situation?).

Regarding evaluation of novelty, a human model can be used to trace her information perception and decision-making processes. In this section, communication concepts from literature are shortly presented which either employ the evaluation of relevance of information for a current task, or adapt communication to the human receiver. In chapter 4, both aspects are combined leading to a new human-centric, information sharing concept.

Besides the questions, when information should be shared or what type of information is relevant, it is important to consider how communication signals are designed. This includes questions like: How can different types of information be shared to a human, what are the cognitive effects of communication, and which efforts or costs are related to it? Explicit as well as implicit communication signals and models were developed for human robot interaction [113], [29], [125], [18], [90]. However, the design of communication signals and modalities will not be the main focus of this thesis. Supposing an available communication interface, the remaining questions, when and what type of information to communicate are investigated with the target of an efficient and natural interaction between human and a robot.

Task relevance of information

The relevance of different types of information can change during a task, depending on the current situation and challenges. As discussed in the introduction, the blind spot warning communicating the presence of another car on the left lane is highly relevant when a lane change is initiated, as it may prevent an accident, but less relevant in other situations.

Task relevance is a dominating factor in critical situations, where the cost of communication is small compared to the risk of bad decisions. Similarly, when agents are spatially separated, such as multi agent exploration teams or shared initiative teleoperation settings, novelty might not play a role. It is rather obvious, that all exclusively perceived information must be new for other agents.

Explicit reasoning for task relevance for information sharing decisions was previously investigated in the area of multi agent research. However, these approaches rely on a synchronized joint policy for all agents. When introducing a cost of communication or bandwidth limitation, it becomes important to reason, "when" communication is beneficial for the joint policy by balancing the cost of communication with expected benefits [37], [120]. Goldman and Zilberstein [37] propose to optimize the time of communication with respect to the global, joint values. Melo et al. [64] further focus on situations or time steps in which interactions occur. Considering communication only in these situations can limit the effort of evaluation. Regarding the question "what" to communicate, Roth et al. evaluate available messages regarding their impact on the joint policy, selecting the message with largest information gain (considered most relevant).

In teleoperation settings, relevance of information can further depend on human operator preferences for different types of information [22]. Renoux et al. also respect the reliability of communication in relation to other possible information sources of the operator [91]. Therefore, the robot explicitly represents possible human beliefs to estimate effects of communication. Due to the spatial separation, it does not receive any feedback or observation of the human, which significantly limits belief estimation.

When agents are physically co-located, their perception will partly overlap and each other's actions can be observed. Consequently, these observations can be used to estimate the overlap in knowledge to avoid unnecessary communication (novelty) and coordinate shared local resources. One approach of respecting perceptual overlap consists in the representation of common knowledge, information of which everybody knows that everybody knows. Common knowledge is useful to coordinate agent's behavior [99]. Foerster et al. [34] extend the concept of common knowledge to learn effective synchronized strategic communication policies.

These approaches from multi agent communication and coordination rely on prior synchronization and the commitment to a known joint policy. When interacting with a human, this is not the case since policies are not synchronized beforehand nor can a robot assume the human to repeat the same behavior every time. Human robot cooperation rather represents an ad hoc setting without prior synchronization and interaction policies are required to be robust to divers human policies [108]. Barrett et al. propose a basic approach for communication in ad hoc settings, where communication is used to estimate others' type and to coordinate behaviors [9].

Human oriented communication

The communication concepts presented so far concentrate on the evaluation of external task relevance, while in the considered situations it was not necessary to account for the receiver's private knowledge. In human robot cooperation and human assistance, this changes. It becomes important to also account for the human knowledge and novelty of potential communication to achieve support while avoid repetitions and annoyance. For example regarding the timing of communication, there can be situations where the human is available or others, where she is engaged in activities. Respecting the human state of engagement will reduce annoyance and may also improve the communication success [18].

To support humans in specific situations, typical assistance approaches compare human behavior to a predefined policy (e.g. optimal behavior). If a deviation is detected, the system proposes an action to the human that is appropriate in the current situations. Such systems are for example proposed for support in health care and automotive contexts [44], [63], [102]. These concepts may detect and support human problems in specific situations. Still, it is not investigated, why the human deviated from expected behavior and the human typically will not be supported to understand her problem in situation awareness.

Communication concepts in human robot interaction research focus on coordination of agents' behaviors e.g. to avoid collisions [107], [74], [114]. Therefore, observed actions are considered in the planning process to detect possible coordination problems. Consequently, the robot may communicate its own intended action or propose an alternative to her [107]. Unhelkar et al. [114] further estimate, if the human seems to know the robot's action.

Explaining action proposals to the human (e.g. the robot is only able to perform one action and asks the human to do another) can help the human to understand its underlying intention improving her trust [116].

These approaches and the investigated tasks, such as stacking a few pieces together or carrying a table, are designed from a robot perspective. The only difficulty for the human typically arises due to the unexpected or non intuitive robot behavior. Communication is e.g. used to tell the human about the robot's plan she needs to adapt, as the robot misses capabilities to understand and seamlessly interact with humans. In this thesis, a different approach is considered towards a human perspective and human support. In complex situations with many different aspects, a human might profit from a robot sharing relevant information at the right time. For this approach, the robot does not only respect the difficulties of the task but further develops and evaluates a human understanding to decide when and what type of information to share.

2.5 Summary and conclusion

In this chapter, concepts, methods, and literature were presented, introducing ideas from human factors and human-human interaction, methods for inference and coping with uncertainty, human modeling and human centric communication.

Information sharing and communication are central to achieve and maintain situation awareness of a team or system in complex situations which is necessary for good decisions. When a robot interacts with humans, it is further important to account for transparency, to support human trust and avoid coordination and responsibility issues. To enable human centric information sharing and efficient interaction modalities, a theory of mind is an important capability to understand causes of her behavior and evaluate relevance of different types of information.

Complex environments as well as other agents introduce uncertainties that the agents need to cope with to successfully solve a task. Bayesian inference can be used for probabilistic inference of hidden states. It can be used to track uncertain system states as well as mental states of other agents. For nonlinear temporal filtering tasks, the Extended Kalman Filter and the particle filter provide useful approximations to achieve tractable inference. When uncertainty of the environment is quantified, the next challenge consists in decision making and action planning, to fulfill the current task or reach a goal. The decision problem can be formalized as partially observable Markov decision process and can be solved approximately by planning in discrete search trees. For filtering as well as decision making in large problem spaces, it is important to choose adequate representations and solution methods to achieve tractable computations.

These methods for coping with uncertainty can be used to plan a robot's behavior but also to understand and model humans acting under uncertainty. A variety of human models were proposed in literature to understand behavior and information processing retrospectively. On the other side, in human robot cooperation, the deployed human models are much simpler often concentrating on a single aspect or mental state such as attentiveness or current subgoal, which is relevant only for a single, investigated scenario. The situation is similar for communication concepts as proposed for multi agent concepts or human assistance approaches. Relevance evaluation and communication planning concepts are proposed for multi agent scenarios. The setting differs from human robot interaction, which additionally requires to respect for collocation and to develop robust asynchronous policies, which prevent the direct application of these concepts. Human robot communication concepts on the other hand commonly focus on the robot while neglecting human knowledge and needs.

3 Artificial Theory of Mind as online inference of human belief

The development of a theory of mind is a fundamental cognitive capability and essential for human interaction. Theory of mind describes a representation of others' states of mind, such as desires, beliefs, or intentions. This is important for understanding and predicting other's behaviors and is necessary to create cooperation and coordination between humans. For effective communication, it is important to consider what a communication partner already knows and what is relevant for her with respect to goals or desires. With this knowledge it is possible to select an appropriate message and avoid unnecessary communication [40].

Others' mental states are not directly accessible, one cannot literally read others' minds. Instead this latent information needs to be inferred from observable behavior, including task progression and information gathering. However, such observations do only represent the result of reasoning and not the process itself and many cognitive configurations could have led to the same observed behavior. As consequence of this sparse feedback, the development of an understanding of others can be difficult and error prone. From a practical view, it is not even necessary to derive a precise model of others, since not every aspect of human belief is of interest in an actual situation. Consequently, it will be sufficient to concentrate on a reduced set of relevant aspects to address in inference [48]. The degree of details can be typically reduced using higher abstraction levels, which are sufficient to explain observed behaviors. Considering the blind spot situation, it is not necessary to infer if the driver knows about weather conditions or the exact relative positions of other cars. It is sufficient to consider if she is generally aware of cars being present on related lanes.

An artificial theory of mind as human understanding will play an important role to enable sophisticated interaction between robots and humans. In contrast to research in cognitive science (targeting to approximate human's ToM [7]), and estimation of long-term human valuations via Inverse Reinforcement Learning [72], the use case of cognitive human robot interaction requires an online evaluation and interpretation. Human beliefs can change dynamically within a situation, which needs to be tracked during interaction to provide the robot a possibility to react and support her.

As a general setting, it is assumed that a human and a robot jointly work to fulfill a given task. Human robot cooperation will especially be useful in complex and uncertain environments, where different types of information need to be processed and where it is difficult to maintain situation awareness. Consequently, the robot needs to infer the human belief of these uncertain aspects to understand and efficiently support the human partner. When the human goal is unknown, belief inference might further be combined with inverse reinforcement learning to estimate possible human goals expressed as a human reward function.

As for a human, an artificial theory of mind will need to be based on observable information in the form of human actions and a model of human perception. An action might change the state of interest and further reveals a decision made by the human on the basis of her belief. By observing human perception (e.g. measuring similar signals or interpreting human information gathering behavior such as eye gaze) one can estimate, which information the partner receives. Human perceptive capabilities might differ from the robot's, which has to be respected in the perception model, additionally to perspective taking as usual in human ToM.

Human perception and decision-making behavior can be modeled as POMDP. Compared to model free approaches (e.g. [86]), it does not require to collect large amounts of interaction data for training [48], which would not be easy to gather, due to the general unavailability of human mental states. Extending prior approaches for Bayesian theory of mind (as in [7]), in this thesis inference is done online during interaction, as this is required to estimate her situation awareness and support her in the current situation. To handle the double inference problem, online approximations of belief representation and inference are introduced. A first version of this approach was published in [15].

In the remainder of the chapter, the formal approach of online inference of human belief is presented, that provides an understanding for the human and is used to estimate her situation awareness. It will be illustrated with an example, and further used and evaluated in a human robot task in chapter 5.

3.1 Belief definition and representation

The human is considered as an agent acting in a partially observable environment, defined as POMDP (see section 2.2.2). Her belief is defined as probability distribution over discrete environmental state $s \in S = \{s_0 \dots s_N\}$,

$$b = p(s),$$

respectively in vector form

$$\mathbf{b} = \begin{pmatrix} p(s_0) \\ \vdots \\ p(s_N) \end{pmatrix}.$$

Starting with a prior belief, she can perceive new observations o and influence the state s with her actions a_H. Both lead to updates in the probability distribution, as described by eq. (2.7). A discrete underlying state space S is considered, as it can result from abstracting low level sensory and motor signals to discrete actions and observations. Beyond physical cooperation (such as carrying a sofa together), interaction will focus on higher levels of abstraction, where a discrete state space is adequate. As example, for the blind spot scenario (Figure 1.1) it is not important (and neither predictable), if a human driver exactly knows the position of another car (continuous state), but rather, if its location is within some critical area. The blind spot situation can be represented with two different state values for the left lane, to be free or occupied $s \in S = \{\text{free}, \text{occupied}\}$. The driver will not always be sure about it, but have an uncertain belief. The belief about the state of the left lane will be updated by human information gathering (e.g. tracking other car's trajectories with eye gaze) and will also be affected by human actions (e.g. adapting speed or changing lanes).

The POMDP formulation models the human acting in isolation. To support coordination processes, it might be extended to also account for the human belief of the robot's intended action or plan. Therefore, a corresponding state aspect needs to be included, similar to a coordination state proposed by [13].

Human uncertainty regarding transition function and other task processes

In a challenging task, a human might not only be uncertain about states of the environment, but also about task processes, e.g. the exact effects of her actions, or the precision of sensory inputs. In the formulation of a POMDP, this could be represented as uncertain transition function T, reward function R, or observation function O.

However, typically these uncertainties do not span over a full function space, but rather address a specific mode or aspect. These aspects are considered by parameters $\theta = (\theta_T, \theta_R, \theta_O)$ determining the related processes, e.g. $T_{\theta_T}(s, a, s') = p(s' \mid s, a, \theta_T)$. All uncertainty can be shifted into an extended state, $s_{\text{ext}} = (s, \theta)$ that also contains these parameters, leading to an equivalent new POMDP with known processes (e.g. $T((s,\theta), a, (s', \theta)) = T_{\theta_T}(s, a, s')$).

3.1.1 Grid world illustrative example

The illustrative example considers a simulated human with limited vision moving in a grid world to reach a goal. The robot, as the second agent, is currently assumed as passive observer estimating human belief and situation awareness and will use it in the next chapter to provide helpful communication. Grid world domains are often used to gain insights and interpretative results. The configuration here was also used by the author in [16]. It is similar to the ones used in [7], [85], [61] regarding agent structure and perception limitations, but is designed to include multiple independent state aspects.

The grid configuration is shown in Figure 3.1 left. The human (H) can move forward, turn left or turn right (action set). Each action leads to the desired state change if the target state is accessible. Moving into wall cells (black) is not possible and the agent instead stays at the current position. The door cell (grey) can be open or closed, which defines its accessibility. An episode ends, when the agent H reaches the desired goal location, which is either cell $g_1 = (0, 2)$ or $g_2 = (3, 5)$. To favor goal directed behavior, every movement of the human, besides reaching the goal, leads to a constant negative reward $R_H(s_{k-1}, a_{k-1}, s_k) = -1$. She starts with a prior belief and can gain new information about cells within a limited field of view, one cell in front, to the left and to the right (blue triangle in Figure 3.1 left). An observation informs about accessibility of the cells in the field of view with a confidence of 90% each. She knows the overall

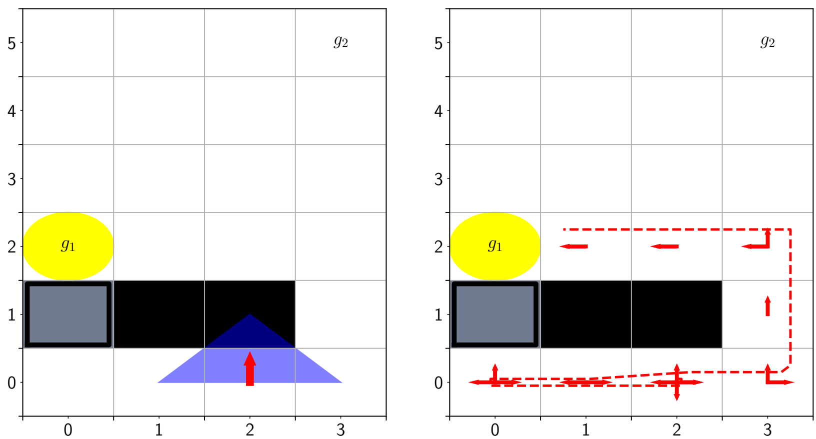

Figure 3.1: Grid world example (left). Human agent needs to move to the goal cell (yellow) while it cannot access to wall cells (black). Human can be unaware of position and orientation, door state (grey cell), or goal location g_1 or g_2. Relative field of view (blue) is shown for the starting position and orientation (red arrow). Resulting trajectory for a rational human agent with false initial belief of an open door (right).

wall configuration (the map), but is uncertain about its own position and orientation, the door state, and the position of the goal $g \in \{g_1, g_2\}$.

The environmental state contains position and orientation of the human agent. In the case of the door cell, the human is uncertain about the state transition T_{θ_T}, as the door state can be considered as parameter of transition function $\theta_T \in \{\text{open}, \text{closed}\}$. As example, the transition success from bottom left corner to door cell is,

$$T_{\theta_T}((0,0,\text{north}), \text{forward}, (0,1,\text{north})) = \begin{cases} 1 & \text{if: } \theta_T = \text{open} \\ 0 & \text{if: } \theta_T = \text{closed} \end{cases}.$$

The door parameter will be interpreted as part of environmental state s_{ext} for which the human knows the transition function T. Similarly, the goal location is considered as state aspect yielding an extended state space of 352 different discrete configurations that contains all aspects of human uncertainty (22 positions $\times$ 4 directions $\times$ 2 door states $\times$ 2 goal locations).

3.1.2 Factorization of human belief

Inferring human belief means calculating a probability density over human state probabilities, $p(\mathbf{b})$ and approximations are required for interesting scenario sizes. The first approximation will be a factorization of human belief to avoid combinatorical increase in its dimensionality.

In typical robotic or every-day environments, many state aspects are independent from each other, such as positions of different objects or agents. A full joint state space $S = S_0 \times \cdots \times S_M$ grows exponentially in the possible combinations of single aspect subspaces S_j. To limit the inference effort, a factorization of the belief representation is considered as also proposed by [14]. The human belief is assumed to factorize within these aspects, corresponding to the factored distribution

$$p(s) \approx \prod_j p(s^j) = \prod_j b^j, \; s^j \in S_j.$$

The factorization allows the independent representation of subbeliefs b^j, leading to an additive increase in belief space of $N_b = \sum_j |S_j|$ instead of $N_{\text{full}} = \prod_j |S_j|$. In the grid world example, the state is factorized according to three aspects, one for combined position and orientation, one for door state, and one for the goal location. This reduces belief dimensionality

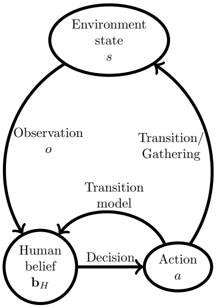

Figure 3.2: Model for human belief and information processing. While interacting with her environment, the human will experience state transitions and gather information about true environmental state s to update her belief $\mathbf{b}_H$.

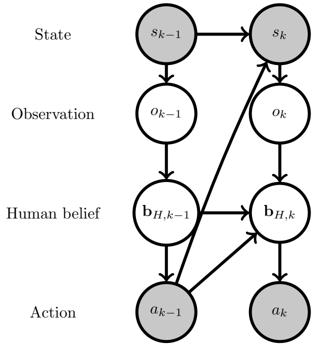

Figure 3.3: Belief filter structure as directed graphical model. Human receives information o over state s to construct a belief $\mathbf{b}_H$. The human belief can be inferred based on observed variables (grey nodes).

from 352 to 92 ($= 88 + 2 + 2$)[1]

This factorization may not hold for all situations, (e.g. when one agent is manipulating some object, agent and object states are coupled) but can provide a pragmatic solution to handle the curse of dimensionality, limit the effective state space size and enable tractable inference. Dynamic factorization might also help to respect such interdependencies [23].

3.2 Human model

A model for human information processing and decision making is the basis for an interpretation of observed behaviors and inference of her beliefs. With the assumption of an (approximately) rational human that wants to achieve a known task, the human model is formalized as POMDP, similar to the approaches of retrospective Bayesian theory of mind discussed in section 2.3.1. An abstract structure for human information processing

[1] In this illustrative example, also the full state space could be handled, as the number of aspects and overall dimensionality is relatively small. Factorization is introduced to demonstrate the principle which becomes necessary for larger, structured spaces as e.g. in chapter 5.

and environment interaction is illustrated in Figure 3.2. Based on her belief and the current task, she will decide for an appropriate action for information gathering or task progression. The action can lead to a state transition of the environment, internally represented as belief change. She will further receive an observation that provides new information of the current environmental state. These processes, state transition, perception and decision making are specified in the next paragraphs. They form the generative model that is afterwards inverted to infer the human belief.

3.2.1 Human belief transition

A human action will lead to a transition of the environmental state s_{k-1} to s_k according to the transition function T, that the human will account for in her mental representation. As for a rational agent in a POMDP (see section 2.2.2) the belief changes accordingly to

$$\begin{aligned} p(s_k \mid a_{k-1}) &= \sum_{s_{k-1}} T(s_{k-1}, a_{k-1}, s_k) p(s_{k-1}) \\ \mathbf{b}_k \mid a_{k-1} &= \mathbf{T}(a_{k-1}) \cdot \mathbf{b}_{k-1} \end{aligned} \tag{3.1}$$

respectively in vector form with the transition matrix $\mathbf{T}(a_{k-1})$, with elements $T_{i,j} = T(s_{k-1} = s_j, a_{k-1}, s_k = s_i)$ for the transition from state s_j to state s_i.

In factored state space, an action may only effect some state aspects (e.g. j and l), and a reduced transition matrix $\mathbf{T}^{j,l}(a_{k-1})$ for the corresponding joint subspace can be used,

$$\mathbf{b}_k^{j \times l} = \mathbf{T}^{j,l}(a_{k-1}) \cdot \mathbf{b}_{k-1}^{j \times l}.$$

The new factorized belief components are obtained by marginalizing out other state aspects.

For the grid world example, the deterministic transition function depends on the human position, orientation and on the door state.

3.2.2 Human perception

The human can sense her surroundings, e.g. by visual perception and extract information to update her belief. Depending on the applied domain, a model of human perception can be static, by respecting her field of view, as in the blind spot case and grid world case, or dynamic, respecting eye gaze. Additionally, uncertainty can be included such as observation noise limiting the human observation gain, e.g. for short glances. Both are combined to yield the stochastic observation function O.

When receiving an observation o_H, the human will extract the information by Bayesian update, inverting the perception model (as specified in eq. (2.7)). The Bayesian update of the human belief is modeled to follow an observation o_H (as introduced for POMDPs in section 2.2.2),

$$
\begin{aligned}
p(s_k \mid o_{H,k}) &= \frac{p(o_{H,k} \mid s_k) p(s_k)}{p(o_{H,k})} \\
\mathbf{b}_k \mid o_{H,k} &= \frac{p(o_{H,k} \mid \mathbf{s})}{p(o_{H,k})} \circ \mathbf{b}_k .
\end{aligned} \tag{3.2}
$$

For the grid world example, the human receives information about the accessibility of the three cells in her field of view, the cell in front, the cells on the left and on the right. The observation set consists of $2^3 = 8$ different observations, $\Omega = \{\text{free}, \text{occupied}\}^3$. The accuracy for each field is 90%, meaning she will perceive the true configuration with 73%, $p(o_H|s) = 0.9^{3-d(s,o_H)} \cdot 0.1^{d(s,o_H)}$ with $d(s, o_H)$ as number of deviations from the true configuration in her field of view.

Observation yields evidence for the human position since she knows the overall wall configuration. Additionally, if she is in the surrounding of the door, she will perceive the door state with 90% certainty.

Process uncertainty Human perception may not be optimally Bayesian [90]. Therefore, process noise will be used to cover suboptimal human inference or other unmodeled effects on human belief, such as forgetting. As one option, process uncertainty can be formulated as Dirichlet distribution for the human belief,

$$
p(\mathbf{b}_k \mid o_{H,k}, a_{k-1}, \mathbf{b}_{k-1}) = \mathrm{Dir}\left(\mathbf{b}_k \mid \tilde{\alpha} \mathbf{f}(o_{H,k}, a_{k-1}, \mathbf{b}_{k-1})\right), \tag{3.3}
$$

with precision parameter $\tilde{\alpha}$ and $\mathbf{f}$ representing the nominal Bayesian belief change of transition and perception model, eq. (3.1) and (3.2),

$$
\mathbf{f}(o_{H,k}, a_{k-1}, \mathbf{b}_{k-1}) = \frac{p(o_{H,k} \mid \mathbf{s})}{p(o_{H,k})} \circ (\mathbf{T}(a_{k-1}) \cdot \mathbf{b}_{k-1})
$$

3.2.3 Human decision making

A human action a_k is the result of the human decision-making process and therefore provides a source of evidence for human belief $\mathbf{b}_k$. The human decision process is assumed to be approximately rational, with an action probability depending exponentially on the expected action values $Q(\mathbf{b}_k, a_k)$ (as it is often used, e.g. [7], [95]).

The action probabilities are described by the softmax function,

$$\begin{aligned} p(a_k \mid \mathbf{b}_k) &= \frac{\exp(\tau Q(\mathbf{b}_k, a_k))}{\sum_{\tilde{a}} \exp(\tau Q(\mathbf{b}_k, \tilde{a}))} \\ &= \operatorname{softmax}(Q(\mathbf{b}_k, a_k)), \end{aligned} \tag{3.4}$$

with rationality parameter τ and action value function Q.

Modeling a stochastic policy accounts for uncertainties in human planning and decision making, and tolerates noise in action execution, as she might not execute the action she intended to do. Solving the POMDP to achieve the action values Q is computationally demanding, as discussed in section 2.2.2. At least, state values will also be needed for a good robot behavior and the POMDP solution could be shared between human model and robot policy.

For the grid world example, the human action values are computed using a tree based online solver (as introduced in section 2.2.2) with depth 2 and evaluating the leaf nodes based on MDP values (averaging values of a blind and a full knowledge policy).

3.3 Belief inference

The generative human model introduced in the previous section provides the basis to infer hidden human belief. It represents a Bayesian filtering problem structure, with the causal relations shown in the directed graphical model in Figure 3.3. Inference is split into the steps of predicting the belief transition followed by an update according to the evidence from the observed human action,

$$\begin{aligned} p(\mathbf{b}_k \mid a_k, s_k) &= p(\mathbf{b}_k \mid s_k) \frac{p(a_k \mid \mathbf{b}_k)}{p(a_k)} \\ &= \int \underbrace{p(\mathbf{b}_k \mid \mathbf{b}_{k-1}, a_{k-1}, s_k)\, \mathrm{d}\mathbf{b}_{k-1}}_{\text{prediction}} \underbrace{\frac{p(a_k \mid \mathbf{b}_k)}{p(a_k)}}_{\text{update}} . \end{aligned} \tag{3.5}$$

Exact inference would require to solve the integrals in prediction and update steps, which is not feasible for meaningful state spaces. Instead, approximate inference methods need to be used. In their offline approach with small examples, Baker et al. discretize the belief space either uniformly or select relevant configurations by hand [7]. With uniform discretization however, the number of considered beliefs increases exponentially in the number of states, becoming intractable for interesting scenarios. Similarly, a specification of relevant belief configurations by experts in advance can be time consuming as well as difficult and error prone. Instead, general inference methods are proposed respectively refined to solve the inference problem of human belief.

A particle filter (see section 2.2.1) follows a sampling-based approach to approximate the resulting belief distributions and can be used to approximate eq. (3.5). Since online planning methods are used to compute human's action values, the evaluation of human action probabilities in update step requires most computation effort, contradicting fast online application. As second deterministic approach, a linearization-based filter is developed based on parameterized distributions for the human belief. As the deterministic approach requires only a few evaluations it provides the potential of very fast inference and further shows benefits when observing unexpected actions.

3.3.1 Human belief particle filter

Particle filtering can be directly applied to the generative human model, using the equations (3.3), (3.4). Starting with an initial set of particles, each particle is transformed and weighted according to prediction and update equations. The weighted particle set represents the approximate distribution of human belief.

In Figure 3.4 left, the inference is visualized for a one-dimensional example of the blind spot case[2]. Inference starts with a distribution for the last belief (blue) favoring human beliefs where the left lane is free. 30 samples are drawn from this distribution and predicted through the model (orange and green crosses) by modeling state transition and two possible observation the human might receive. When observing the human action (staying on lane), the action likelihood (violet) is used to compute the new weights (black circles). The weighted samples approximate the posterior distribution.

The approximation quality depends on the number of samples. Especially for unexpected human actions (human changes lane) this can be a

[2]For which the full belief space can be shown and the true posterior can be computed

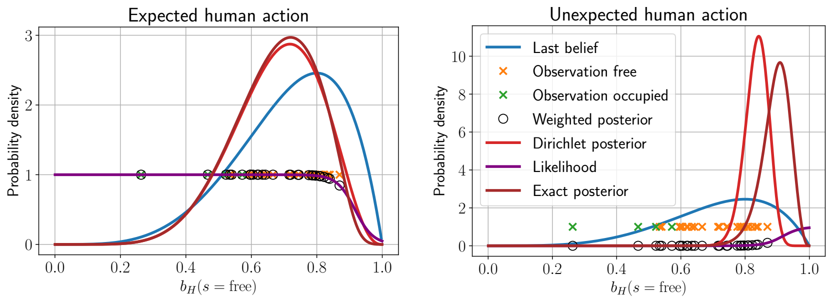

Figure 3.4: Illustration of one inference step for human belief inference (blind spot example). Starting with a distribution for last belief (blue), 30 belief samples are drawn, predicted, and updated according to the action likelihood function (violet). A remapping to a Dirichlet distribution is drawn in red. In contrast to the expected human action (left) for an unexpected action (right) only few samples remain with relevant likelihood.

problem, as the likelihood eq. (3.4) will be high at low probability regions. Only few samples contribute to the posterior approximation reducing inference quality, see Figure 3.4 right. Increasing the number of samples will increase the computation effort and a trade off of accuracy and resources is required.

3.3.2 Linearization-based inference

As an alternative to sampling-based particle filter, a deterministic approach based on linearization is developed for approximate belief inference. Like the extended Kalman filter (section 2.2.1), it is based on a parameterized distribution to represent uncertain human belief combined with a Taylor series approximation for the human model.

Parameterized distributions for human belief

The belief itself represents a probability distribution over states and the space of possible belief configurations is the probability simplex[3] $\mathbf{b} \in \mathbb{R}^N \mid b_i \geq 0, \sum_{i=1}^N b_i = 1$.

[3]Due to the constraint, the effective state dimension is $N-1$ as $b_N = 1 - \sum_{i=0}^{N-1} b_i$. For factorized belief representations, each subspace represents one probability simplex.

The Dirichlet distribution and the logistic normal distribution are candidates for parameterized distributions to cover the probability simplex. The Dirichlet distribution is parameterized by N parameters α. Since it is member of the exponential family, a Bayesian belief update can be formulated by linearizing the log likelihood function [10]. In contrast, the prediction step of filtering is more complex, as the combination of Dirichlet distributions in general does not yield another Dirichlet distribution. One option consists in a hybrid approach, using sampling-based methods for the prediction step (which leads to fast evaluation) combined with a linearization-based update. However, the Dirichlet distribution further limits the belief representation as the set of parameters cannot represent correlations between the different belief dimensions.

More flexibility is provided by the logistic normal distribution [2]. It is based on the additive logistic transform

$$\mathbf{b} = \begin{pmatrix} \frac{\exp(z_1)}{1+\sum_j \exp(z_j)} & \cdots & \frac{\exp(z_{N-1})}{1+\sum_j \exp(z_j)} & 1 - \sum_{j=1}^{N-1} b_j \end{pmatrix}^T =: \operatorname{alt}(\mathbf{z}) \tag{3.6}$$

and its inverse

$$\mathbf{z} = \operatorname{alt}^{-1}(\mathbf{b}) = \begin{pmatrix} \log\left(\frac{b_1}{b_N}\right) & \cdots & \log\left(\frac{b_{N-1}}{b_N}\right) \end{pmatrix}^T,$$

transforming the belief $\mathbf{b}$ to an $N-1$ dimensional real number $\mathbf{z} \in \mathbb{R}^{N-1}$. After transformation, a normal distribution $\mathcal{N}(\mathbf{z} \mid \mu, \Sigma)$ is considered. Correlations of belief dimensions can be represented in the covariance matrix Σ.

With the factorized belief representation, each subbelief is individually transformed, $\mathbf{z}^j = \operatorname{alt}^{-1}(\mathbf{b}^j)$ and stacked together to yield one vector of real numbers,

$$\mathbf{z} = \begin{pmatrix} \mathbf{z}^1 \\ \vdots \\ \mathbf{z}^M \end{pmatrix}.$$

The additive logistic transformation further allows application of other distributions and algorithms for real valued random variables, e.g. a mixture of Gaussian representation within the multi hypothesis filter, which supports multi modal distributions [69].

Approximate inference based on linearization and a logistic normal distributions

To reduce the effort of sampling-based inference, deterministic approximations based on the logistic normal distribution can be used. Using the additive logistic transform, the latent belief distribution is approximated as a normal distribution in the transformed space, $p(\mathbf{z}) \approx \mathcal{N}(\mathbf{z} \mid \mu, \Sigma)$. Consequently, the prediction step of filtering can be approximated as for the extended Kalman filter,

$$p(\mathbf{z}_k \mid \mathbf{z}_{k-1}) = \mathcal{N}\left(\mathbf{z}_k \mid \tilde{\mathbf{f}}(\mu_{k-1}), \nabla\tilde{\mathbf{f}}(\mu_{k-1})\Sigma_{k-1}\nabla\tilde{\mathbf{f}}(\mu_{k-1})^T + \Sigma_{pn}\right),$$

by linearizing the transition function

$$\mathbf{z}_k = \tilde{\mathbf{f}}(\mathbf{z}_{k-1}) = \mathrm{alt}^{-1}(\mathbf{f}(\tilde{o}_{H,k}, a_{k-1}, \mathrm{alt}(\mathbf{z}_{k-1}))),$$

using eqs. (3.3), (3.6).

For each possible human observation o, a different prediction follows, leading to a mixture of Gaussians. For the case, that a single observation $\tilde{o}_{H,k}$ dominates (as in the grid world model), it is sufficient to focus on this observation and subsume other observations within process uncertainty. The process noise is respected directly in the transformed space with covariance Σ_{pn}.

The update step of filtering is based on the likelihood function $l(\mathbf{z}_k) = p(a_k \mid \mathbf{b}_k = \mathrm{alt}(\mathbf{z}_k))$ (human decision model, eq. (3.4)). To regain a normal distribution (which is central for subsequent steps), the likelihood function needs to be approximated by a normal distribution.

In the following, this approximation is derived by first order Taylor approximation, yielding a one-dimensional Gaussian approximation of the likelihood function.

Therefore, the likelihood function is represented as the exponential of a squared nonlinear function[4] h,

$$\begin{aligned} l(\mathbf{z}) &= \exp(-0.5h(\mathbf{z})^2) \\ h(\mathbf{z}) &= \sqrt{-2\log(l(\mathbf{z}))} \end{aligned}$$

[4]For this definition of h, the argument of the logarithm needs to be in the range $(0, 1]$. As the likelihood is a probability function of human action, $0 \leq l(\mathbf{z}) \leq 1$. From a general, logical point of view, the human model should guarantee, that the likelihood of an observed action is greater zero. This is the case for the softmax decision model eq. (3.4).

This function h is approximated by a first order Taylor series around a point $\mathbf{z}_0$, $h(\mathbf{z}) \approx h_0 + \nabla h_0(\mathbf{z} - \mathbf{z}_0)$, where $h_0 = h(\mathbf{z}_0), \nabla h_0 = \nabla h(\mathbf{z}_0)$. This yields a one-dimensional Gaussian approximation of likelihood,

$$\begin{aligned}
\log l(\mathbf{z}) &\approx -0.5 \left[h_0 + \nabla h_0^T (\mathbf{z} - \mathbf{z}_0)\right]^2 \\
&= -0.5 \Big[\mathbf{z}^T \nabla h_0 \nabla h_0^T \mathbf{z} + \\
&\qquad + 2 \left(h_0 - \nabla h_0^T \mathbf{z}_0\right) \nabla h_0^T \mathbf{z} + \\
&\qquad + \left(h_0 - \nabla h_0^T \mathbf{z}_0\right)^2 \Big],
\end{aligned}$$

with precision matrix $\nabla h_0 \nabla h_0^T$.

With this approximation and the prior distribution $p(\mathbf{z}) = \mathcal{N}(\mathbf{z} \mid \mu, \Sigma)$, the log posterior becomes

$$\begin{aligned}
\log(p(\mathbf{z} \mid a) &= \log(p(\mathbf{z}) + \log(l(\mathbf{z})) + \text{const} \\
&\approx -0.5 \big[\mathbf{z}^T \left(\Sigma^{-1} + \nabla h_0 \nabla h_0^T\right) \mathbf{z} + \\
&\qquad - 2 \left(\left(\nabla h_0^T \mathbf{z}_0 - h_0\right) \nabla h_0^T + \mu^T \Sigma^{-1}\right) \mathbf{z} + \\
&\qquad + \text{const}\big]
\end{aligned}$$

From the quadratic form, the posterior normal $p(\mathbf{z} \mid a) = \mathcal{N}(\mathbf{z} \mid \mu_{\text{post}}, \Sigma_{\text{post}})$ can be obtained with

$$\begin{aligned}
\Sigma_{\text{post}} &= (\Sigma^{-1} + \nabla h_0 \nabla h_0^T)^{-1} \\
\mu_{\text{post}} &= \Sigma_{\text{post}} \left(\left(\nabla h_0^T \mathbf{z}_0 - h_0\right) \nabla h_0 + \Sigma^{-1} \mu\right).
\end{aligned}$$

The point of linearization can be selected as the prior mean $\mathbf{z}_0 = \mu$. As alternative approach, one could start with the prior mean and iteratively refine the point, based on computed posteriors (iterated extended Kalman filter [124]).

As for the particle filter, linearization-based inference is illustrated using the one-dimensional blind spot example in Figure 3.5. In contrast to the particle filter, it shows a good performance nevertheless of the action likelihood. The derived linearization-based inference method is significantly faster, as it only requires a single evaluation of the likelihood function and its gradient, in contrast to an evaluation per particle of the particle filter.

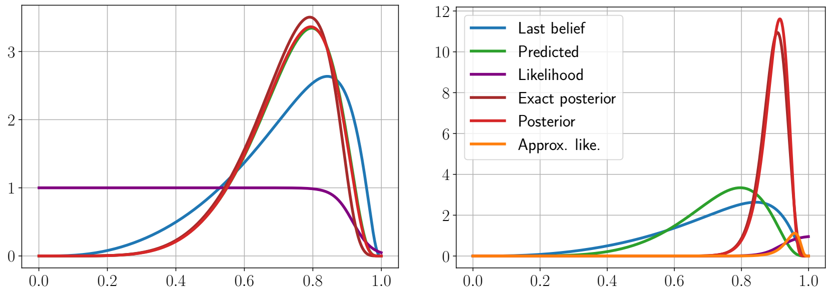

Figure 3.5: Illustrating one step of linearization-based inference (blind spot example). The last belief (blue) is predicted for the most probable observation and updated according to the logistic action likelihood. The left plot shows an update for an expected human action, the right one for an unexpected. In contrast to particle filtering, the approximate posterior is close to the true one in both cases.

3.4 Quantitative formulation of situation awareness

The inference of human belief can provide insights about human reasoning and explanations for her behavior. It can further be used to quantitatively evaluate human situation awareness (SA). In contrast to classical SA measurement methods like SAGAT or SART [32], this method does not require interruptions nor subjective assessments. In this work, situation awareness will be formalized as a representation in the human mind which is sufficient to enable good decision making. Hence, every environmental state aspect that is relevant to cope with the current situation should be represented with sufficient certainty in the human belief.

The task relevance of information can be expressed in terms of action values, by stating that action values, given true state, should be similar to action values based on human belief. The optimal action values $Q^*(s, a)$ based on true state s are used to evaluate possible human decisions. If the human belief contains all relevant aspects, her subjective action value function should be similar enough to the optimal one to allow for good decisions $p(a \mid \mathbf{b})$. Situation awareness is considered as belief configuration with small expected value loss,

$$\mathcal{L}(\mathbf{b}, s) = \max_a Q^*(s, a) - \sum_a p(a \mid \mathbf{b}) Q^*(s, a). \tag{3.7}$$

The value loss is a graded quantity (a measure for situation "unawareness"). For a binary decision, it can be compared to a threshold δ of acceptable value loss[5]. Consequently, an agent with belief $\mathbf{b}$ at true state s is situation aware if

$$\mathcal{L}(\mathbf{b}, s) \leq \delta.$$

A lack of situation awareness does not mean, that the human will actually select a bad action. However, it is expected, that she considers suboptimal actions in her decision making, since it is based on an inadequate situation understanding. The use of the value function also accounts for long term effects. Even if the current decision does not cause severe consequences immediately, it might lead to worse situations in the future.

The evaluation of situation awareness and options to support a human will depend on the quality and the timing of the belief inference results. It might not always be possible to detect problems in the human belief representation before a suboptimal human decision is made due to limited observability through sparse feedback. However, within a longer sequence, an early unexpected action can provide the required evidence to draw corresponding conclusions that intervention might be helpful for later situations.

3.5 Illustration

In the following, the approach of repeated inference of human belief is demonstrated on the grid world example for multiple time steps starting with one characteristic situation. It is used to demonstrate and evaluate the process of double inference with the different approximate inference approaches.

By simulating an agent in place of real human, it is possible to compare the inference result to the true agent belief for different configurations. The true agent belief is not used in the algorithm itself, e.g. in contrast to learning based approaches, where artificial agents are used to generate training data.

[5]The choice of δ needs to respect the scale of a reward function

3.5.1 Belief inference

In the following, one specific scenario is considered, where agent H starts in position (2, 0, north), as shown in Figure 3.1 left. She is unaware of her position and orientation within the set {(1, 0, north), (2, 0, north), (2, 2, south), (1, 2, south)}. She starts with a belief for an open door (while it is closed) and goal location g_1 (true location). The robot observer R starts with a uniform belief of the H belief.

The human model uses a rationality parameter of $\tau = 10$. Regarding process noise, the particle filter uses an $\tilde{\alpha} = 200$, while process noise of linearization-based approach is chosen to be diagonal, $\Sigma_{pn} = 0.09I$. Both introduce uncertainty for the human belief in the range of 0.06. The agent H rationally updates her belief while she receives observations of the true surrounding. For the policy, the action value function is evaluated using online planning with depth 2 search combined with MDP based leaf evaluation.

The trajectory of H is shown in Figure 3.1 right and the inference results in Figure 3.6 regarding position belief, respectively Figure 3.7 for door and goal belief. As H does not know her position, she starts with a right turn. The following observation allows her to disregard positions (2, 2, west), (1, 2, west), see Figure 3.6, $k = 1$. Her position belief focuses on (0, 1, east) and (0, 2, east). Since H believes that the door is open, she turns right two more times, heading towards the door in time step $k = 3$. Afterwards, H moves forward until reaching the door, where her perception yields a correction of door belief within time steps 5 and 6, see Figure 3.7. Afterwards she takes the longer but open path to reach goal g_1 (trajectory shown in Figure 3.1, right).

At the beginning, the robot is mainly uncertain about the human belief. The second human decision (second right turn at $k = 1$) allows the robot to correctly infer the human door and goal beliefs (Figure 3.7). From there on, estimated and true human beliefs remain close together.

3.5.2 Estimation of human situation awareness

At the beginning of this episode, the human agent is uncertain about state as well as holding a false belief regarding door state. After the first time step, observation reduces position uncertainty for H which is also expected by R due to its perception model of H. However, the erroneous door belief remains unknown to the robot.

The expected human belief is used to evaluate situation awareness ac-

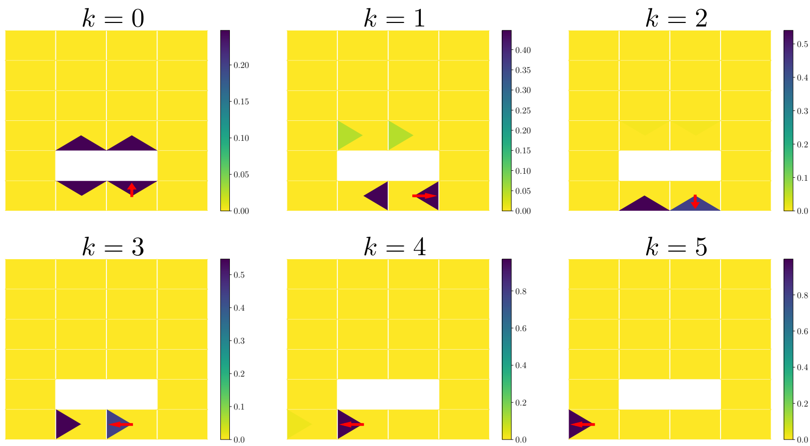

Figure 3.6: Mean of inferred belief of position and orientation. Each grid cell is divided according to possible human orientations. True agent position is marked by red arrow. In the forth time step $k = 4$, the belief converges to the true position.

cording to the quantitative criteria of expected value loss, eq. (3.7), shown in Figure 3.8. In the beginning, the estimate of human belief is uncertain (see Figure 3.7), leading to an imprecise estimation of her awareness (not visible in the expected value). The second human decision however provides sufficient evidence to infer her goal and door belief, and the estimated situation awareness converges to the true value. The suboptimal second human decision (moving forward towards the closed door) could not be predicted in advance, since there were not enough observations to infer a meaningful belief. However this decision provides an indication for the next time steps, where H will still not be aware of the situation. In time step 4, a good human decision is expected even though she has an uncertain door belief (in this special situation door belief is not relevant). After the 5th action, the human corrected her initial false belief and from there on, she maintains situation awareness.

3.5.3 Inference methods and computation times

Both approximate inference methods are able to infer the hidden human belief, shown in Figure 3.9. They slightly differ in the speed of convergence

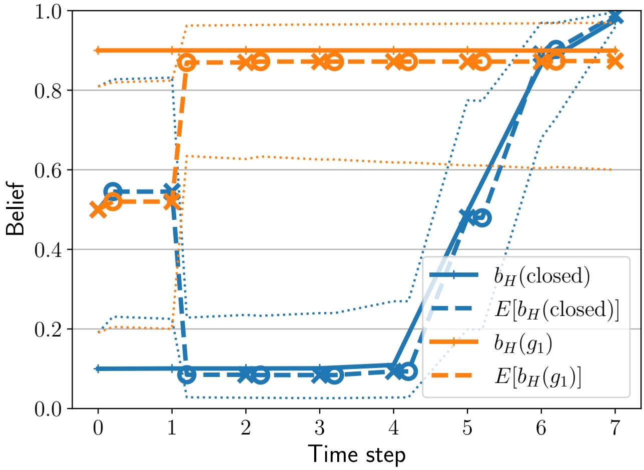

Figure 3.7: Human belief and linearization-based inference results (dashed) regarding door (blue) and goal (orange) state aspects. For belief inference, prediction (cross) and update (circle) as well as standard deviation (dotted) are shown. Second action ($k = 1$) reveals false door belief of human, which she updates later on when perceiving true door state in step 3 and 4.

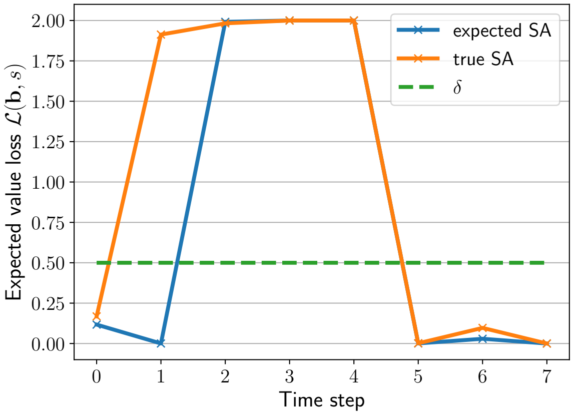

Figure 3.8: Evaluation of Situation Awareness as expected value loss for expected human belief (blue) and true human belief (orange), low values mean H is situation aware. The second action reveals that the human is not aware of the current situation (due to her false door belief).

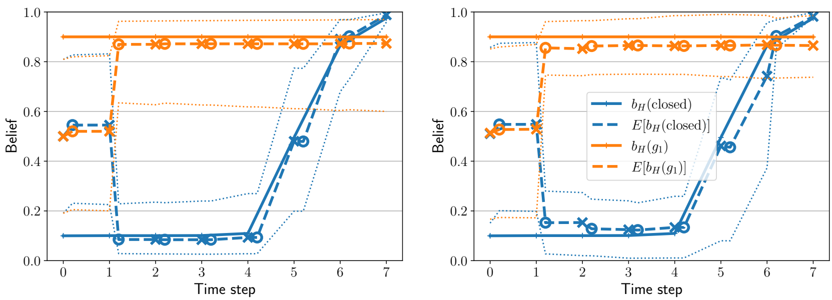

Figure 3.9: Belief inference with update based on linearization (left) and sampling (right, with $K = 1000$ particles).

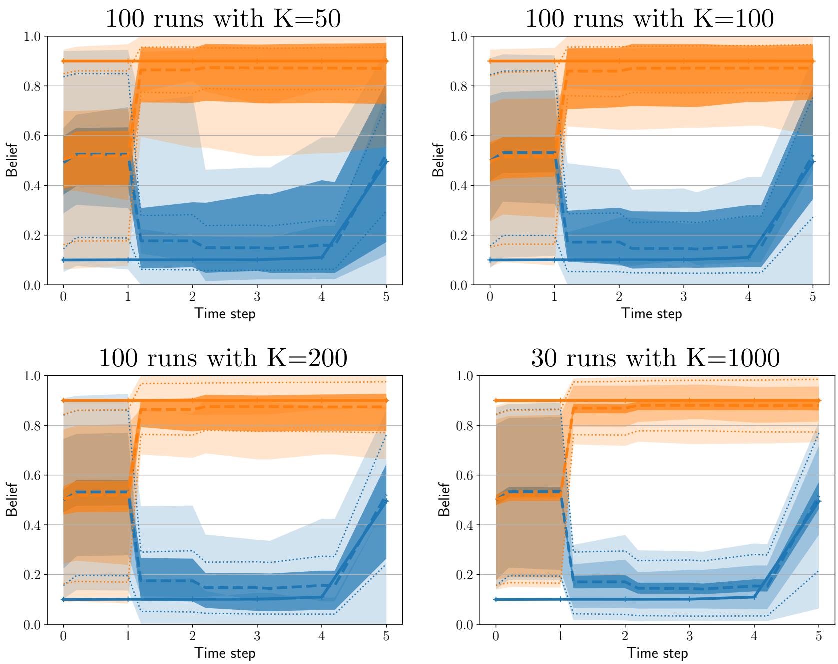

Figure 3.10: Comparison of different sample sizes for particle filter inference. Curves show range of expectations of different runs (dark colors) and the range of standard deviation (light colors). With a higher sample number, the results of different runs stay closer together.

(the linearization-based approach becomes faster certain about the false door belief). Differences might originate from the unimodal model of the linearization-based approach, as well as from the problem of the particle filter to process unexpected actions. Nevertheless, both curves contain the true H belief in the confidence interval from the second time step and can provide useful insights for the interaction.

Regarding the number of samples, Figure 3.10 shows a statistic visualization of inference results for several runs with different numbers of samples.

The difference in computation effort however can gain more impact. The sampling-based inference becomes slow with an increasing number of particles, even for this relatively simple, intuitive example. One evaluation

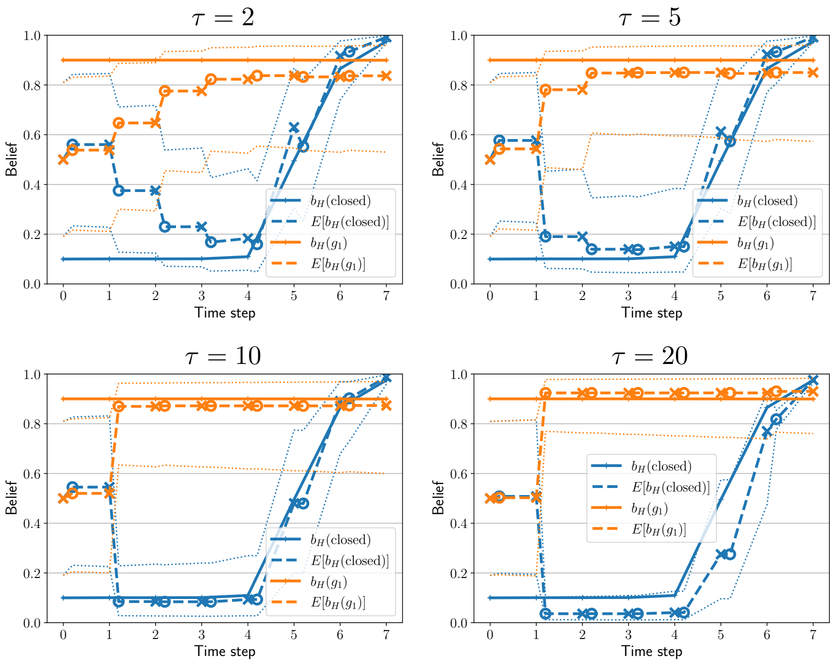

Figure 3.11: Inference results (linearization-based) for different rationality parameters, $\tau \in \{2, 5, 10, 20\}$. With increasing rationality, belief updates (circles) become more confident.

of human action values takes about 50 ms, which is required to update each belief sample (programmed in python, on core i5u processor with 2 GHz). For a number of 100 samples, evaluation takes about 5 seconds. This is a limiting factor for the target of inferring and supporting a human online during the interaction (as one action need to be processed until the next is executed).

The linearization-based update requires only one evaluation of the action values and its gradients, leading to a much faster inference step of about 0.16 s. Consequently, this method is suited for much more scenarios with shorter sample respectively human decision times.

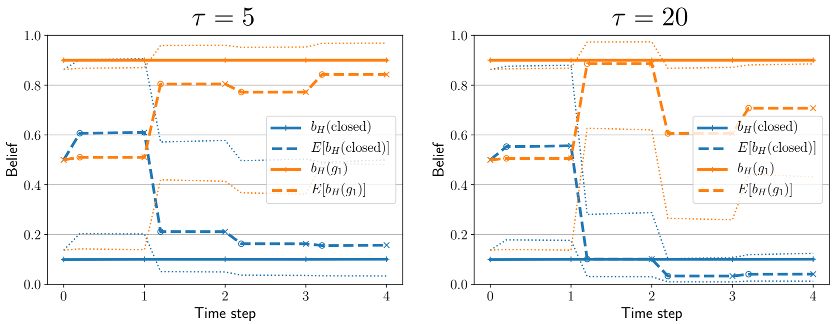

Figure 3.12: Inference results (linearization-based) for an unexpected human action at time step $k = 2$ for rationality parameters, $\tau \in \{5, 20\}$. The higher the modeled human rationality the less it tolerates noise, and inference instead searches for alternative explanations.

3.5.4 Rationality parameter

The rationality parameter τ (eq. (3.4)) specifies the expected randomness of human behavior. Assuming a stochastic policy is important to tolerate human noise, e.g. uncertainties in planning, decision making, or action execution. A higher rationality parameter implies higher confidence that the human takes the optimal action and hence allows faster updates of her belief (see Figure 3.11).

On the other hand, human actions deviating from the optimal policy will easily mislead inference of her belief when the rationality parameter is high. Let's assume, that the human is taking as third action "move forward" instead of "turn right" while she is heading to the grid border ($k = 2$). This action does not have an effect on the state, is unnecessary and unexpected. The inference process tries to explain the observed action by searching for reasonable beliefs, for example a different human goal belief. Lower rationality assumptions are more tolerant to noise and the influence of unexpected human actions is less severe, Figure 3.12. If the human is taking an unexpected action, which is reasonable for a different belief configuration, this belief configuration will be considered more likely in all configurations.

3.6 Summary and conclusion

A framework for online inference of human belief was presented to form an artificial theory of mind for human robot cooperation. This human understanding is developed combining available observations of information gathering as well as task progressing behavior to form a complete view on human information processing. It is based on modeling the human as approximately rational POMDP agent, representing tasks within complex and uncertain environments. To account for the computational complexity of second order inference, first an approximate representation of human belief, based on factorization, is introduced. Second, two approximate methods were proposed for actual inference, a stochastic particle filter and a newly derived efficient linearization-based approach.

The estimate of human belief provides an understanding of human behavior. It is used to propose quantitative formulation of situation awareness, which provides a non-intrusive estimation of human awareness according to expected value loss of her belief configuration. The proposed concepts and methods were demonstrated and evaluated on a grid world example with different uncertain aspects.

The developed online inference of human belief provides a basis for human centric human robot cooperation. In the next chapter, it is used to develop an intelligent information sharing strategy that enables a robot to support a human according to her needs.

4 Theory of mind based assistive communication

An artificial theory of mind, as derived in the last chapter, provides a rich base for human centric interaction. When it is possible to detect a problem in human situation awareness online during interaction, this is valuable and can be used to actively support a human partner.

Therefore, a robot could inform the human about aspects if she holds a false or uncertain belief of it and if the information is relevant in the current situation. This is a core idea of the developed human centric assistive communication concept regarding the decisions when and what type of information to share to support a human partner. Theory of mind, the understanding of interaction partners, is an important cognitive capability for efficient human communication [40]. Already three years old children take receivers' knowledge into account and decide to tell them unknown information [35]. Sharing information that is new for a human can support her situation awareness. Still communication requires perception and processing resources including attention mechanisms. Receiving too much information will therefore distract and overload her and provoke annoyance. Instead, only currently relevant information should be exchanged to balance these costs of communication against expected benefits.

Combining theory of mind with an evaluation of information relevance regarding the current task, leads to the new concept of theory of mind based assistive communication (ToM-Com), introduced in the author's publication [16].

It specifies a human centric assistance concept with the target to enable a human to make good decisions by supporting her situation awareness. This contrasts typical approaches that simply tell her what to do, when some deviation is detected.

The concept of theory of mind based communication considers sharing environmental information from a robot partner to a human. Other functions of communication, e.g. for coordination purposes, are not in the focus here. A fully cooperative setting is assumed, meaning that the robot wants to maximize human reward while minimizing cost of communication. The basic principle is illustrated in Figure 1.2. Based on second level reasoning (belief inference as introduced in last chapter) the robot can anticipate and evaluate possible outcomes to decide whether, when and what information to communicate to a human partner. In this chapter, a principled formulation for this communicative assistive communication is presented. The problem of when and what type of information to share is formulated as a POMDP with human belief as uncertain underlying state. Robot's action planning uses the estimated human belief for evaluating uncertainty and relevance.

4.1 Effects and efforts of communication

For planning communication decisions, it is necessary to account for effects and efforts of possible communication. In the considered setting, a robot has access to different actions with communicative effects, $a_c \in A_c$. In each time step, it can choose to apply one of these or to avoid communication, deciding when and what information it should share with a human. Communication can be direct (such as speech) or indirect (via expressive motions). Different concepts regarding the design of communication actions are presented in section 2.4. Here, it is assumed, that a communication interface is available providing a set A_c of different actions. Each communication action will contain some type(s) of information that can be transmitted with a corresponding confidence or success rate.

Informative effects of communication

The communication initiated by a robot action a_c will generate an additional human observation o_c, as she receives the communication signal. This observation can include information regarding the true environmental state s. Reliability of communication and the contained information will shape the communicative observation function, $p(o_c \mid \mathbf{s}, a_c)$ for some communication action a_c. Receiving a robot's communication can be modeled as Bayesian belief update of the human, as for an environmental observation, eq. (3.2),

$$\begin{aligned} \mathbf{b} \mid a_c, o_c &= \frac{p(o_c \mid \mathbf{s}, a_c)}{p(o_c \mid a_c)} \circ \mathbf{b} \\ &= \mathbf{f}_{\text{comm}}(\mathbf{b}, o_c, a_c). \end{aligned} \tag{4.1}$$

Extending the grid world example from the previous chapter, the robot observer now obtains the capability to support the human agent with information. It has the choice between three communication actions, additionally to the option of not communicating. It can tell the human about one aspect of the true state, regarding position and direction, door state, or goal location. The robot communicative action set is formed as, $A_R = \{a_\emptyset, a_{\text{pos}}, a_{\text{door}}, a_{\text{goal}}\}$.

With a reliability p_{comm} as probability of communication success, the aspect based communication function becomes

$$p(o_c \mid a_{c,i}, s) = \begin{cases} p_{\text{comm}} + p_{\text{noise}} & \text{if: } o_c = s^i \\ p_{\text{noise}} & \text{else} \end{cases} \tag{4.2}$$

For the grid example, reliable communication is assumed, with $p_{\text{comm}} = 0.99$.

In this example, communication actions in A_R only differ in the type of information they contain. It is further possible, to use messages regarding the same state aspect that instead differ in their reliability (e.g. short utterance compared to long explanations). Depending on situation, it can be worth to spend more effort for reliable exchange or less for uncertain communication, which is both covered in the proposed problem formulation.

Strategic effects of communication

Besides the information that is included in the communication signal, a human can further interpret it as the result of robot decision making. To account for such interpretations, a robot would need even higher orders of theory of mind (the robot reasons what the human thinks that the robot wanted to achieve) or strategic reasoning of possible equilibria of a signaling game (section 2.2.3). For example, a human could be used to be supported by a robot, e.g. it always warns her before her making an error. Receiving no communicative warning then tells her that everything is fine, reducing uncertainty and allowing “riskier” behaviors. Strategic effects of communication consequently allow for more efficient communication (as no communication can transport information) which is interesting in settings with sparse exchange options, e.g. considered in the card game Hanabi, [34]. However, such implicit communication only works if both agents perform similar strategic reasoning and otherwise may lead to misunderstandings.

Strategic effects might be included in the model as a human can not only processes the communicated information o_c, but also accounts for an estimated robot intention or communication policy, $p(a_c \mid s, R_R)$. It leads to a third human belief update, where the robot's communication a_c serves as additional human observation

$$\mathbf{b} \mid a_c = \frac{p(a_c \mid \mathbf{s})}{p(a_c)} \circ \mathbf{b}$$

This can be directly included in the belief inference framework. The derivation of a robot policy that the human might use in her reasoning ($p(a_c \mid \mathbf{s})$) however remains for future research.

Costs of communication

For each communication action, a cost of communication $-R_{\text{comm}}(a_c, s)$ is considered to respect cognitive efforts of both, sender and receiver and related effects such as distractions or time delay in task execution.

The cost of communication can depend on the current situation (state s), e.g. in high stress conditions, distraction is more severe than at more relaxed times. Depending on the domain, it might be often sufficient to consider a constant cost of communication for each communication action, that e.g. represents the typical delay introduced through communication.

The communication effort might depend on the amount of information transferred, the reliability of communication or also on the current situation, external state, or human engagement. In contrast to explicit communication actions, implicit communication may additionally effect the environmental state s, such as altered robot task behavior or demonstrative movements. For example, demonstrative robot movement can reveal information to the human at the cost of slower task progress (see e.g. [29]).

For the grid world example with explicit communication, a constant communication cost is considered for all communication actions $R_{\text{comm}}(s, a_c) = -1.5 \; \forall a_c \neq a_\emptyset$. Compared to human movements with a negative cooperative reward of -1, communication is beneficial if it saves more than one unnecessary human action, as it may introduce processing efforts and delays.

4.2 Communication planning for information sharing decisions

With the communication model from the previous section, the decision problem of the robot can be formalized as another Partially Observable Markov Decision Process (POMDP). This provides a principled approach to decide when and what type of information to communicate to support a human partner. This communication POMDP builds upon the base POMDP for human decision making as introduced in section 3.2.

For the robot, a main source of uncertainty consists in the latent human belief, as considered in the previous chapter. Additionally, the robot could be uncertain about environmental state s. The state of the communication POMDP s_R combines the uncertain aspects, $s_R = \begin{pmatrix} s & \mathbf{b} & a_H \end{pmatrix}$, where the human action is required to fulfill the Markov property. The robot receives an observation o_R that it uses to estimate the communication state s_R.

The full robot state transition combines the transition model for human belief, with environmental state transition function and communication effects. The external state transition may depend on both agents' actions, $p(s_{k+1} \mid s_k, a_H, a_R) = T(s_k, a_H, a_R, s_{k+1})$. The model of human belief transition, eq. (3.3), is extended by the effects of robot communication actions according to eq. (4.1). Lastly, human decision making is modeled according to the policy, eq. (3.4). These result in the robot state transition function T_R as

$$\begin{aligned} T_R(s_{R,k-1}, a_{R,k-1}, s_{R,k}) =& p(s_k, \mathbf{b}_k, a_{H,k} \mid s_{k-1}, \mathbf{b}_{k-1}, a_{H,k-1}, a_{R,k-1}) \\ =& \underbrace{p(s_k \mid s_{k-1}, a_{H,k-1}, a_{R,k-1})}_{\text{state transition } T} \cdot \\ & \underbrace{p(\mathbf{b}_k \mid s_k, a_{H,k-1}, a_{R,k-1}\mathbf{b}_{k-1})}_{\text{belief prediction and communication}} \cdot \\ & \underbrace{p(a_{H,k} \mid \mathbf{b}_{k-1})}_{\text{human decision}} \end{aligned} \tag{4.3}$$

The fully cooperative reward function for the robot decision process considers the joint task reward while also accounting for communication efforts,

$$R_R(s_k, a_{H,k}, a_{R,k}) = R(s_k, a_{H,k}, a_{R,k}) + R_{\text{comm}}(a_{R,k}, s_k)$$

With a focus on human behavior and support, in the latter examples it is assumed, that the robot has full access to the current human action

$a_{H,k}$ and true environmental state s. Consequently, the robot observation deterministically becomes

$$o_{R,k} = \begin{pmatrix} s_k & a_{H,k-1} \end{pmatrix}^T .$$

With access to environmental state and human action, it remains sufficient for the robot to hold a belief over human belief and the configuration represents a mixed observability setting [75].

The inference of human belief based on its observations represents the first step for solving the POMDP. The result is used for planning under uncertainty to compute robot action values Q_R to select its best action, as discussed in section 2.2.2. Planning with the robot belief requires high computation efforts. First, it must plan in its continuous state space, as it contains the human belief. Second, each robot state transition, eq. (4.3), includes the evaluation of human policy, which itself is computed by planning in the human POMDP, leading to a hierarchical evaluation. More efficient approximation schemes, that account for the special POMDP structure, e.g. by directly estimating the influence of communication on the human decisions via gradients, remain for future work. For the grid world example, the approximate POMDP solver uses, as for the human model, depth 2 planning with MDP leaf evaluations and 30 belief samples.

4.3 Illustration

The developed assistive communication concept is applied to and discussed for the previously introduced grid world example, to demonstrate the resulting behavior and show principle benefits compared to other communication concepts. Therefore, characteristic situations will be chosen to illustrate the decisions about what type of information should be communicated when. For the belief inference, the linearization-based inference approach is used, providing the robot belief for communication planning.

4.3.1 Alternative communication concepts

This new human centric communication (ToM-Com) concept is compared to a simpler state of the art communication concept and a variation using a theory of mind without communication planning. As discussed in section 2.4, typical assistance concepts provide information when a deviation from an expected behavior is detected (e.g. [44], [63]). This first baseline concept using detected human deviations (Dev) can trigger a warning, propose the next optimal action, or communicate all available information (for all state aspects). Warning or proposing good actions may help the human at the current time, but not for future decisions, as it does not support her situation understanding. Communication based on deviations from expected behavior does further require a threshold or definition of which actions are considered as deviating (especially for continuous actions).

As second concept for comparisons, it is proposed to reliy only on the artificial theory of mind (ToM) presented in chapter 3 without the evaluation of situation relevance. When a false or uncertain human belief is detected, e.g. a deviation of expected human belief to true state, the robot could directly decide to share information related to the state aspect. Compared to the full ToM-Com concept, this does not use the POMDP planning for evaluating the relevance of information. Instead, the robot communicates whenever an uncertain or false belief is detected, according to a threshold, to correct it. Since this approach does not consider task relevance of information, it may produce irrelevant communication and avoidable interruptions with the risk to annoy and distract the human.

4.3.2 What to communicate

Regarding the decision "what" to communicate, the scenario from the last chapter is considered again. From the start position (Figure 3.1 left, repeated in Figure 4.1) the human agent should reach her goal g_1. She correctly beliefs her goal to be g_1, but falsely beliefs the door to be open. Regarding her position, she is uncertain between possible start positions in $\{(0,1,\text{north}),(0,2,\text{north}),(2,1,\text{south}),(2,2,\text{south})\}$. The robot initially has a uniform belief over possible human beliefs. As first action, the human turns right since she is not aware of her current position. With the resulting observation, her belief shifts towards a position in $(0,1)$ or $(0,2)$. Due to the false door belief, the human tries to take the short path to g_1 passing the door cell, although this is actually not accessible (this would lead to a long detour of gathering door state information, as in Figure 3.1).

The second human action reveals this false door belief and her missing situation awareness (Figure 3.8). Robot's planning in the communication POMDP evaluates the importance of the door belief and its influence on future human behavior. Consequently, the best robot action is to communicate door information in the second time step (see Figure 4.2 for the robot's expected action values). This information sharing helps the human to become situation aware (Figure 4.3 right) and she changes her movement to the available path through the right passage, reaching her goal much faster (Figure 4.1).

In this scenario, the question what to communicate is of main interest. The deviation-based method could also detect a deviation at $k = 2$, as the human turns towards the closed door in the second time step (deviation from optimal path). However, generally warning the human or telling the human what to do will not solve her problem. She might recognize a problem of herself but could not distinguish problems of door belief or goal location belief. Instead, it would require further communication at later time steps. In contrast, ToM-Com (and ToM) interpret the available information to detect the probable error cause. This is used to precisely communicate the required information according to estimated human needs. Specific communication of door state requires less effort than communication of the full state with all state aspects.

The concept ToM, without relevance evaluation, would also detect an uncertain human position belief (see Figure 3.6 for $k = 2$) and, besides the relevant door state, further share position information. However, in this situation exact position knowledge is not relevant for the human behavior and communication represents an avoidable interruption (compare action values in Figure 4.2).

4.3.3 When to (not) communicate

Besides decisions what to communicate, the information sharing approach also robustly addresses the question when to communicate. There are situations, where a human might deviate from optimal behavior without a need for support, or, on the other hand, dangerous situations, where proactive intervention can be useful due to a high risk. There are even situations, where communication can have a negative impact on the joint reward. In the following, characteristic situations are discussed to illustrate these effects.

Human information gathering

As the human faces environmental uncertainty, she needs to trade off information gathering and task progress (a general problem in POMDPs, section 2.2.2). Consequently, information gathering actions can be a reason for a human to deviate from the optimal path given full knowledge. For such a situation, the human starts in position and orientation $(0, 1, \text{west})$, knowing her position and orientation, and goal location g_1 (Figure 4.4 left). She holds a uniform belief over door state. In this situation it is

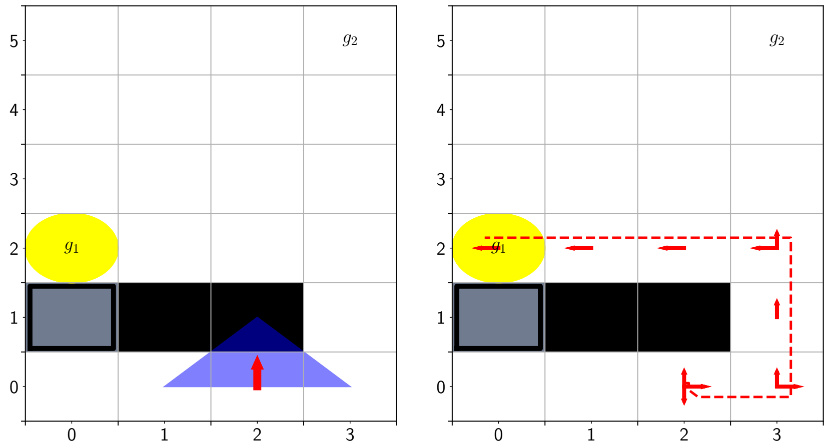

Figure 4.1: Scenario "what to communicate". Starting state (left), human agent falsely believes the door to be open. With communication support by robot (ToM-Com), the human can recover optimal behavior after the second time step and reaches her goal faster than without intervention.

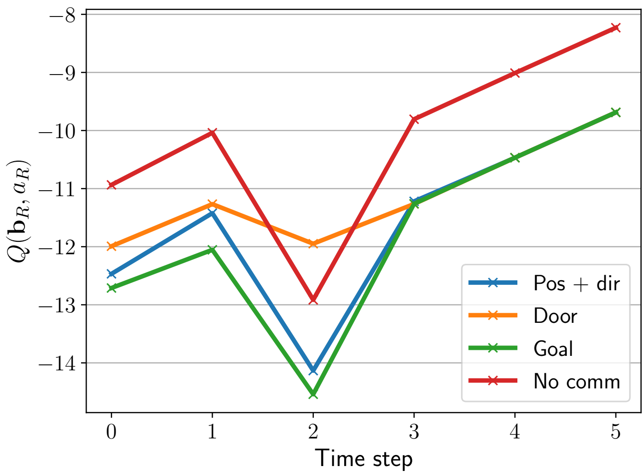

Figure 4.2: Expected action values for the scenario what to communicate, with false door belief. After the second human action, her false belief can be detected, and communication action a_{door} for door information is expected beneficial.

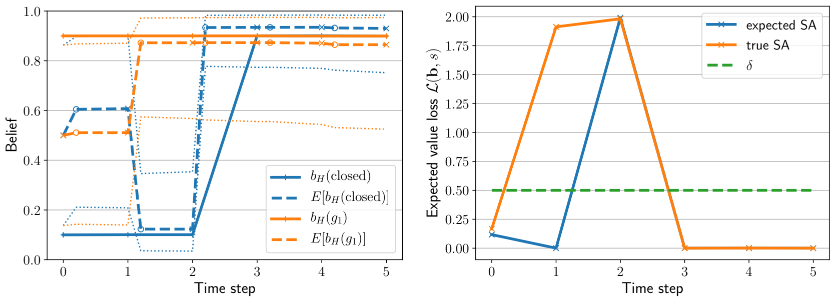

Figure 4.3: Belief inference (left) and situation awareness estimate (right) for the scenario what to communicate. After the second action, false human door belief is inferred. Supported by the information sharing action of the robot, she can correct her door belief and achieves situation awareness.

cheap to gather door information as it only requires one action to move forward and perceive the door state, which she does in the first time step.

Afterwards, she gains information regarding true door state (Figure 4.4 top right) and communication is not needed anymore (see action values, Figure 4.4 bottom right). Still, a deviation-based concept would intervene and disturb the human, as she deviated from the optimal behavior, while ToM and ToM-Com would stay silent. Neither warning, proposing next actions or communicating all state information would help the human, as she is already situation aware.

Human action execution noise and dangerous situations

The concept ToM-Com intrinsically accounts for noise in human decision making or action execution, as it is included in the human model, eq. (3.4), and respected during inference. A deviation-based approach instead needs to define a (fixed) threshold to classify observed actions as deviations to trigger communication (in the discrete grid world, the threshold can be zero).

In contrast, ToM-Com reasons about the causes of human decisions, especially if they are unexpected. If the observed behavior can be explained by a false human belief, this is a likely cause which can be supported by communication. In other situations, where no reasonable explanation exists, it will be considered as noise (see e.g. Figure 3.12). Accordingly, ToM-Com does not use a fixed threshold, but rather adapts communica-

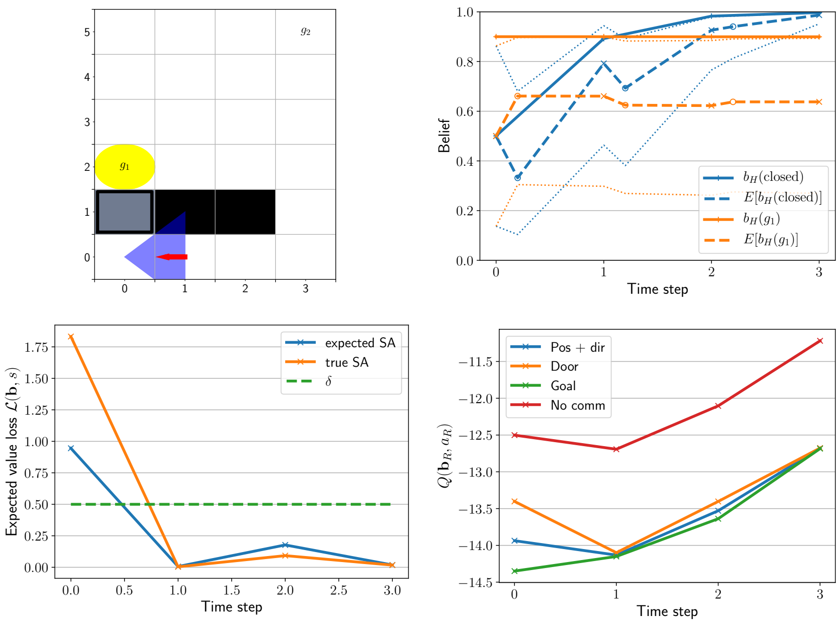

Figure 4.4: Human information gathering scenario. Human start position (top left) makes information gathering of the door state relative cheap. Estimated belief (top right), situation awareness (bottom left) and communication action values (bottom right) are shown for the first time steps. Though the human deviates from the optimal path (by moving forward), it is not beneficial to share any information since she already gathers it herself.

tion decisions to the current situation and expected human needs.

If a potential human problem is however not represented in the task model (e.g. human goal belief is another cell g_3), ToM-Com would probably interpret observed human behavior as noise and could not support the human (the robot would not be able to detect the causes and stay passive). In such corner cases, it is in principle not possible to accurately support another agent and ToM-Com avoids disturbances as it tolerates unknown human behaviors. In close cooperation, it probably would be useful to tell the human about the fact that the robot is uncertain as it does not understand her behavior (promoting transparency).

The respect of human and situation is further helpful for dangerous situations, where a bad human action might occur with a large negative outcome. It will make sense to proactively communicate before deviations are detected due to the related risk. This is the case when the human starts in the configuration in Figure 4.5 left. The true goal is at g_1, and her position is close to the branching point where paths to the goals split. Independent of a true human belief for the goal position, an uncertain robot using ToM-Com will proactively inform the human about the true goal location (see Figure 4.5 bottom right).

This evaluation of task relevance is not done in neither of the comparison concepts, which cannot recognize the high risk and the corresponding benefits of proactive assistance.

Concluding, the human centric communication concept of ToM-Com can cover a wide range of assistive situations to flexibly evaluate the questions when and what types of information should be shared to support the human. This cannot be achieved by simpler concepts such as the deviation-based approaches or human belief approaches without task relevance evaluation.

Negative effects of communication

In some special situations, communication can not only be irrelevant but even have a negative effect on the joint behavior. Although the robot could know about a false human belief, it could decide to not communicate, when this information is not relevant, or it could lead to a worse human behavior.

For such scenario, the human starts in location $(2, 0, \text{north})$ (Figure 4.1 left). She falsely believes the door to be open (which is closed) and the goal at g_2. Let's assume that the robot knows the true human belief. If the robot would inform the human about the true goal location g_1 immediately, she would try to get there via the shorter path through the (actually

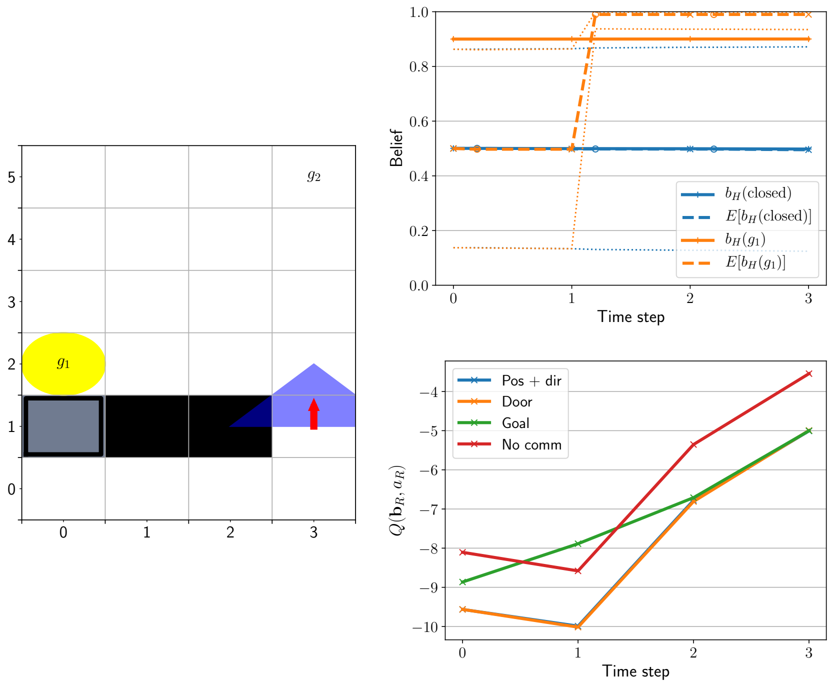

Figure 4.5: Branching situation with high risk. Human start position (left) is close to the branching point between the two goals. Estimated belief (top right) and communication action values (bottom right) are shown for the first time steps. Since the robot does not know the true human goal belief, it is better to proactively communicate, before a possible human deviation can happen. If she would move forward, three additional steps were required (compared to the cost of communication of 1.5) representing a high risk.

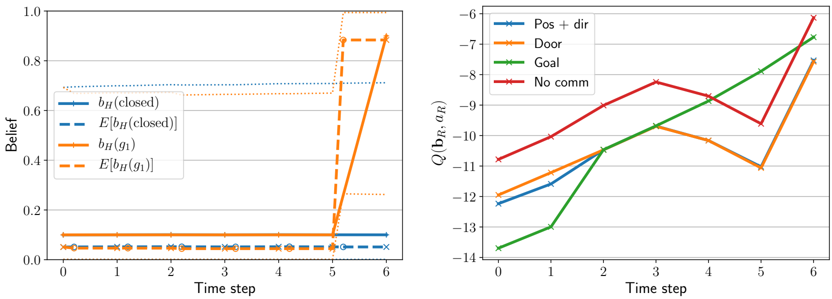

Figure 4.6: Situation with negative effects of communication, start position as in Figure 4.1. When the human falsely beliefs that the door is open and her goal is g_1 (left), communicating the true goal in the first two time steps would lead to worse human behavior (right), as she would move to the door. Later, in time step 5, it becomes important to tell the human about true goal position.

closed) door. Consequently, the robot should wait a few time steps and communicate later, e.g. just before the branching point is reached. Earlier communication would correct a false human belief but lead to worse joint reward, which is visible in the action values, Figure 4.6 right. In contrast, the concept ToM without relevance evaluation would not be able to account for such effects.

Going beyond – Strategic reasoning and white lies

Including strategic interaction into ToM-Com, as sketched in section 4.1, would allow for another solution of the previous situation. In the last scenario, the human is uncertain about door state and holds a false goal belief. Besides waiting until the human needs the goal information, the robot could use indirect strategic communication and only share information of the closed door. If she understands and considers the intention of the robot (trusts its capabilities and intentions) she can conclude that the door state is relevant for the joint task and consequently that her goal must be g_1. Hence strategic communication effects can transport more information than included in the message itself, while the receiver needs to consider the intention of the sender (the robot in this case). Especially in repeated interaction with a cooperative robot, the human might get used to the robot's communication and develop strategic interpretations. Still, strategic reasoning, as another recursion level, increases the complexity and processing load for both agents and imposes the risk of misunderstandings.

The decision to not tell the human about her false belief (as it is beneficial for her to stay ignorant), can be seen as controversial silent lie, as the robot intentionally leaves her in ignorance. In this case, the robot's behavior is not transparent for the human, and it might further effect her trust.

4.4 Summary and conclusion

In this chapter, theory of mind based communication, as principled concept for assistive communication, was presented to support the human with information according to her needs. It integrates both, the novelty of information for the human receiver as well as its relevance for the current situation and task. The decision problem was formulated as POMDP to systematically balance uncertainty, cost of communication and expected benefits for the joint behavior. It was based on the inference of human belief and a communication model, specifying available communication actions, expected effects on human belief together with related efforts and costs. The robot behavior is computed by solving the POMDP with planning under uncertainty, yielding effective support for the human.

The approach was applied to and discussed for an illustrative example, clearly showing principled benefits compared to alternative communication strategies. It can support the human awareness through well informed decisions on when and what type of information to share. In contrast to deviation-based approaches that instruct the human what to do, ToM-Com addresses the causes of human behavior to enable her making good decisions herself.

The computational effort of hierarchical planning is still a challenge that need to be addressed in the future for online application in larger domains. This might be achieved by either using better offline approximations of POMDP value functions or by more efficient online planning, e.g. based on available gradient information.

The illustration used a general human and communication model demonstrating structural benefits of human centric communication concepts. To further account for individual human preferences regarding annoyance or desire for support, one could allow the human to select the communication costs herself. It could even depend on the current situation and workload, enabling dynamic alerts or hints as desired by Schutte

[100].

5 Human centric human–robot cooperation

In the last chapters, theoretical concepts and required methods were introduced to understand and support a human partner according to her needs. Based on an artificial theory of mind and a situation awareness evaluation, the relevance of different types of information was evaluated yielding decisions for communicative support. Principles and systematic benefits were demonstrated on simulated, illustrative examples. Simulation explicitly makes it possible to compare inference results with a ground truth of the simulated agent's beliefs. To achieve a robust robot interaction policy, a broad probabilistic human model is used, respecting action and perceptual uncertainties.

Regarding potential applications, it is important to test the approaches in real interaction with humans, since humans show diverse behaviors and problem solving strategies. For quantitative evaluations and comparison to baseline concepts, two user studies are examined regarding the two targets of estimating human situation awareness respectively supportive communication. Human belief inference and awareness estimation is applied to a sequential task, orienting on cooperative manufacturing settings. A study is conducted to test and evaluate the concept and methods for the inference of human belief and situation awareness, presented in the publication [15]. The designed task could generate interesting situations where the participants missed important pieces of information leading to a lack of situation awareness. With the proposed artificial theory of mind, the human belief could be inferred online and problems in awareness could be detected. A further quantitative evaluation based on the prediction of human actions shows better performance compared to baselines.

The insights from the first study were used to design a more complex human robot sushi making task for evaluating human centric theory of mind based communication (ToM-Com). The task was chosen and parameterized to challenge human participants and generate interesting situations where support can be evaluated. It was used for a second user study, to investigate human support by ToM-Com, presented in publication [17].

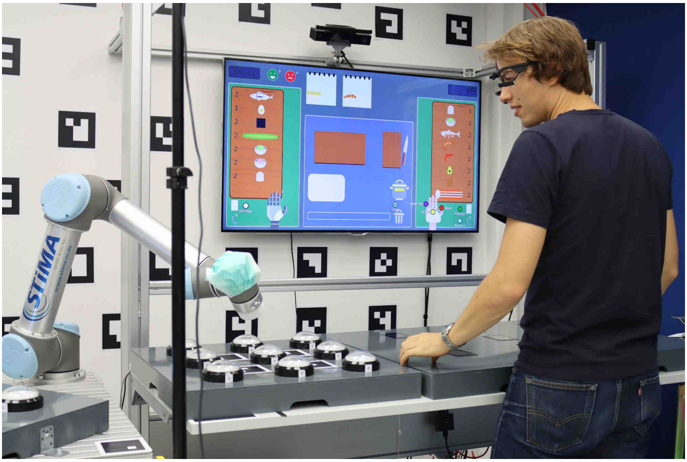

Figure 5.1: Human robot cooperative setup. Actions are mapped to discrete button presses on three boards: robot board on the left, human board on the right and shared workspace in the middle. A large screen on the wall is used to show the task and related information. Human eye gaze is measured by gaze tracking glasses.

Human–robot cooperative setup

The experimental setup consists of a human robot cooperative workspace, shown in Figure 5.1, where human and robot can act jointly to reach a common goal. A UR5 cooperative robot[1] is used, which is designed to work together with humans in a common workspace. The robot detects collisions which trigger an emergency stop. Together with reduced speed and weight, compared to classical industry robots, safe operation together with humans is possible. Additionally, emergency stop buttons were always in reach for participants and the operator.

Concentrating on cognitive human-robot interaction, typical robotic problems such as grasping and manipulation are not in the focus. Consequently, such manipulation actions are abstracted by button presses to allow for flexible design and deployment of different cooperative tasks. Each action of the agent is represented by pressing a physical button or touch display. Three boards are placed on the workspace, each equipped with 9 buttons respectively touch displays with similar usage. The buttons

[1] see `https://www.universal-robots.com/products/ur5-robot/`

are illuminated in different colors, which is used to distinguish them and map them to actions they represent. One board is placed so that it could only be used by the robot, one by the human, and one board placed in a shared workspace could be accessed by both agents.

Human actions are detected by the corresponding button presses (as the robot further knows its own actions). A screen on the wall is used to display task related information to the human, to visualize the task and to provide visual hints as communication actions. Regarding human information gathering, two variants are used in the different studies, the use of distinct information gathering actions (via some of the buttons) and an eye tracking device. Therefore, gaze tracking glasses from pupil labs[2] are used. They can measure the human pupils and calculate current gaze direction in relation to the glasses with infrared eye cameras. A world camera of the glasses serves for absolute localization of gaze. Aruco markers on the walls at known positions allowed to compute camera position and orientation in respect to screen and button boards.

5.1 Belief inference and situation awareness estimation in a sequential human–robot cooperative task

Inspired by cooperative manufacturing where robot and human jointly assemble pieces of a product, a sequential task is considered, where both agents contribute to reach a final configuration. In such a task, it is important, that both agents are aware of the current situation and the other's behavior. Lack of human awareness might arise from different underlying problems. The task structure can generate problems of human awareness regarding the plan to follow, the robot's current action, or regarding what has already been achieved. The following section is based on the author's publication [15].

5.1.1 Sequential cooperative task

The sequential task is designed to meet the following requirements. For human–robot cooperation, the task should enforce interdependence of the actors to generate interactive situations and interaction effects. To detect

[2]https://pupil-labs.com/

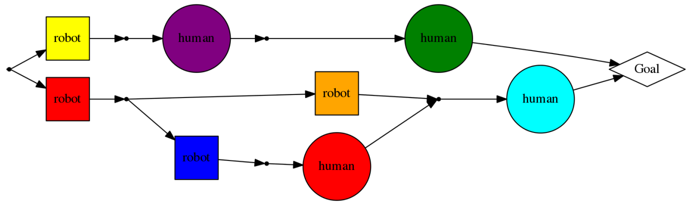

Figure 5.2: Task structure and visualization used in the user study. Small black circles represent states, squares (robot actions) and large circles (human actions) contain the color of the corresponding buttons for a state transition. Multiple paths lead to the goal state requiring different levels of involvement for the agents.

and evaluate situations, where a human partner is unaware, it is necessary to have a task that is complex enough to provoke human errors.

Similar to cooperative manufacturing, the task consists of specified sequences of actions to execute. There can be different paths to solve the task, to represent e.g. that the order of assembly might be interchangeable for some parts. Task structures are automatically generated and visualized by task graphs as in Figure 5.2. Colored nodes represent required actions of either robot (square) or human (large circle) to progress to the next node respectively state towards reaching the goal state[3]. At some states, the task branches and the agents can choose on which path to proceed. The structure shown in Figure 5.2 represents the task during a single random episode. This visualization is not available throughout the task, but the human may need to explicitly request for it (information gathering). Before the start of an episode, the graph is displayed on the screen for 3 seconds. This time span is designed to give the human the possibility extract some information to start with, while it is insufficient to memorize every aspect and plan behavior in detail. Instead, she will typically need to gather task information later by pressing the corresponding information gathering button.

Both actors may have multiple possible actions in a state, giving them the choice to decide on which path to proceed. This is e.g. the case for the robot in the first node in Figure 5.2. The robot can choose between the

[3]For this task, the shared action space is not used, and each action is directly mapped to one agent.

upper (via its yellow button) or lower path (red button). Consequently, the human needs to trace the robot's decision as it determines if and how she needs to react afterwards. These branching states are central for task complexity. The concept was introduced to encourage human planning needs (searching for shortest path) over purely reactive behavior, and to increase interdependence and interaction, as the human needs trace the robot's decisions in branching states. Executing any action that does not fit to the task (if there is no outgoing edge in the current state), is considered as error and leads to a negative reward. The state respectively node remains unchanged. The human participant does not receive direct feedback on errors, such that follow up errors can provide further hints regarding underlying error cause.

Information gathering For the sequential task, human information gathering is made explicit via discrete information gathering actions. Therefore, three buttons in the human space are reserved. While holding one of the gathering buttons, a corresponding type of information is shown on the screen. This can be information of the robot's current action "robot gathering", current task to accomplish "task gathering" or the actual state in the current task, including last and next required actions "state gathering". During robot gathering a visualization is shown for the robot approaching its next button[4]. This is useful in branching situations to distinguish between different possible paths in the task.

To represent gathering efforts, a time delay is used, until the requested information is displayed. The duration of time gathering delay depends on the type of information. Since the human should always be aware and track task progress, requesting task state information is delayed with largest waiting time of 1 second. Task gathering is delayed by half a second. In contrast, gathering of robot's movement is not delayed as it is considered as normal interaction and the human needs to dynamically adapt to its behavior. The participants were informed about the gathering costs and were further able to experience these during the habituation phase.

Robot behavior The robot has access to the task and current state. It starts to perform a valid action whenever there is one available. In

[4]To control the human perception of the robot's actions, the robot is simulated during the main part of the study. Without the use of robot gathering, the human only receives a sound indicating that the robot finished an action, similar to the noise of a button press.

cases with multiple opportunities, the action is selected at random, which was explicitly communicated to the participants. This forces the human to actively track the robot's decision as it cannot be inferred from other information. The robot needs 2 to 6 seconds to finish an action, from decision until the button is pressed, depending on button location. This time span is typically long enough for the human to gather information about it.

5.1.2 Application of belief inference

To apply belief inference and situation awareness estimation as introduced in chapter 3, it is necessary to formulate the task as a POMDP from the human perspective. Therefore, state set, action set, transition function, observation function and rewards are specified. For this study, an early version of the belief inference algorithm was used, where the human belief regarding a given aspect, b^j is parameterized by the most probable value together with a precision parameter, instead of using the full belief distribution [15].

Representing state and transition In this task, the human can be uncertain about task structure, progress state and the current robot action. These uncertain aspects form the environmental state of the human POMDP. The progress is specified by the number of the current state in the graph. The human belief representation further contains the current robot action, which is relevant for states, where the robot has multiple action opportunities.

The task is represented as graph defined by its adjacency matrix. The human task belief is split into connectivity of nodes and the correct actions for each transition. After the initial task gathering opportunity, the human may select an adequate path that she wants to proceed to reach the goal (connectivity). On this path, she further needs to know, which actions she has to do. The task parameters are chosen, that the human is normally not able to memorize all required actions in the beginning, but has to gather task information during the task.

While the task adjacency matrix is constant during one episode, task progress changes according to the selected actions and the robot action is subject to its stochastic policy. The transition function of valid actions is directly given by the task adjacency matrix, while errors do not alter the task progress.

Perception model The observation function is based on the explicit information gathering actions and information shown on the screen. For robot and state gathering, it is assumed, that the human does perceive all shown information. For state gathering, the last action as well as the next required actions are displayed to visualize the current progress. Additionally to the current state, the human can also perceive the next required action. For task gathering, the whole graph is shown, typically representing too much information for the human to perceive and memorize. Consequently, a memory parameter is introduced in the observation function regarding the number of actions, that the participant can remember.

Reward function and human policy Two performance measures are considered, the number of false button presses and time need to complete the task. Accordingly, the human has to balance between performing fast with the risk of making errors and performing safely with more frequent information gathering. She is told about both aspects without an explicit weighting (to limit complexity). At the end of an episode, feedback is provided in form of a score. Since total time of an episode significantly depends on the slow robot action execution, the time needed by robot actions is not counted, to gain better comparability between episodes. Error count c_{error} and human time needs t_H are combined to yield the final outcome $J = -c_{\text{error}} - wt_H$, with a weight $w = 3/s$, meaning that one human error is considered equivalent to a delay of one third of a second. To infer the approximated human belief, the action value function Q is needed. For the sequential task and the belief representation, it was specified by hand, respecting the risk of failure for task actions (if she is uncertain about task or state) and the effects and costs of information gathering [15].

5.1.3 User study

In this study, 9 people participated, well-educated with mostly technical background. It served for application of belief inference in the interaction with human participants.

After consent and instructions, multiple episodes of the sequential task were performed. In a habituation phase, the participants had the chance to experience the overall task structure, interfaces and get used to the gathering buttons and abstractions. After the completion of five training episodes, in the main phase 20 episodes should be completed as fast as

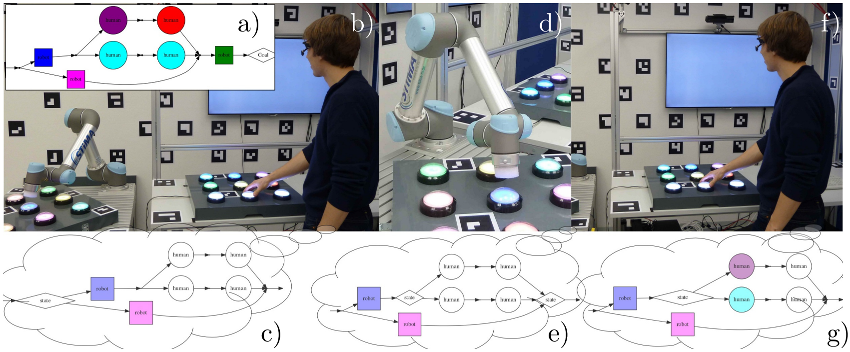

Figure 5.3: Illustration of a recorded example, where the human is unaware due to a missed robot action (reconstructed). Task a); three consecutive button presses b), d) and f); and human belief estimates after respective actions, c), e) and g).

possible while avoiding false actions. Depending on the participant's speed, this took about 30 minutes.

False buttons were selected 21 times by the participants, that is 5% of all task actions. The execution time varied strongly between different episodes and participants (mean 2.2 s, std 1.8 s). The task seems to be complex enough to challenge the human, since errors and situations with slow task progress occurred. It generates situations, in which the participant is unaware of the true required action, although it is not always possible to distinguish awareness related from action execution errors.

On average, participants performed 2.4 task actions and 3 gathering actions per episode. The participants developed different strategies in using task gathering versus state gathering. Indeed, there is information overlap, because gathering the state also provides the information over the immediate next action. However, the strategy of relying mainly on state gathering is not globally optimal, because it neither allows to choose the best path nor to prepare a sequence of consecutive human actions for faster progress. Instead, it is a simple strategy that seem to be a local optimum in the human learning process. This strategy was observed mainly for participants 2, 3 and 4 and partly for 5.

Qualitative evaluation The principle of belief inference is demonstrated in Figure 5.3 for a recorded situation. The relevant part of the task structure is shown in 5.3 a). After an initial human action b), one of two possible robot actions is required a). In c), the result of the inference process after this action is visualized. The inferred state belief is visualized by the width of the diamond for potential human state beliefs. In c), the probability for the human state estimate is concentrated on the first state. It is further estimated that the human is aware of the next robot actions, but does not know the actions required afterwards (empty circles). In this state, the robot could select the dark blue or rose buttons. Since the robot randomly selects the action, the human should gather its movement to track the actual state transition. However, the human misses to gather the (simulated) robot's decision d). The inferred human belief e) contains relevant probability for multiple states, as these could follow from the unknown robot action. Additionally, a low human state certainty is estimated. Accordingly, it is expected, that the human collects state information which she needs to further progress the task. This is actually the next human action, g) and her belief updates to the current task state g), and similarly the belief regarding the next required action. The inference of human belief, based on the tracking of her action and information gathering behavior, was able to detect the concrete problems in human awareness. At the beginning, the human is not aware of the robot's current action. Consequently, she cannot know which state transition occurs and loses awareness for the current state.

Quantitative evaluation

In contrast to the simulated agent scenario before, there is no ground truth available to compare the estimated human belief. As indirect measure, the human belief estimate is used to predict the next human action which is compared to the actual selection. The action prediction is directly given by the human stochastic policy model, eq. 3.4, using the action values for the belief estimate. This prediction is compared to that of a human expert observer, as well as against a heuristic rule based approach without explicit belief representation. Prediction of human action is still not the target of belief inference, as the insights into human reasoning can be used to evaluate human problems and to support cooperation in other ways, including assistive information sharing.

For action prediction, the ratio of correctly predicted human actions is evaluated. The action with the highest predicted probability is compared to the actually observed human action. Additionally, the action likelihood, the probability with which the observed action was predicted, was

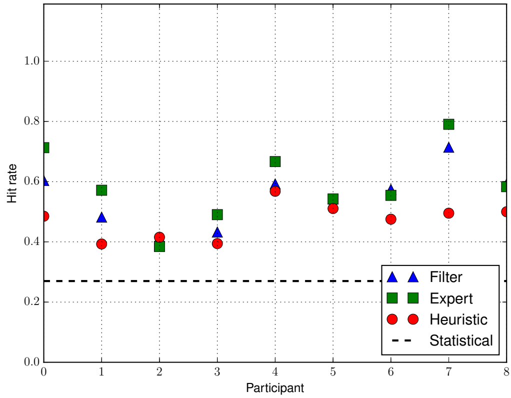

Figure 5.4: Action prediction for the evaluation of belief inference. Prediction results of belief inference filter lie between human expert rating and a heuristic.

evaluated. The results are very similar, why only firs hit rate is reported here.

The hit rate based on the belief inference filter is shown in Figure 5.4 for each participant (blue triangles). The average prediction rate is at 56%, varying between participants. Lower prediction accuracy for participants 2 and 3 may arise from the suboptimal strategy that they have used.

A statistical baseline (dotted line) results from always predicting the most frequent human action, which is task gathering with 26%. The prediction result is further compared against a human expert (the author) predicting participants' actions using the same available information of action history and task information (green squares). The expert's performance is slightly better with an average accuracy of 59% while similarly varying for the different participants. For action prediction, a simpler heuristic based approach is also introduced. It is specified by a few rules, predicting human taking task progressing actions alternating with gathering behavior according to the memory model, e.g. predicting task gathering in a fixed interval. This heuristic leads to a hit rate of 47% (red circles).

The inference of human belief could yield an action prediction with accuracy close to a human expert, while it is almost always better than the heuristic. However, it seems that human behavior is partly unpredictable in this task, as even a human expert has problems to correctly predict her actions in 40% of the cases. This is partly induced through the branching situations, where a participant could choose between two valid options. It is further the task structure that limits predictability. Human errors are hard to predict as there are often no hints or indications in her behavior in advance. In most situations, an error cause could only be found retrospectively, where it is too late to intervene. Prediction of errors were sometimes possible, when they followed in a sequence. However, also when one human error occurred, she may recognize it herself and assistive support is rarely necessary. In this task, neither human expert nor artificial theory of mind based prediction could significantly profit from their theory of mind abilities to predict human errors.

5.2 Assistive communication in a complex human robot cooperative sushi task

The discrete sequential task was designed and used to apply belief inference and estimate situation awareness in general. Regarding the second main contribution, to support the human based on the inferred information according to the concept of theory of mind based communication, the requirements for the task changed. A new task was developed to generate more interesting situations, where a human could profit from a robot's assistance, to evaluate the concept of theory of mind based communication. It is required that the task challenges the human to regularly generate situations, where she missed an important aspect, leading to unawareness and a need for support. However, the assistance concept is not limited to domains with frequent human problems, as it is not depending on training data. It can further be relevant to cover rare but severe events as it is the case for many types of warning or emergency support systems. The task should further generate complex situations with many uncertain aspects. Accordingly, depending on the cause or aspect that the human missed, different communication actions will be appropriate to support her decisions and the robot will need to reason what information to provide. As additional requirement, it is necessary that human problems can persist over a longer period, providing the robot the chance to detect errors and related causes, to address them in communication. For example, in the blind spot scenario from the introduction, an unaware driver might indicate and initiate a steering maneuver long before the situation becomes critical. These first, indicative actions only have small negative outcomes, but point to an underlying awareness problem.

Tasks in human robot literature

In human robot interaction literature, many tasks are designed from a robot perspective and provide little challenge for the human participant, as example the case for cooperative pick and place tasks, as in [74], [114]. The human can oversee the task (the current state and the required steps) and manipulation actions provides low challenges. Typically, the only uncertain aspect in the environment is introduced by the robot's behavior. This is interesting from a coordination or robotic planning perspective, but not regarding human support. Examples with multiple uncertain aspects that an agent needs to consider, are often grid world settings similar to the illustrative example in chapter 3, e.g. [7], [86], [85].

Other tasks challenge a human regarding one specific difficulty, as for example a lunar lander task which requires the human to hold a precise dynamics model for the environment [89]. The card game Hanabi represents a challenging cooperative task with high uncertainty. It is proposed as benchmark for agents to cooperate and coordinate with human partners [8]. The game restricts communication to the exchange of very small amounts of information. The challenge consists in finding strategic signaling strategies to efficiently share available information. In contrast to general support situations considered in this thesis, the relevance of information and the other's uncertainty is directly known from the game mechanism.

5.2.1 Sushi task description

As there was no task in literature according to these requirements, a cooperative sushi making task was designed. In this task, a human needs to assemble different types of sushi together with a robot partner to fulfill customer orders. The task visualization, Figure 5.5, is presented to the human on the large screen.

The task concept is similar to the video game Overcooked[5], as it challenges human situation awareness regarding multiple aspects and activities in a dynamic kitchen environment. The game Overcooked was independently proposed for the evaluation in human–robot settings due to its interactive and challenging structure [11].

[5] `https://www.team17.com/games/overcooked`

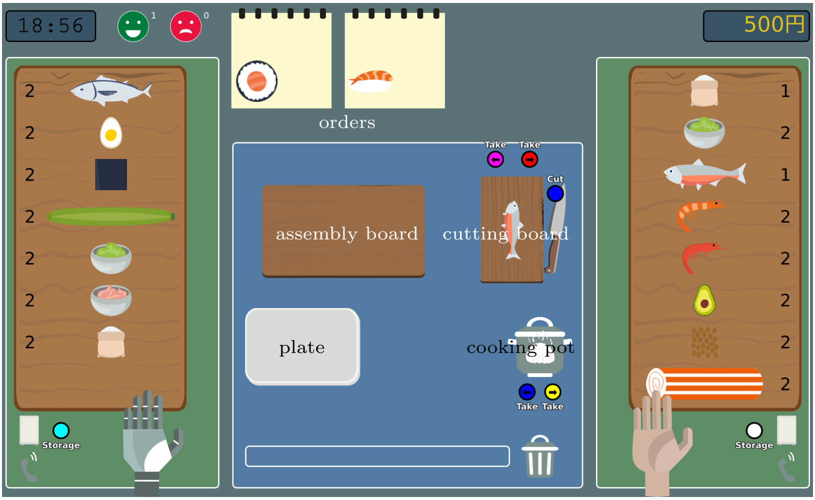

Figure 5.5: Sushi task visualization. The robot can take ingredients from the left, the human from the right storage. In the shared area in the middle, both agents can cooperatively work to create the final sushi. Current customer orders are presented at the top.

In the sushi task, human and robot each have access to a set of base ingredients (Figure 5.5, left for robot, right for human), from which some are available for both while some are exclusively to one side. These ingredients need to be processed and assembled to finally yield a sushi as ordered by the “customers”. Some ingredients need to be cooked at the cooking pot or cut at the cutting board, shaped by the human and finally stacked together on the assembly board. For each recipe, up to 4 different ingredients and different processing steps are necessary to successfully create the product. Sequences of in average 15 actions are necessary to serve a final sushi from the plate. In case of an error in the processing, there is the possibility to trash an item and free the corresponding location. This is necessary, as all locations, besides the assembly board, can only carry a single item. On the assembly board, the ingredients are stacked together to the final sushi. Due to these limitations, planning of the agents is required to coordinate the use of locations and which order to process first. Further, both agents need to contribute at some point, as ingredients from their private storage are required.

Using the cooperative setup with action abstraction (see Figure 5.1), all available actions are mapped to the button boards via colors, shown on the screen (Figure 5.5).

A set of 6 different sushi recipes is used leading to an adequate level of difficulty for the participants. During the experiment, up to two orders are sampled from this set. When one sushi is served on a plate, a new order is generated after a short delay. The orders do only show the final sushi visualization but not all required ingredients nor the detailed processing steps.

Task difficulties

The task is designed to challenge human participants in different aspects. One type of difficulty is induced by the visual design of ingredients, as similar looking fish or shellfish types. So, a human might take a different type of fish and process it, being unaware of the error. Depending on the current customer orders, the required ingredients and processing actions change, requiring flexible planning. The cooperation with the robot provides the chance to distribute work between the agents and progress faster. But it also induces challenges. Actions need to be coordinated on action level, as well as on a strategic level. Since only one sushi can be made on the assembly board at a time, it is necessary to coordinate which order to prepare first. Further, if the robot is taking rice as required by the current order, the human should not do the same. These location constraints, as well as the complexity of the recipes challenge human action planning.

In this task, the human needs to respect different aspects and face different types of difficulties. It is required to differentiate ingredient symbols, memorize complex recipes, plan action sequences while coordinating with the robot.

5.2.2 Human model and robot behavior

For the application of the human centric concept of ToM-Com, proposed in chapter 4, it is necessary to formulate the task as POMDP for the human, including state representation, transition and perception models. For a robot's supportive information sharing, communication options need to be specified that contain different types of useful information.

State aspects and state space

In the sushi task, the human can be uncertain about the content of different locations, the recipes for the different sushi types, and the current robot action. Consequently, the location contents are considered as first part of the state space and each location is represented by one state aspect (leading to one subbelief). For the assembly board (that can hold more items) there is one state aspect per item that it can hold. The robot hand is also considered as location, as it can hold an item that the robot will place or process somewhere. This belief aspect (robot hand content) is used instead of explicitly representing the current robot action.

Regarding recipe beliefs, a full representation of all possible recipe combinations with up to 4 ingredients would produce a very large state space, which would not be feasible. Instead, recipe belief errors are considered as modifications to a true recipe. For each recipe, there are a few typical confusions that were intended by design respectively observed in a prestudy. A human can miss on required ingredients (especially as not all are visible from the order), she can take additional ingredients, or exchange them, as there are similar looking fish types. Further she can confuse processing steps, e.g. forget to cook an ingredient. For every recipe, one state aspect is introduced which considers 4 respectively 5 recipe variants.

The 10 locations can each carry, depending on their use case, 8, 15, 21 or 37 different possible contents. Together with the 6 sushi recipes, the combinatorical state space has a size of $|S| = \prod_j |S_j| = 37^5 \cdot 21 \cdot 15 \cdot 8^2 \cdot 6 \cdot 5^4 \cdot 4^2 \approx 8.4 \cdot 10^{16}$. Due to factorization according to the state aspects, the human belief representation can be reduced to $N_b = \sum_j |S_j| = 271$ dimensions.

Transition and reward function

The transition function is directly specified by the task mechanism. While the recipe definition remains constant, the location contents change according to the agents' actions. Depending on the current state configuration, different action opportunities are available, hence the actions set depends on the current state, $A(s)$.

To favor fast and efficient task completion, a reward of $R = -1$ is used for every action besides serving a completed sushi. When the agents send out an order according to the customer's wish, a benefit of $R = 10$ is given, if the plate content differs, the customer is expected to be unsatisfied and a reward of $R = -10$ is used.

Model of human perception

To complete the POMDP model, the observation function needs to be specified. Visual perception represents the main source of information for participants in the task. It is used to gather information about the task displayed on the screen or by observing robot's behavior. Human eye gaze direction is measured by gaze tracking glasses and used to estimate her information gain.

The content of each location is visualized on the screen and can be perceived by human gaze. For each fixation, information provided in surrounding locations might be perceived. Therefore, a Gaussian is considered locally around the gaze point to describe a probability p_{Hp} of human state perception according to gaze distance,

$$p_{Hp} = p_{Hp0} \exp\left(-0.5 \frac{d^2}{\sigma_{Hd}^2}\right).$$

Here, d describes the distance (in pixel on the screen or in gaze angle) between gaze measurement and a task location center and σ_{Hd} a typical distance, representing the angle of foveal vision together with measurement uncertainties. The maximum observation probability p_{Hp0} describes an amount of information that can be perceived from a single gaze sample, respectively considers the time required to fully perceive some information. This spatial configuration is visualized in Figure 5.6 for a few gaze samples collected during one action execution. The accumulated Gaussian mask

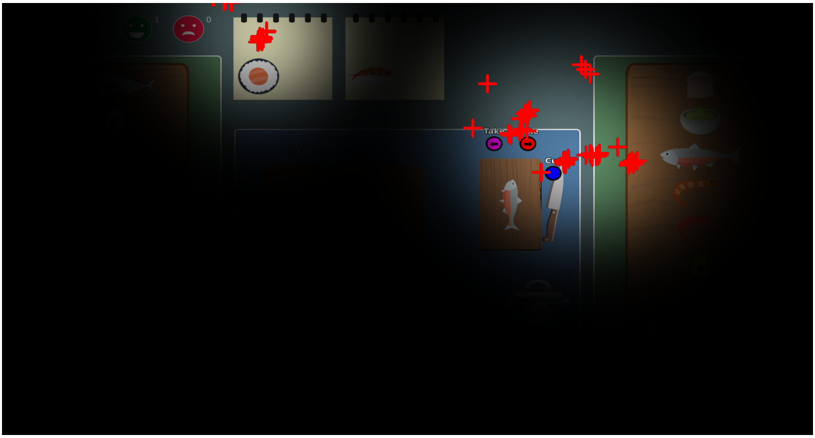

Figure 5.6: Visualization of human information gathering model. Measured eye gaze samples (red crosses) and Gaussian masks around. In this example, human might have perceived information about the first order and the cutting board content.

around gaze samples (red crosses), represents the areas of the screen that the human might have perceived.

For each location that the human might have perceived, a perception update is considered for the corresponding state aspect s^i according to the observation function

$$p(o^i \mid s^i) = \begin{cases} p_{Hp} + p_u & \text{if } o^i = s^i \\ p_u & \text{else} \end{cases},$$

with uncertainty $p_u = (1 - p_{Hp})/N_i$.

The maximum probability $p_{Hp0} = 0.03$ of perceiving some information is chosen according to gaze measurement sample time (120Hz) and typical fixation durations (around 330ms).

Action set as observation

In the sushi task, not all actions are available in every state, e.g. only if a location contains an item, the agents can move the item to their hands (take it). Further, a sushi assembly action is only available (assembling is only supported by the task mechanism), if the correct ingredients are

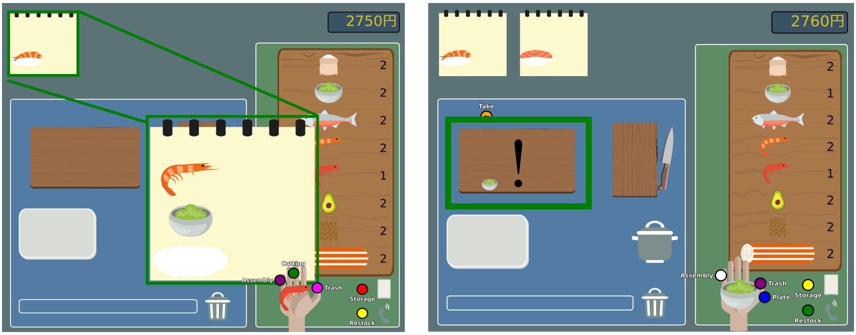

Figure 5.7: Communication signals: Display the current recipe (left) or direct attention to a specific location and corresponding content (right).

located on the board. This mechanism can be used by the human to update her recipe belief. Correspondingly, the available action set represents an additional observation for the human.

It is modeled, that the human perceives the actual action set with a probability of 99%, while the action set is generated according to the task mechanism.

Design of communication signals

To support human awareness according to ToM-Com, the robot needs to be equipped with a set of communication actions transmitting different types of information. These are realized as visual hints, displayed prominently on the screen. To cover typical human awareness problems, communication signals are designed to inform about the recipe of a current order, or direct the human focus to relevant locations. Regarding recipe communication, the required base ingredients are shown as in Figure 5.7 left. Although the content of locations on the screen are always visible for the human, it can be helpful to highlight a specific location, as the robot hand or the assembly board (Figure 5.7 right). These locations are especially relevant to coordinate behavior with the robot.

To illustrate a human awareness problem and its consequences, an example situation is shortly described. The human has access to two different shellfish types. If the human falsely believes that a recipe requires the second type, this false belief can start a longer sequence of bad human actions. The human might take the wrong shellfish, cut it on the cutting board and place it onto the assembly board. All the suboptimal actions of this sequence are caused by the false recipe belief. By displaying the true required ingredients as in Figure 5.7 left, the human can recover from the false belief. This need for communication can be estimated in the beginning of the sequence and communicative intervention has the potential to prevent most unnecessary actions of this sequence of errors.

Besides missing recipe knowledge, awareness problems can further result from coordination failures. If the robot started with one order while the human is working for another, the usage of locations might lead to a conflict and need to be resolved. Similar on action level, coordination conflicts appear when both agents want to do the same subtask, e.g. bringing the same ingredient (Fig. 5.7 right).

As for the grid world example, reliable communication ($p_{\text{comm}} = 0.99$) is modeled for the effects of these communication actions (eq. (4.2)). To respect negative effects of communication, such as distraction and delays of task progressing, a constant cost $R_{\text{comm}} = -1.5$ is used, corresponding to the effort of 1.5 additional task actions.

Action planning and robot task behavior

To solve the task and to serve the ordered sushi, several steps are necessary. A planning module searches for possible trajectories to reach the target configuration from the current state, evaluates the number of required actions and provides the set of best next actions for both agents. These results are used for the robot's task behavior. Further, for the human decision model POMDP planning is achieved by evaluating planning trees with a depth of two, as in the illustration 4.3. Leaf nodes are evaluated based on the number of further required actions.

The robot selects an action from the set of required actions, as computed by the planning module. If multiple actions are possible, it randomly chooses one. This means, it reacts to the human partner's past actions (which lead to the current state) without explicitly considering coordination. Implicitly, this results in a mixed initiative setting. When the human is passive and does not start preparation, the robot will choose some action to progress one of the customer orders. At some point, it will still require human engagement, as she has to contribute her private ingredients. If the human starts preparing one sushi type, the robot adapts as it selects actions best for the current situation. It still can happen that both agents act at the same time and block each other.

Planning communication

With the POMDP model for the sushi task, inference of human belief and communication planning can be applied as generally introduced in chapters 3 and 4. For this user study, the sampling-based particle filter approach was used for inference. To allow faster inference and planning, transition and observation functions were precomputed and cached. Still, belief inference and communication planning are not fast enough to allow online application in the used configuration. Especially planning in the human model takes time, which need to be repeated for inference as well as communication planning. Depending on the numbers of samples in the different stages of the algorithms, a computation time of about $8s$ (single core computation @ 3.70GHz, implemented in python) is achieved per observed action. In contrast, a typical human action lasts two seconds and can be much faster. In a prestudy, it was tested to limit action execution speed to a duration of 8 seconds by introducing a time delay for each action. However, this significantly changed the nature of the task, effecting human behavior, difficulty and performance. The participants had more time to evaluate the current situation leading to fewer and less predictable errors. Regarding the inference of human belief, the linearization based method was introduced in section 3.3.2 which significantly reduces the computation effort compared to the particle filter approach.

For this study, a "Wizard of Oz" setting is used, where the available communication actions are selected by a human expert (the author) in the assisted condition. The wizard followed the concept of theory of mind based communication with the same information available, meaning access to system state and gaze measurements. The wizard was trained in advance during training runs to detect and evaluate task situations fast enough for online intervention (even for a human it is hard to fulfill real time requirements). For evaluations of the concept, the similarity of decisions by wizard and offline decisions of the robot will be evaluated.

5.2.3 User study

A user study is done to evaluate benefits and opportunities of the developed human centric communication concept based on an artificial theory of mind, ToM-Com.

The target of the user study is to evaluate benefits and chances of the human centric assistive communication concept. Therefore, the following hypotheses are investigated: Communicative assistance based on a theory of mind (wizard) improves joint task performance (H1). The decisions of the robotic communication assistant (ToM-Com) are close a human expert (wizard) (H2). ToM-Com assistant supports a human partner respecting relevance and efficiency criteria, leading to fewer interruptions and higher acceptance, compared to alternative communication concepts (H3).

To account for the high variability in performance of different participants, a within subject design is used, where all participants experienced two conditions, one with and one without support by information signals.

For the study, 14 participants, well-educated and with mostly technical background, cooperated with the robot on the sushi task. Participants were randomly divided into two groups, one started with an assisted trial, the other started unassisted. The duration per participant was 1 hour and limited the total number of recorded actions. The participants had no prior experience with the task and started with instructions and a familiarization phase. After they got used to the task (they were able to successfully complete orders), two subsequent trials were recorded with both assistance conditions. After each trial, subjective impressions were collected regarding general task understanding including types of difficulty, and regarding effectiveness and acceptance of assistance. The developed questionnaire and the results are provided in annex A.

Human task performance

The task challenged the participants and generated situations where humans were not aware of important aspects. Out of 8515 recorded actions, 587 are considered as errors, as they are progressing the current orders. 481 of those were retrospectively classified by a human expert (author) to be caused by different types of belief related awareness problems, by also considering subsequent human actions (which is not possible in the situation itself). The remaining errors (e.i., seemingly random actions that were directly reverted by H) were probably caused by color mismatch or erroneous button presses (which corresponds to action execution noise in the human model). A false belief of one important aspect (as the current recipe) normally leads to a longer sequence of errors, e.g. where the participant works with wrong ingredients. Accordingly, the 481 belief related bad actions were clustered into 153 error sequences of actions with an expected common cause.

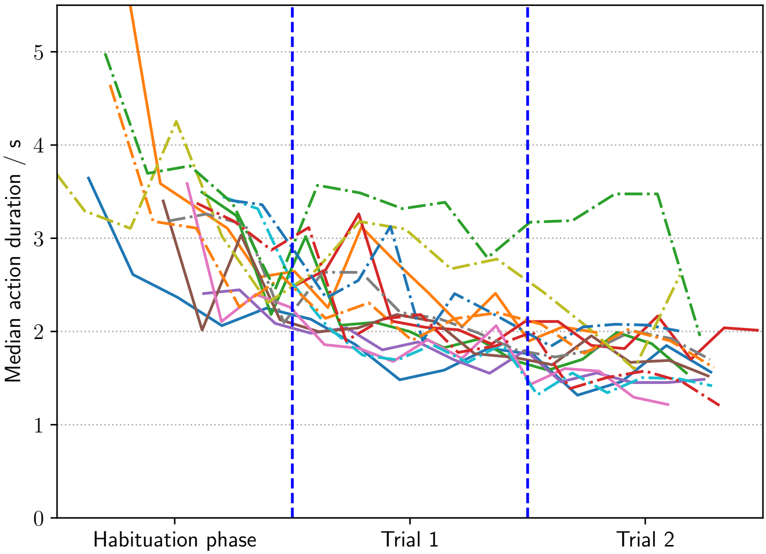

Figure 5.8: Action duration change during the experiment. The decrease in action duration shows the learning process of participants. During habituation phase, action duration drops significantly as the participants get used to the task, although it slightly decreases until the end of experiment.

Learning effect

Although the task is inspired by an everyday cooking scenario, the exact principles and mechanisms are new to the participants. Accordingly, in the first habituation phase, the participants could experience the task and get used to the mechanisms. This habituation respectively learning can be seen in the median action duration shown in Figure 5.8. The action duration drops significantly and the variance between participants decreases. Still, during the whole experiment, a learning effect could be observed. With more experience, the participants became faster. Learning represents an unwanted side effect, as it not only effects the human action execution speed, but also the number of errors respectively the performance. To balance this learning effect, it is important to split participants into the two groups with assistance conditions in differing orders.

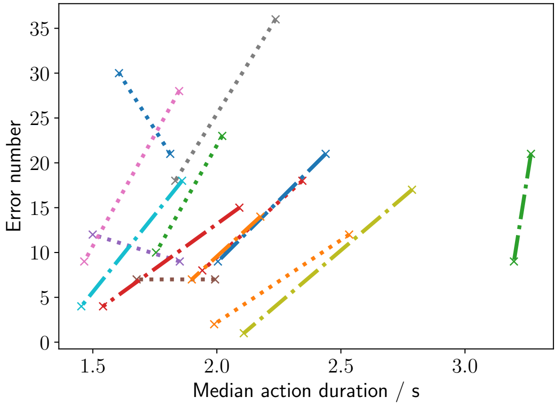

Figure 5.9: Variation in participant performance. For each participant (different color), median action duration and number of errors are shown for both experimental runs.

Diversity of human behaviors

The participants showed diverse behaviors and a high variance in performance. Some had problems to cope with the situation and to understand their task, while others seemed to interact intuitively with a low number of errors. The diversity can be seen in Figure 5.9, where error number and action duration are visualized. The large individual differences also make it hard to select an appropriate level of difficulty. Instead of a fixed task design, it could be useful to adapt difficulty according to the human performance as in different game levels.

Understanding communication signals

In the assisted condition, the human participant received support by communication signals triggered by the wizard. Most of the time, the participants could understand the received communication signals and recover from awareness problems. However, information sharing was not always successful and few participants had problems to extract the communicated information and update their beliefs accordingly. As consequence, not all error sequences could be solved by communication. Issues in understanding the communication were also reported in the questionnaires, as 4 par-

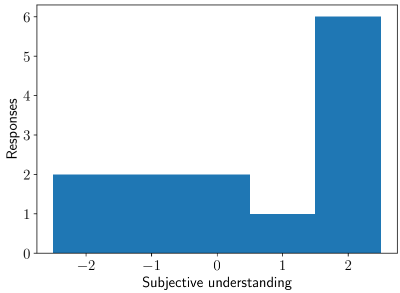

Figure 5.10: Subjective understanding of communication signals, reported by the participants after the trial. Responses to the statement "The signals were understandable" from -2 to 2.

ticipants mentioned that they did not fully understand the communication signals ("The signals were understandable", Figure 5.10).

Hence it seems, that communication was not always as reliable as expected, especially for some of the participants and not all awareness problems could be solved. Different designs of communication signals (which is not in the focus of this thesis) might be investigated to better support human understanding. Also prior explanations of communications and the information contained might help to improve participants' understanding.

5.2.4 Performance improvement by human centric assistance

Wizard similarity

Before investigating the performance improvement by communicative support, the similarity between the communication decisions of the wizard and the communication planning ToM-Com is considered (H2). The recorded data is played back and evaluated retrospectively to compare communication decisions. In general, the artificial assistant ToM-Com would communicate more often and more proactively. For 74% of the communication actions by the wizard, a similar decision is selected by the assistant, meaning the agent would communicate in nearby time steps.

To understand the differences better, characteristic example situations were analyzed, where the wizard communicated while the agent assistant would not. Within these, there are false positives of the wizard as well as situations, where the wizard reacted lately to an awareness problem, while the assistant would have intervened earlier. Still there are cases where the assistant does not intervene. These occurred often during short error sequences where it might have intervened later on, i.e., evidence accumulation was slower than for the wizard. Regarding the similarity hypothesis (H2), behaviors do not match exactly, but still show similar patterns in many situations. Consequently, it is expected that similar positive effects of communication could be achieved by the automated system. Besides performance effects, similarity to the advice of a human expert itself is desirable, as humans are good at and used to interaction with others. Approaching human capabilities in theory of mind and supportive communication will yield more intuitive interfaces for an efficient and natural interaction.

Number of errors

Hypothesis H1 states that joint performance is improved by supportive communication respecting estimated human error causes within a theory of mind. Therefore, the results in both conditions, with and without communication assistance, are compared. For this analysis, the assistance data from the human wizard assistance is used, for which a similar behavior could be found. As performance measures, the number of errors and the lengths of error sequences are considered. A human error is defined as a suboptimal action, which decreases the achievable collected reward. Similar results are obtained for other performance measures, such as the time needed for fulfilling orders or the number of actions exceeding optimal behavior.

The number of errors for each participant is shown in Figure 5.11 left for the two conditions with and without support. Besides the condition, also the learning effect influences the number of errors. Differences between the two groups are clearly visible. When the first run was assisted (solid lines), the participants' experience was higher in the second unassisted condition and the error number even decreased for some participants. Hence, only considering error numbers does not allow to distinguish effects of experience from those of assistance condition. The strong learning effect partially hides the influence of assistance on the number of errors. To distinguish the effects, experience gain is explicitly considered as independent effect. The experience of human participants is further directly related to the action duration, as visualized in Figure 5.8. The action duration for the different runs of the participants, Figure 5.12, in contrast shows no depen-

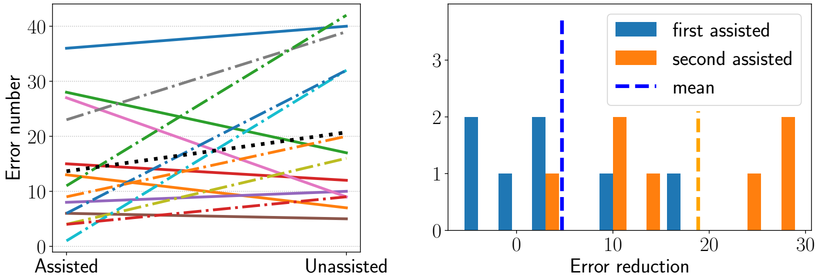

Figure 5.11: Performance for assistance conditions: Error number of participants (left), with first assisted group (solid lines), first unassisted group (dash-dotted), and mean (dotted). The difference of the runs (first run minus second run) is shown as histogram for both groups (right).

dence to the assistance condition. For another visualization, in Figure 5.12 right, the duration difference from the first to the second trial is shown, which is similar for both groups. This supports the assumption, that experience (measured by action execution duration) represents an effect that is independent of the assistance condition.

To separate both effects on the performance, Figure 5.11 right shows the difference of error counts from the first to the second trial, split according to the assistance condition. The experience effect reduces the error count for all participants independent of their group, while the assistance condition affects the groups in opposite directions. Accordingly, the clear difference in group means (dashed lines) shows the positive effect of assistance on the participants performance. The Pearson correlation factor between assistance condition and error reduction is calculated as $r = 0.64$ (with $p = 0.014$).

As statistical evaluation, a mixed analysis of variance (mixed ANOVA) is applied. The assistance condition represents the within-subjects variable while the group respectively the order of trials is the in-between-subject variable. The analysis shows a significant influence from the assistance aspect, with $p = 0.014$. The in–between effect of experience is also significant, with $p = 0.0004$.

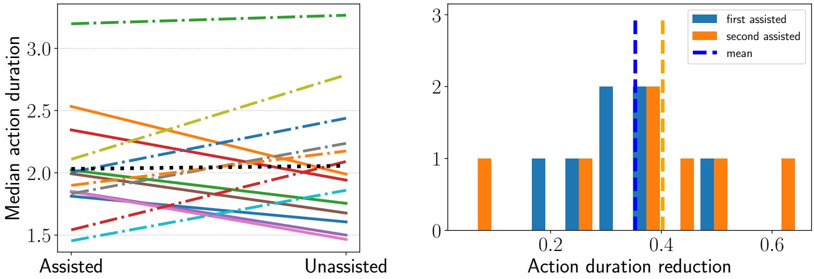

Figure 5.12: Assistance condition and experience effects on human action duration. The human action duration is split according to the participants groups (left). Taking the difference from first to second run (histogram on the right), the action duration is unaffected by the assistance condition.

Error sequence length

Human centric communication aims to support human awareness by sharing relevant information. An influence will consequently be visible, when looking at the sequences of errors, which are caused by the same underlying awareness problem. The positive effect is visible when looking at the error sequence lengths which is the number of errors in one sequence. Figure 5.13 shows the distributions of error sequence lengths for both conditions. Assisting the human with information reduces the length of human error sequences respectively the impact of a false belief situation. Compared to the unassisted case, she can recover much earlier from belief related problems. When sharing information adapted to her needs, most sequences end after one or two errors. This performance measure is not affected by the experience effect.

The sequence length distributions also provide two more insights. Even in the unassisted condition, 25% of the error sequences end after only one error where communication would not help, demonstrating the importance of relevance evaluation. The state-of-the-art deviation-based approach would always communicate after the detection of a human error even when the sequence would end nevertheless which happened in 25% of the cases. Considering relevance and the human needs instead safes unnecessary communications in many cases. As similar number, 20% of the error sequences of the assisted condition, the sequence length is one while there is no communication triggered. In Figure 5.14 the sequence lengths

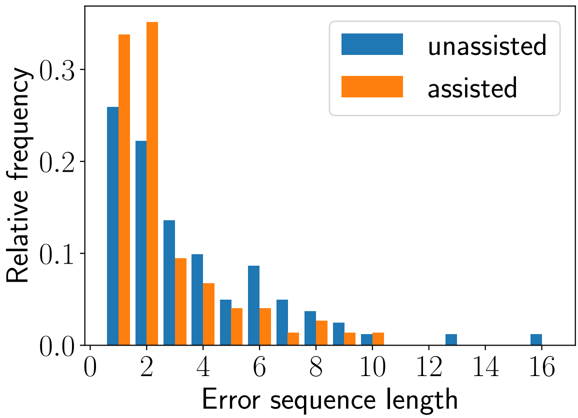

Figure 5.13: Relative frequency of error sequence lengths, the error numbers per sequence. Histograms are split according to the assistance condition. Communicative assistance reduces error sequence length and allows the human to recover earlier from false belief situations.

are shown according to the number of communication interventions. Assuming similar distributions for assisted and unassisted runs, it can be expected, that about 80% of the 25% sequences of length one are correctly classified as the human is aware of the situation, and communication is not necessary, saving 80% of potential interventions. The situation seems to be similar for the length 2 error sequences, although the numbers cannot be compared directly (due to the assistance effects on the distribution).

The second insight considers the success of communication. In some situations, the shared information could seemingly not achieve the intended effects, as there are still longer sequences in the assisted condition (Figure 5.13). Even repeated communication seems not to be helpful in some cases, with long error sequences despite 2-5 communication events (Figure 5.14). This can be partly explained by some participants' difficulty to understand the communication signals, as already discussed and reported.

Human acceptance

One central motivation for information sharing according to human needs is the acceptance by the human partner. This is expected to depend on

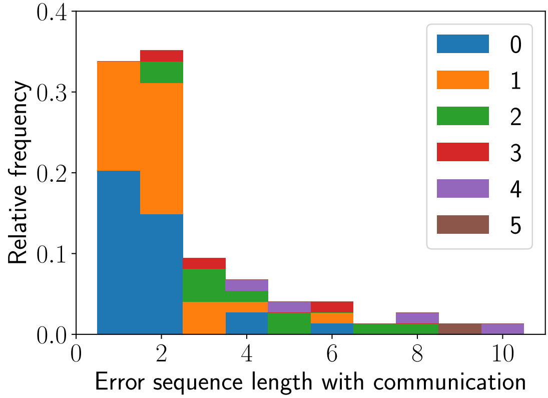

Figure 5.14: Relative frequency of error sequence lengths for the assistance condition. Histogram split according to the count of communication interruptions. Hence, in 20% of the sequences unnecessary communication is avoided, since there is no communication at a length of 1.

the efficiency of communication, as it should be limited to relevant aspects while avoiding unnecessary interruptions. The results from the questionnaire support a corresponding hypothesis. Combining the responses for the statements "The signals were given too often" and "The signals annoyed me", an acceptance measure is formed, shown in Figure 5.15. 7 participants reported high acceptance (they highly disagreed for both statements), while 6 seem to be uncertain. Lower acceptance may result from the difficulty of understanding communication signals. The subjective responses show correlations between the understanding of communication signals and acceptance (Pearson correlation $r = 0.53$, $p = 0.06$, Figure 5.16).

5.2.5 Comparison of information sharing concepts

In the last section, the influence of communication on the joint performance was considered, comparing results to the condition without assistance. It is further interesting to compare the communication decision to other communication concepts, as it was done for the grid world example in chapter 4. As in section 4.3, a state-of-the-art deviation-based concept

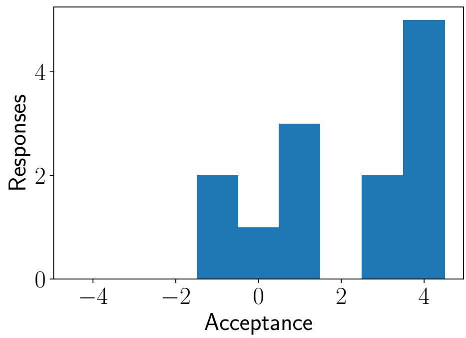

Figure 5.15: Participants acceptance as reported in the questionnaire. Combining the questions regarding annoyance and frequency of communication.

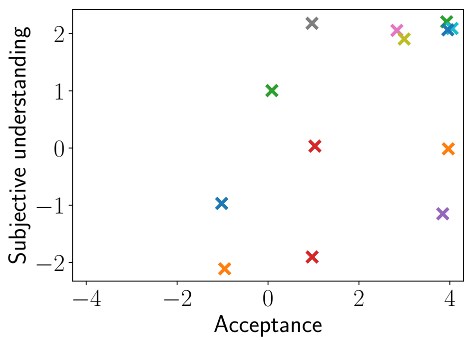

Figure 5.16: Responses regarding acceptance and understanding of communication signals by the participants. Issues in understanding may lead to lower acceptance, independently from the communication concept.

and a theory of mind concept without communication planning are considered. Comparison and evaluation is performed offline using the trials without assistance.

The deviation-based approach (DEV), as state-of-the-art concept, intervenes when the human deviates from the expected behavior, which is here considered as human error. Consequently it could warn the human or propose a good next action. Considering the action proposal, it can help the human to select a good next action, but will in general not support the human to understand the situation and the error she made. For a quantitative evaluation, it is modeled that participants would follow the action proposal and that the deviation-based concept would prevent a following human error. To improve the support, the robot can not only propose one next action but intervene for a longer sequence proposing k good next actions.

As a second approach, a theory of mind based alternative without communication planning (ToM) is considered. It communicates when a false or uncertain human belief is estimated. This policy should support the human and prevent errors similar to ToM-Com, but may lead to unnecessary communication, when a communicated false belief aspect is not relevant for current decisions. Here, a threshold needs to be selected to classify the inferred belief as false.

As analyzed in the last section, information sharing could achieve a significant reduction of error sequence length. Consequently, it is assumed, that providing information in a situation with a false human belief would prevent the subsequent errors in the current sequence (neglecting partial problems of understanding communication, which could be addressed by other means). This is modeled as effect of communication for both concepts using a theory of mind for information sharing (with and without communication planning).

The different concepts, ToM-Com, DEV and ToM, are compared regarding their potential in preventing human errors as well as the number of interruptions for the human (cost of communication). By varying the decision parameters of the concepts, a receiving operator characteristic (ROC) is created, drawing the true positive rates (potentially prevented errors divided by total errors) against the false positive rate (unnecessary interruptions divided by number of optimal human actions), shown in Figure 5.17. For ToM-Com, the cost of communication represents an explicit parameter to handle the false positives against false negatives. For the deviation-based communication the number k of next actions proposed is varied and for ToM, the decision threshold serves as variable.

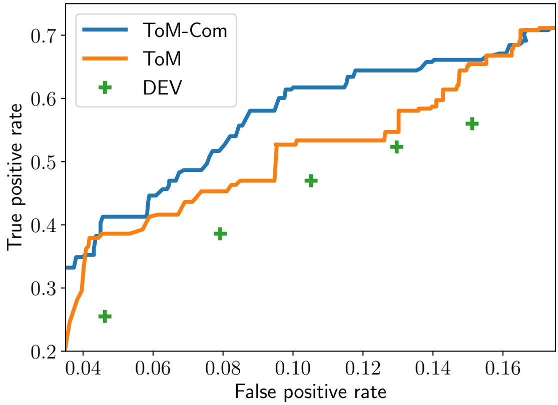

Figure 5.17: ROC curves for communication concepts. ToM-Com with communication costs between 0.5 and 2.4, a theory of mind approach without communication planning (ToM) with erroneous belief thresholds between 0.4 and 0.99, and a simple deviation-based approach (DEV) proposing the next 1 to 5 actions after a human error.

Comparing the different concepts, ToM-Com outperforms the other concepts. It can prevent more human errors while reducing unnecessary disturbances for the human partner. Communication with belief inference but without communication planning can achieve better results than the purely reactive deviation-based approach but would sometimes lead to avoidable interruptions.

5.3 Summary and conclusion

In this chapter, the results for two user studies were presented, considering a sequential, human robot cooperative manufacturing and a complex sushi making task. It could be demonstrated, that the developed theoretical concepts can be applied to true human interaction data. It is possible, to interpret human behavior to infer reasonable human beliefs and to detect false or uncertain configurations leading to a lack of situation awareness. A challenge that remains is the real time application, as related computations are extensive. Depending on task and sample time, an online application could not be achieved so far. A theory of mind provides a good basis for information sharing decisions to support a human partner. The human participants could profit from targeted communication signals and recover much faster from a lack of situation awareness. Retrospective concept comparisons could show large benefits of the concept of theory of mind based assistive communication compared to a state-of-the-art deviation-based approach and a concept without communication planning. It could support the human more efficiently, reducing unnecessary human interruptions.

6 Conclusion

For efficient and natural human robot interaction, a robot or technical system should not only concentrate on the task but also consider and understand its human partner. The topic of human centric human robot cooperation is framed from a conceptual side by considering research on human factors and human human interaction, including related challenges, effects and processes. To consider complex situations and information exchange, a formal consideration of uncertainty is important, as it is provided by the areas of Bayesian inference and reinforcement learning. Uncertainty is present in complex environments as well as introduced by other agents. Literature from human modeling in human robot interaction and existing communication concepts complement the background.

A human understanding as artificial theory of mind is essential for a complex cooperative robot. A new approach is introduced for inferring the human belief during interaction, as second order inference of what the human inferred of her environment. To handle the complexity of second order inference while allowing the online use during interaction with a human, approximations of belief representation as well as inference were developed and tested. Estimates of human belief provide insights into possible human problems and can be evaluated regarding her awareness for the current situation.

When it is inferred that a human partner missed an important piece of information, it will be a good idea to share such information with her. Accordingly, supportive information sharing is formalized in the concept of theory of mind based assistive communication. It represents a human centric communication concept deciding when and what type of information to provide. This decision is based on the evaluation of novelty for the human receiver together with task relevance in the current situation, yielding expected benefits of communication. Further, costs and efforts of communication need to be respected and balanced against benefits.

The developed concepts and methods of human belief inference, situation awareness estimation and communication planning were applied to interaction data, collected in user studies. It could be shown that they are capable to detect problems in human belief representations leading to human errors, and to improve the joint performance by human centric communication. Compared to alternative state of the art communication concepts, theory of mind based communication is more efficient and instead of instructing, it enables the human to become situation aware and make good decisions herself.

As challenge remains the computation effort, which is essential for real time application, due to the complexity of planning in large uncertain spaces. Regarding the inference of human belief, significant improvements were achieved by introduced methods. To similar improve communication planning, first ideas were given. Theoretical extensions of the concepts could further consider strategic effects of information sharing, that occur when the human also interprets robot's intentions. There, it is important to make the robot's decisions transparent to allow robust signaling and avoid misunderstandings. Besides for sharing relevant environmental information, theory of mind will be important for coordination of agents' behaviors. To support the process of coordination, information about the robot's actions and plans might be communicated to find a joint solution. For initiate such interaction strategies it may be further useful to also include coordination states (what does the human believe the robot wants to do) in the representation of human belief for the artificial theory of mind.

The concepts of belief inference as artificial theory of mind and communication planning are formulated in a general way independent of a specific scenario or domain. The application to different scenarios (the user studies and illustrations) required a specification of task and perception model, while the belief inference and intervention concepts and implementation remains the same. This demonstrates the flexibility of the approach opening a broad area of potential applications. The concepts and methods will enable complex human centric assistance approaches in diverse domains. Such intelligent assistance will not replace nor instruct the human what to do, but efficiently support a human specifically according to her needs.

A Questionnaire and results

During the second human robot interactive user study, subjective responses of the human participants were collected after each of the two trials. The developed questionnaire is shown in Figure A.1 and the responses are given in the tables A.1 and A.2.

Questionnaire: Assistance in collaborative tasks

Please select the answer, that best matches your experience in the last trial.

	Disagree	Slightly disagree	Neutral	Slightly agree	Agree
General task					
I understood the task	☐	☐	☐	☐	☐
I was capable to solve the task	☐	☐	☐	☐	☐
I had enough information to solve the task	☐	☐	☐	☐	☐
I often had to wait for the robot	☐	☐	☐	☐	☐
Task difficulty					
It was difficult to plan the actions for the current recipe	☐	☐	☐	☐	☐
I planned multiple orders in parallel	☐	☐	☐	☐	☐
It was difficult to respect location availability	☐	☐	☐	☐	☐
I looked for the robot‘s current action	☐	☐	☐	☐	☐
It was difficult to coordinate with the robot	☐	☐	☐	☐	☐
It was difficult to manage all subtasks in parallel	☐	☐	☐	☐	☐
Assistance with visual signals					
I noticed visual communication signals	Yes ☐	No ☐			
If yes:	Disagree	Slightly disagree	Neutral	Slightly agree	Agree
The signals were understandable	☐	☐	☐	☐	☐
The signals were helpful	☐	☐	☐	☐	☐
The signals were given too often	☐	☐	☐	☐	☐
The signals helped to understand the current situation	☐	☐	☐	☐	☐
The signals helped to chose the next action	☐	☐	☐	☐	☐
The signals distracted me	☐	☐	☐	☐	☐
The signals annoyed me	☐	☐	☐	☐	☐
The signals appeared when I missed some aspect of the situation	☐	☐	☐	☐	☐
The signals informed about an aspect that I missed	☐	☐	☐	☐	☐

Experiment ID: Trial:

Figure A.1: User study questionnaire

Participant	1		2		3		4		5		6		7	
Trial	1A	2B	1A	2B	1A	2B	1A	2B	1A	2B	1A	2B	1A	2B
Q1	1	1	1	2	2	2	-1	1	1	2	1	2	1	2
Q2	1	1	1	2	2	2	0	2	2	2	2	2	1	2
Q3	1	1	2	2	0	2	-1	1	2	2	1	0	2	2
Q4	1	1		0	-1		-2	1	0	1	-1	0	0	2
Q5	-1	-1	1	1	1	1	1	-1	1	1	0	-1	1	-2
Q6	-1	-1	-1	-2	-2	1	-2	0	-1	1	2	2	0	-1
Q7	-1	1	1	1	1	0	0	-1	1	1	-1	-1	2	-1
Q8	2	2	1	1	-1	-1	1	1	1	1	1	1	2	2
Q9	-1	-1	1	0	0	1	-1	-2	1	2	-1	0	1	1
Q10	-1	-1	2	2	2	2	1	0	1	1	-1	-1	1	0
Q11	1		1		1		1		1		0		1	
Q12	-1		-2		2		0		-1				2	
Q13	-1		-2		2		1		0				1	
Q14	0		0		-2		-1		-2				-1	
Q15	0		-1		2		0		0				2	
Q16	0		-2		2		1		-1				1	
Q17	2		1		-2		-1		-2				-1	
Q18	1		1		-2		0		-2				-2	
Q19	1		2		1		0		1				1	
Q20	-1		0		2		1		1				2	

Table A.1: Questionnaire results for the first group with assistance in the first trial.

Participant	8		9		10		11		12		13		14	
Trial	1B	2A	1B	2A	1B	2A	1B	2A	1B	2A	1B	2A	1B	2A
Q1	1	1	1	2	2	2	2	2	2	2	2	1	2	2
Q2	0	0	0	2	2	2	1	2	1	2	1	1	1	2
Q3	0	1	-1	1	1	2	-2	1	1	2	2	-1	-1	2
Q4	2	1	0	0	1	1	-1	-1	-1	-1	-2	-2	-1	1
Q5	2	1	0	-1	0	1	-1	1	-1	-1	1	1	1	-1
Q6	-1	0	2	2	2	2	-1	1	0	0	2	2	-1	1
Q7	-1	2	0	0	2	1	-1	-1	0	0	1	1	-1	1
Q8	2	2	1	2	2	2	1	1	1	-1	-2	1	0	1
Q9	-1	1	-1	-1	0	1	0	-1	-1	-1	1	1	-1	-1
Q10	0	1	1	-1	-1	0	0	1	-1	-1	1	-1	1	1
Q11		1		1		1		1		1		1		1
Q12		2		2		2		2		0		1		-2
Q13		1		2		2		2		0		2		-2
Q14		0		-1		-2		-2		-2		2		-2
Q15		2		2		2		2		1		2		-2
Q16		2		2		0		2		0		2		-2
Q17		0		-2		-2		-2		-2		-1		1
Q18		-1		-2		-2		-2		-2		-2		1
Q19		2		2		2		1		-1		1		2
Q20		2		2		2		2		-1		2		1

Table A.2: Questionnaire results for the second group with assistance in the second trial.

Bibliography

[1] M. J. Adams, Y. J. Tenney, and R. W. Pew. Situation awareness and the cognitive management of complex systems. *Human Factors*, 1995. ISSN 00187208. doi: 10.1518/001872095779049462.

[2] J. Aitchison and S. M. Shen. Logistic-normal distributions: Some properties and uses. *Biometrika*, 67(2):261–272, 1980.

[3] S. V. Albrecht and P. Stone. Autonomous agents modelling other agents: A comprehensive survey and open problems. *Artificial Intelligence*, 258:66–95, 2018.

[4] I. A. Apperly, D. J. Carroll, D. Samson, G. W. Humphreys, A. Qureshi, and G. Moffitt. Why are there limits on theory of mind use? evidence from adults' ability to follow instructions from an ignorant speaker. *Quarterly Journal of Experimental Psychology*, 63:1201–1217, 6 2010. ISSN 1747-0218. doi: 10.1080/17470210903281582.

[5] C. L. Baker and J. B. Tenenbaum. Modeling human plan recognition using bayesian theory of mind. *Plan, activity, and intent recognition: Theory and practice*, 7:177–204, 2014.

[6] C. L. Baker, R. Saxe, and J. B. Tenenbaum. Action understanding as inverse planning. *Cognition*, 113(3):329–349, 2009.

[7] C. L. Baker, J. Jara-Ettinger, R. Saxe, and J. B. Tenenbaum. Rational quantitative attribution of beliefs, desires and percepts in human mentalizing. *Nature Human Behaviour*, 1:0064, 2017. ISSN 2397-3374. doi: 10.1038/s41562-017-0064.

[8] N. Bard, J. N. Foerster, S. Chandar, N. Burch, M. Lanctot, H. F. Song, E. Parisotto, V. Dumoulin, S. Moitra, E. Hughes, et al. The hanabi challenge: A new frontier for ai research. *Artificial Intelligence*, 280:103216, 2020.

[9] S. Barrett, N. Agmon, N. Hazon, S. Kraus, and P. Stone. Communicating with unknown teammates. In *ECAI*, 2014.

[10] C. M. Bishop. *Pattern recognition and machine learning.* Springer Science+Business Media, LLC, 2009.

[11] J. Bishop, J. Burgess, C. Ramos, J. B. Driggs, T. Williams, C. C. Tossell, E. Phillips, T. H. Shaw, and E. J. Visser. Chaopt: A testbed for evaluating human-autonomy team collaboration using the video game overcooked!2. In *2020 Systems and Information Engineering Design Symposium (SIEDS)*, 4 2020. ISBN 9781728171456. doi: 10.1109/SIEDS49339.2020.9106686.

[12] A. Bobu, D. R. R. Scobee, J. F. Fisac, S. S. Sastry, and A. D. Dragan. Less is more: Rethinking probabilistic models of human behavior. In *2020 ACM/IEEE International Conference on Human-Robot Interaction*, 2020. doi: 10.1145/3319502.3374811.

[13] C. Boutilier. Sequential optimality and coordination in multiagent systems. In *IJCAI*, 1999. URL `http://dl.acm.org/citation.cfm?id=1624218.1624287`.

[14] X. Boyen and D. Koller. Tractable inference for complex stochastic processes. In *Fourteenth conference on Uncertainty in artificial intelligence*, pages 33–42, 1998.

[15] M. C. Buehler and T. H. Weisswange. Online inference of human belief for cooperative robots. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2018. ISBN 978-1-5386-8094-0. doi: 10.1109/IROS.2018.8594076.

[16] M. C. Buehler and T. H. Weisswange. Theory of mind based communication for human agent cooperation. In *International Conference on Human-Machine Systems*, 2020. ISBN 9781728158716. doi: 10.1109/ICHMS49158.2020.9209472.

[17] M. C. Buehler, J. Adamy, and T. H. Weisswange. Theory of mind based assistive communication in complex human robot cooperation. *arXiv preprint arXiv:2109.01355*, 2021. URL `https://arxiv.org/abs/2109.01355`.

[18] E. Cha, E. Meschke, T. Fong, and M. J. Matarić. A probabilistic approach to human-robot communication. In *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 6217–6222. IEEE, 2019.

[19] A. Chandrasekaran, D. Yadav, P. Chattopadhyay, V. Prabhu, and D. Parikh. It takes two to tango: Towards theory of ai's mind. *arXiv preprint arXiv:1704.00717*, 2017. URL `http://arxiv.org/abs/1704.00717`.

[20] J. Y. C. Chen and M. J. Barnes. Human–agent teaming for multirobot control: A review of human factors issues. *IEEE Transactions on Human-Machine Systems*, 44:13–29, 2 2014. ISSN 2168-2291. doi: 10.1109/THMS.2013.2293535.

[21] M. Chen, S. Nikolaidis, H. Soh, D. Hsu, and S. Srinivasa. Planning with trust for human-robot collaboration. In *2018 ACM/IEEE International Conference on Human-Robot Interaction - HRI '18*, pages 307–315. ACM Press, 2018. ISBN 9781450349536. doi: 10.1145/3171221.3171264.

[22] R. Chitnis, L. P. Kaelbling, and T. Lozano-Perez. Learning what information to give in partially observed domains. In *2nd Conference on Robot Learning*, pages 724–733, 10 2018. URL `http://proceedings.mlr.press/v87/chitnis18a.html`.

[23] R. Chitnis, L. P. Kaelbling, and T. Lozano-Pérez. Integrating human-provided information into belief state representation using dynamic factorization. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 3551–3558. IEEE, 2018.

[24] J. Choi and K.-E. Kim. Inverse reinforcement learning in partially observable environments. *Journal of Machine Learning Research*, 12:691–730, 2011.

[25] H. H. Clark. *Using language.* Cambridge university press, 1996.

[26] J. W. Crandall. When autonomous agents model other agents: An appeal for altered judgment coupled with mouths, ears, and a little more tape. *Artificial Intelligence*, 2020. ISSN 00043702. doi: 10.1016/j.artint.2019.103219.

[27] J. Daunizeau, H. E. Den Ouden, M. Pessiglione, S. J. Kiebel, K. E. Stephan, and K. J. Friston. Observing the observer (i): meta-bayesian models of learning and decision-making. *PloS one*, 5(12): e15554, 2010.

[28] A. Dini, C. Murko, S. Yahyanejad, U. Augsdörfer, M. Hofbaur, and L. Paletta. Measurement and prediction of situation awareness in human-robot interaction based on a framework of probabilistic attention. In *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 4354–4361. IEEE, 2017.

[29] A. Dragan and S. Srinivasa. Integrating human observer inferences into robot motion planning. *Autonomous Robots*, 37:351–368, 12 2014. ISSN 0929-5593. doi: 10.1007/s10514-014-9408-x.

[30] Y. Du, S. Tiomkin, E. Kiciman, D. Polani, P. Abbeel, and A. Dragan. Ave: Assistance via empowerment. *Advances in Neural Information Processing Systems*, 33, 2020.

[31] M. Endsley. Toward a theory of situation awareness in dynamic systems. human factors journal 37(1), 32-64. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 37:32–64, 1995. ISSN 0018-7208. doi: 10.1518/001872095779049543.

[32] M. R. Endsley, S. J. Selcon, T. D. Hardiman, and D. G. Croft. A comparative analysis of sagat and sart for evaluations of situation awareness. *Human Factors and Ergonomics Society Annual Meeting*, 42:82–86, 10 1998. ISSN 1541-9312. doi: 10.1177/154193129804200119.

[33] E. Ferreira, G. Milliez, F. Lefèvre, and R. Alami. Users' belief awareness in reinforcement learning-based situated human–robot dialogue management. In *Natural Language Dialog Systems and Intelligent Assistants*, pages 73–86. Springer, 2015. doi: 10.1007/978-3-319-19291-8.

[34] J. N. Foerster, F. Song, E. Hughes, N. Burch, I. Dunning, S. Whiteson, M. Botvinick, and M. Bowling. Bayesian action decoder for deep multi-agent reinforcement learning. *36th International Conference on Machine Learning*, 11 2019. URL `http://arxiv.org/abs/1811.01458`.

[35] C. D. Frith and U. Frith. Implicit and explicit processes in social cognition. *Neuron*, 60:503–510, 2008. ISSN 08966273. doi: 10.1016/j.neuron.2008.10.032.

[36] P. J. Gmytrasiewicz and P. Doshi. Interactive pomdps: Properties and preliminary results. In *International Conference on Autonomous Agents*, 2004.

[37] C. V. Goldman and S. Zilberstein. Optimizing information exchange in cooperative multi-agent systems. In *Second international joint conference on Autonomous agents and multiagent systems - AAMAS '03*, page 137. ACM Press, 2003. ISBN 1581136838. doi: 10.1145/860575.860598.

[38] O. Görür, B. S. Rosman, G. Hoffman, and S. Albayrak. Toward integrating theory of mind into adaptive decision-making of social robots to understand human intention. 2017. URL `http://researchspace.csir.co.za/dspace/handle/10204/9653`.

[39] D. Hadfield-Menell, S. Milli, P. Abbeel, S. J. Russell, and A. Dragan. Inverse reward design. In *Advances in neural information processing systems*, pages 6765–6774, 2017.

[40] F. G. Happé. Communicative competence and theory of mind in autism: A test of relevance theory. *Cognition*, 48:101–119, 8 1993. ISSN 0010-0277. doi: 10.1016/0010-0277(93)90026-R.

[41] M. Hassenzahl. Experience design: Technology for all the right reasons. *Synthesis lectures on human-centered informatics*, 3(1):1–95, 2010.

[42] H. He, J. Boyd-Graber, K. Kwok, and H. Daumé III. Opponent modeling in deep reinforcement learning. In *International conference on machine learning*, pages 1804–1813. PMLR, 2016.

[43] M. Herman, T. Gindele, J. Wagner, F. Schmitt, and W. Burgard. Inverse reinforcement learning with simultaneous estimation of rewards and dynamics. In *Artificial Intelligence and Statistics*, pages 102–110. PMLR, 2016.

[44] J. Hoey, C. Boutilier, P. Poupart, P. Olivier, A. Monk, and A. Mihailidis. People, sensors, decisions. *ACM Transactions on Interactive Intelligent Systems*, 2:1–36, 12 2012. ISSN 21606455. doi: 10.1145/2395123.2395125.

[45] C. Hubmann, J. Schulz, M. Becker, D. Althoff, and C. Stiller. Automated driving in uncertain environments: Planning with interaction and uncertain maneuver prediction. *IEEE Transactions on Intelligent Vehicles*, 3(1):5–17, 2018.

[46] C. P. Janssen, S. F. Donker, D. P. Brumby, and A. L. Kun. History and future of human-automation interaction. *International Journal of Human Computer Studies*, 131:99–107, 11 2019. ISSN 10959300. doi: 10.1016/j.ijhcs.2019.05.006.

[47] N. Jaques, A. Lazaridou, E. Hughes, C. Gulcehre, P. Ortega, D. Strouse, J. Z. Leibo, and N. De Freitas. Social influence as intrinsic motivation for multi-agent deep reinforcement learning. In *International Conference on Machine Learning*, pages 3040–3049. PMLR, 2019.

[48] J. Jara-Ettinger. Theory of mind as inverse reinforcement learning. *Current Opinion in Behavioral Sciences*, 29:105–110, 10 2019. ISSN 2352-1546. doi: 10.1016/J.COBEHA.2019.04.010.

[49] L. P. Kaelbling and T. Lozano-Pérez. Integrated task and motion planning in belief space. *The International Journal of Robotics Research*, 32:1194–1227, 8 2013. ISSN 0278-3649. doi: 10.1177/0278364913484072.

[50] L. P. Kaelbling, M. L. Littman, and A. R. Cassandra. Planning and acting in partially observable stochastic domains. *Artificial Intelligence*, 101:99–134, 1998. ISSN 0004-3702. doi: 10.1016/S0004-3702(98)00023-X.

[51] KNKT. Final knkt.18.10.35.04 aircraft accident investigation report, 2019. URL `http://knkt.dephub.go.id/knkt/ntsc_aviation/baru/2018%20-%20035%20-%20PK-LQP%20Final%20Report.pdf`.

[52] M. Krüger, T. Driessen, C. B. Wiebel-Herboth, J. C. de Winter, and H. Wersing. Feeling uncertain—effects of a vibrotactile belt that communicates vehicle sensor uncertainty. *Information*, 11(7): 353, 2020.

[53] P. Kulms and S. Kopp. A social cognition perspective on human–computer trust: The effect of perceived warmth and competence on trust in decision-making with computers. *Frontiers*

in Digital Humanities, 5:14, 6 2018. ISSN 2297-2668. doi: 10.3389/fdigh.2018.00014.

[54] C.-P. Lam and S. S. Sastry. A pomdp framework for human-in-the-loop system. In *53rd IEEE Conference on Decision and Control*, pages 6031–6036. IEEE, 12 2014. ISBN 978-1-4673-6090-6. doi: 10.1109/CDC.2014.7040333.

[55] J. Z. Leibo, V. Zambaldi, M. Lanctot, J. Marecki, and T. Graepel. Multi-agent reinforcement learning in sequential social dilemmas. In *16th Conference on Autonomous Agents and MultiAgent Systems*, pages 464–473, 2017. URL `https://dl.acm.org/citation.cfm?id=3091194`.

[56] S. Levine and V. Koltun. Continuous inverse optimal control with locally optimal examples. In *29th International Conference on Machine Learning (ICML)*, 2012.

[57] S. C. Levinson. On the human 'interactional engine'. *In NJ Enfield and SC Levinson, Roots Of*, 2006.

[58] M. L. Littman and P. Stone. Implicit negotiation in repeated games. In *International Workshop on Agent Theories, Architectures, and Languages*, pages 393–404. Springer, 2001.

[59] S.-Y. Lo, B. Fernandez, and P. Stone. Iterative human-aware mobile robot navigation. In *RSS'17 Workshop on Human Centered Robotics*, 2017.

[60] D. P. Losey and D. Sadigh. Robots that take advantage of human trust. In *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 7001–7008. IEEE, 2019.

[61] R. Lowe, Y. WU, A. Tamar, J. Harb, O. P. Abbeel, and I. Mordatch. Multi-agent actor-critic for mixed cooperative-competitive environments. In *Advances in Neural Information Processing Systems 30 (NIPS 2017)*, pages 6379–6390, 2017.

[62] T. Makino and J. Takeuchi. Apprenticeship learning for model parameters of partially observable environments. In *29th International Conference on Machine Learning*, 2012.

[63] J. C. McCall and M. M. Trivedi. Driver behavior and situation aware brake assistance for intelligent vehicles. *Proceedings of the IEEE*, 95:374–387, 2 2007. ISSN 0018-9219. doi: 10.1109/JPROC.2006.888388.

[64] F. S. Melo, M. T. J. Spaan, and S. J. Witwicki. Querypomdp: Pomdp-based communication in multiagent systems. In *EUMAS 2011: European Workshop on Multi-Agent Systems*, pages 189–204, 2012. doi: 10.1007/978-3-642-34799-3.

[65] C. A. Miller. Delegation and transparency: Coordinating interactions so information exchange is no surprise. In *International Conference on Virtual, Augmented and Mixed Reality*, pages 191–202. Springer, Cham, 2014. doi: 10.1007/978-3-319-07458-0.

[66] J. B. Misyak and N. Chater. Virtual bargaining: a theory of social decision-making. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369:20130487–20130487, 9 2014. ISSN 0962-8436. doi: 10.1098/rstb.2013.0487.

[67] V. Mnih, A. P. Badia, M. Mirza, A. Graves, T. Lillicrap, T. Harley, D. Silver, and K. Kavukcuoglu. Asynchronous methods for deep reinforcement learning. In *International conference on machine learning*, pages 1928–1937. PMLR, 2016.

[68] H. Moll and M. Tomasello. Cooperation and human cognition: the vygotskian intelligence hypothesis. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 362(1480):639–648, 2007.

[69] K. P. Murphy. *Machine learning: a probabilistic perspective*. MIT press, 2012.

[70] R. R. Murphy and D. Schreckenghost. Survey of metrics for human-robot interaction. In *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 197–198, 2013. ISBN 978-1-4673-3101-2. doi: 10.1109/HRI.2013.6483569.

[71] R. Nair, M. Tambe, M. Yokoo, D. Pynadath, and S. Marsella. Taming decentralized pomdps: Towards efficient policy computation for multiagent settings. In *IJCAI*, volume 3, pages 705–711, 2003.

[72] A. Y. Ng and S. J. Russell. Algorithms for inverse reinforcement learning. In *Seventeenth International Conference on Machine Learning (ICML)*, 2000.

[73] T.-H. Nguyen, D. Hsu, W.-S. Lee, T.-Y. Leong, L. Kaelbling, T. Lozano-Perez, and A. Grant. Capir: Collaborative action planning with intention recognition. In *AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, volume 6, 2011.

[74] S. Nikolaidis, M. Kwon, J. Forlizzi, and S. Srinivasa. Planning with verbal communication for human-robot collaboration. *ACM Transactions on Human-Robot Interaction*, 7:1–21, 11 2018. ISSN 25739522. doi: 10.1145/3203305.

[75] S. C. Ong, S. W. Png, D. Hsu, and W. S. Lee. Planning under uncertainty for robotic tasks with mixed observability. *The International Journal of Robotics Research*, 29(8):1053–1068, 2010.

[76] M. J. Osborne. *An introduction to game theory.* Oxford University Press, 2004. ISBN 9780195128956.

[77] O. Palinko, F. Rea, G. Sandini, and A. Sciutti. Robot reading human gaze: Why eye tracking is better than head tracking for human-robot collaboration. In *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 5048–5054, 2016. doi: 10.1109/IROS.2016.7759741.

[78] A. Panella and P. J. Gmytrasiewicz. Nonparametric bayesian learning of other. In *Workshops at the Twenty-Ninth AAAI Conference on Artificial Intelligence*, 2015.

[79] R. Paulin, T. Fraichard, and P. Reignier. Human-robot motion: Taking human attention into account. In *Workshop on Assistance and Service Robotics in a Human Environment, IROS*, 2018.

[80] T. Peltola, M. M. Çelikok, P. Daee, and S. Kaski. Modelling user's theory of ai's mind in interactive intelligent systems. *arXiv preprint arXiv:1809.02869*, 9 2018. URL `http://arxiv.org/abs/1809.02869`.

[81] C. Pérez-D'Arpino and J. A. Shah. Fast target prediction of human reaching motion for cooperative human-robot manipulation tasks using time series classification. In *2015 IEEE international conference on robotics and automation (ICRA)*, pages 6175–6182. IEEE, 2015.

[82] J. Perner, U. Frith, A. M. Leslie, and S. R. Leekam. Exploration of the autistic child's theory of mind: Knowledge, belief, and communication. *Child Development*, 60:689, 6 1989. ISSN 00093920. doi: 10.2307/1130734.

[83] J. Pineau, G. Gordon, and S. Thrun. Anytime point-based approximations for large pomdps. *Journal of Artificial Intelligence Research*, 27:335–380, 2006.

[84] D. V. Pynadath and S. C. Marsella. Psychsim: Modeling theory of mind with decision-theoretic agents. In *IJCAI*, volume 5, pages 1181–1186, 2005.

[85] J. Pöppel and S. Kopp. Satisficing models of bayesian theory of mind for explaining behavior of differently uncertain agents: Socially interactive agents track. In *17th International Conference on Autonomous Agents and MultiAgent Systems*, pages 470–478, 2018. URL `https://dl.acm.org/citation.cfm?id=3237455`.

[86] N. Rabinowitz, F. Perbet, F. Song, C. Zhang, S. M. A. Eslami, and M. Botvinick. Machine theory of mind. In *International conference on machine learning*, pages 4218–4227, 7 2018.

[87] R. Raileanu, E. Denton, A. Szlam, and R. Fergus. Modeling others using oneself in multi-agent reinforcement learning. In *International conference on machine learning*, pages 4257–4266. PMLR, 2018.

[88] D. Ramachandran and E. Amir. Bayesian inverse reinforcement learning. In *IJCAI*, volume 7, pages 2586–2591, 2007.

[89] S. Reddy, A. D. Dragan, and S. Levine. Where do you think you're going?: Inferring beliefs about dynamics from behavior. In *Advances in Neural Information Processing Systems 31 (NIPS 2018)*, 2018.

[90] S. Reddy, S. Levine, and A. D. Dragan. Assisted perception: Optimizing observations to communicate state. *arXiv preprint arXiv:2008.02840*, 8 2020. URL `http://arxiv.org/abs/2008.02840`.

[91] J. Renoux, T. S. Veiga, P. U. Lima, and M. T. J. Spaan. A unified decision-theoretic model for information gathering and communication planning. In *2020 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, 2020.

[92] T. Rezvani, K. Driggs-Campbell, D. Sadigh, S. S. Sastry, S. A. Seshia, and R. Bajcsy. Towards trustworthy automation: User interfaces that convey internal and external awareness. In *2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)*, pages 682–688. IEEE, 2016.

[93] S. Ross, J. Pineau, S. Paquet, and B. Chaib-draa. Online planning algorithms for pomdps. *Journal of Artificial Intelligence Research*, 32:663–704, 7 2008. ISSN 1076-9757. doi: 10.1613/jair.2567.

[94] C. A. Rothkopf and D. H. Ballard. Modular inverse reinforcement learning for visuomotor behavior. *Biological Cybernetics*, 107:477–490, 8 2013. ISSN 0340-1200. doi: 10.1007/s00422-013-0562-6.

[95] D. Sadigh, S. Sastry, S. A. Seshia, and A. D. Dragan. Planning for autonomous cars that leverage effects on human actions. In *Robotics: Science and Systems XII*, 2016. ISBN 9780992374723. doi: 10.15607/RSS.2016.XII.029.

[96] D. Sadigh, S. S. Sastry, S. A. Seshia, and A. Dragan. Information gathering actions over human internal state. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2016)*, pages 66–73. IEEE, 10 2016. ISBN 978-1-5090-3762-9. doi: 10.1109/IROS.2016.7759036.

[97] K. Sakita, K. Ogawara, S. Murakami, K. Kawamura, and K. Ikeuchi. Flexible cooperation between human and robot by interpreting human intention from gaze information. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, volume 1, pages 846–851. IEEE, 2004. ISBN 0780384636. doi: 10.1109/iros.2004.1389458.

[98] P. M. Salmon, N. A. Stanton, and D. P. Jenkins. *Distributed situation awareness: Theory, measurement and application to teamwork*. CRC Press, 2017.

[99] C. Schroeder de Witt, J. Foerster, G. Farquhar, P. Torr, W. Boehmer, and S. Whiteson. Multi-agent common knowledge reinforcement learning. *Advances in Neural Information Processing Systems*, 32:9927–9939, 2019.

[100] P. C. Schutte. How to make the most of your human: design considerations for human–machine interactions. *Cognition, Technology & Work*, 19(2):233–249, 2017.

[101] J. Schwehr and V. Willert. Multi-hypothesis multi-model driver's gaze target tracking. In *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, pages 1427–1434. IEEE, 2018.

[102] J. Schwehr, S. Luthardt, H. Dang, M. Henzel, H. Winner, J. Adamy, J. Fürnkranz, V. Willert, B. Lattke, M. Höpfl, et al. The proreta 4 city assistant system: Adaptive maneuver assistance at urban intersections using driver behavior modeling. *at-Automatisierungstechnik*, 67(9):783–798, 2019.

[103] B. Sendhoff and H. Wersing. Cooperative intelligence-a humane perspective. In *2020 IEEE International Conference on Human-Machine Systems (ICHMS)*, pages 1–6. IEEE, 2020.

[104] D. Silver and J. Veness. Monte-carlo planning in large pomdps. *Advances in Neural Information Processing Systems*, 23, 2010.

[105] A. Somani, N. Ye, D. Hsu, and W. S. Lee. Despot: Online pomdp planning with regularization. In *Advances in Neural Information Processing Systems 26*, 2013.

[106] D. Sperber and D. Wilson. *Relevance: Communication and cognition*, volume 142. Harvard University Press Cambridge, MA, 1986.

[107] A. St. Clair and M. Mataric. How robot verbal feedback can improve team performance in human-robot task collaborations. In *Tenth annual acm/ieee international conference on human-robot interaction*, pages 213–220, 2015.

[108] P. Stone, G. A. Kaminka, S. Kraus, and J. S. Rosenschein. Ad hoc autonomous agent teams: Collaboration without pre-coordination. *Twenty-Fourth AAAI Conference on Artificial Intelligence*, 7 2010.

[109] L. D. Strater, M. R. Endsley, R. J. Pleban, and M. D. Matthews. Measures of platoon leader situation awareness in virtual decision-making exercises. Technical report, TRW INC FAIRFAX VA SYSTEMS AND INFORMATION TECHNOLOGY GROUP, 2001.

[110] R. S. Sutton and A. G. Barto. *Reinforcement learning: An introduction.* MIT press, 2014.

[111] R. Taylor. Situation awareness rating technique: the development of a tool for aircrew system design. In *AGARD conference proceedings*, volume 478, 1990.

[112] K. A. Thomas, P. DeScioli, O. S. Haque, and S. Pinker. The psychology of coordination and common knowledge. *Journal of Personality and Social Psychology*, 107:657–676, 2014. ISSN 00223514. doi: 10.1037/a0037037.

[113] C. Torrey, S. R. Fussell, and S. Kiesler. How a robot should give advice. In *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 275–282. IEEE, 2013. doi: 10.1109/HRI.2013.6483599.

[114] V. V. Unhelkar, S. Li, and J. A. Shah. Decision-making for bidirectional communication in sequential human-robot collaborative tasks. In *ACM/IEEE International Conference on Human-Robot Interaction*, pages 329–341, 2020. doi: 10.1145/3319502.3374779.

[115] S. Vinanzi, M. Patacchiola, A. Chella, and A. Cangelosi. Would a robot trust you? developmental robotics model of trust and theory of mind. *Philosophical Transactions of the Royal Society B*, 374(1771): 20180032, 2019.

[116] N. Wang, D. V. Pynadath, and S. G. Hill. Trust calibration within a human-robot team: Comparing automatically generated explanations. In *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 109–116. IEEE, 2016.

[117] Z. Wang, K. Mülling, M. P. Deisenroth, H. Ben Amor, D. Vogt, B. Schölkopf, and J. Peters. Probabilistic movement modeling for intention inference in human–robot interaction. *The International Journal of Robotics Research*, 32(7):841–858, 2013.

[118] B. Wilder, E. Horvitz, and E. Kamar. Learning to complement humans. In *Twenty-Ninth International Joint Conference on Artificial Intelligence (IJCAI-20)*, 2020.

[119] H. Wimmer and J. Perner. Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, 13:103–128, 1 1983. ISSN 0010-0277. doi: 10.1016/0010-0277(83)90004-5.

[120] F. Wu, S. Zilberstein, and X. Chen. Multi-agent online planning with communication. *Nineteenth International Conference on Automated Planning and Scheduling*, 10 2009.

[121] W. Yoshida, R. J. Dolan, and K. J. Friston. Game theory of mind. *PLoS Comput Biol*, 4, 2008. doi: 10.1371/journal.pcbi.1000254.

[122] S. Young, M. Gasic, B. Thomson, and J. D. Williams. Pomdp-based statistical spoken dialog systems: A review. *Proceedings of the IEEE*, 101:1160–1179, 5 2013. ISSN 0018-9219. doi: 10.1109/JPROC.2012.2225812.

[123] L. Zettlemoyer, B. Milch, and L. Kaelbling. Multi-agent filtering with infinitely nested beliefs. *Advances in Neural Information Processing Systems*, pages 1–8, 2009.

[124] Z. Zhang. Parameter estimation techniques: A tutorial with application to conic fitting. *Image and vision Computing*, 15(1):59–76, 1997.

[125] A. Zhou, D. Hadfield-Menell, A. Naaabandi, and A. D. Dragan. Expressive robot motion timing. In *2017 12th ACM/IEEE International Conference on Human-Robot Interaction (HRI*, pages 22–31. IEEE, 2017.

[126] B. D. Ziebart, A. Maas, J. A. D. Bagnell, and A. Dey. Maximum entropy inverse reinforcement learning. In *Proceeding of AAAI 2008*, 2008.

Index

Action value function, 27
Active perception, 28
Additive logistic transform, 59

Bayesian Inference, 17
Belief, formal definition, 28

Capabilities, 12
Communication, 14
Communication effects, 73
Coordination, 11, 15, 32
Cost of communication, 75

Delegation, 11
Discount factor, 24

Extended Kalman Filter, 21

Human role, 11

Intention estimation, 39
Inverse Reinforcement Learning, 25, 35

Logistic normal distribution, 59

Markov Decision Process, 24
Markov property, 25

Observation, 27
Observation function, 27

Partially Observable Markov Decision Process, 27
Particle filter, 21
Policy, 25
POMDP, 27

Reinforcement learning, 25
Relevance, 14
Reward function, 24

Sampling based inference for filtering, 21
Situation awareness, 9, 62
State value function, 26
Strategic communication, 33
Strategic interaction, 31

Theory of Mind (TOM), 13
Transition function, 24
Transparency, 11
Trust, 12

Value alignment problem, 12
Value function, 26