Reasoning Under Uncertainty: Towards Collaborative Interactive Machine Learning

Sebastian Robert[1(✉)], Sebastian Büttner[2], Carsten Röcker[1,2,3],
and Andreas Holzinger[3]

[1] Fraunhofer-Institute of Optronics, System Technologies and Image Exploitation,
Application Center Industrial Automation (IOSB-INA), Lemgo, Germany
`sebastian.robert@iosb-ina.fraunhofer.de`
[2] Ostwestfalen-Lippe University of Applied Sciences, Lemgo, Germany
`{sebastian.buettner,carsten.roecker}@hs-owl.de`
[3] Holzinger Group, HCI-KDD, Institute for Medical Informatics,
Statistics and Documentation, Medical University Graz, Graz, Austria
`a.holzinger@hci-kdd.org`

Abstract. In this paper, we present the current state-of-the-art of decision making (DM) and machine learning (ML) and bridge the two research domains to create an integrated approach of complex problem solving based on human and computational agents. We present a novel classification of ML, emphasizing the human-in-the-loop in interactive ML (iML) and more specific on collaborative interactive ML (ciML), which we understand as a deep integrated version of iML, where humans and algorithms work hand in hand to solve complex problems. Both humans and computers have specific strengths and weaknesses and integrating humans into machine learning processes might be a very efficient way for tackling problems. This approach bears immense research potential for various domains, e.g., in health informatics or in industrial applications. We outline open questions and name future challenges that have to be addressed by the research community to enable the use of collaborative interactive machine learning for problem solving in a large scale.

Keywords: Decision making · Reasoning · Interactive machine learning · Collaborative interactive machine learning

1 Introduction and Motivation

Disregarding the application domain, i.e. whether in the medical domain or in the industrial context, current developments such as the rapidly growing communication infrastructure, the internet of things and increasing processing power with services and applications on top of those lead to massive amounts of data and new possibilities. Traditional analytic tools are not well suited to capturing the full value of "big data". Instead ML is ideal for exploiting opportunities hidden in data. Highly complex small batch production and personalized medicine (precision medicine [1]) are two of many possible target scenarios. Both depend

on computer-intensive data processing prior to its analysis and decision making processes.

However, to handle and exploit the required data, besides computer algorithms, human capabilities are strongly needed as well. For example, classical logic in ML approaches permits only exact reasoning, i.e. if A is true then A is non-false and if B is false then B is non-true. However, even though modern sophisticated automatic ML approaches can hardly cope with such situations, human agents can deal with such deficiencies.

Moreover, many ML approaches are based on *normative* models such as formal probability theory and expected utility (EU) theory. EU theory accounts for decision under uncertainty and is based on axioms of rational behavior described by von Neumann and Morgenstern (1944) [2]. Based upon the fact that information available in daily problem solving situations is most of the time imperfect, imprecise and uncertain due to time pressure, disturbance of unknown factors or randomness outcome of some attributes [3,4], the interaction between human and computer has to be designed in an optimal way in order to realize the best possible output. Given that, a combined approach of human and computer input can be a sustainable approach for effectively revealing structural or temporal patterns ("knowledge") and make them accessible for decision making.

At this point, decision theory comes into play and helps us to deal with bounded rationality and the problem of which questions to pose to human experts and how to ask those questions. Therefore, new types of human-computer interaction (HCI) will arise and shape the ecosystem of human, technology and organization. In particular, adaptive decision support systems that help humans to solve complex problems and make far-reaching decisions will play a central role in future work places.

In this paper, we will focus on decision making under uncertainty and bridge it to ML research and particularly to interactive ML. After discussing the state-of-the-art in ML and decision making under uncertainty, we provide some practical aspects for the integration of both approaches. Finally, we discuss some open questions and outline future research avenues.

2 Glossary and Key Terms

Bias refers to a systematic pattern of deviation from rationality in decision making processes.

Bounded Rationality – introduced by Herbert A. Simon [5] – is used to denote the type of rationality that people resort to when the environment in which they operate is too complex relative to their limited mental abilities [6].

Decision Support Systems (DSS) are intended to assist decision makers in taking full advantage of available information and are a central part of health informatics [7] and industrial applications [8].

Decision Theory is concerned with goal-directed behaviour in the presence of options [9]. While *normative* decision theory focuses on identifying the optimal decision to make, assuming a fully-rational decision maker who is able to

compute with perfect accuracy, *descriptive* decision theory deals with questions pertaining to how people actually behave in given choice situations. *Prescriptive* decision theory is the logic consequence and tries to exploit some of the logical consequences of normative theories and empirical findings of descriptive studies to make better choices [10].

Expected Utility (EU) Theory consists of four axioms, that define a rational decision maker: completeness, transitivity, independence, and continuity; if those are satisfied, then the decision making is considered to be rational and the preferences can be represented by a utility function, i.e. one can assign numbers (utilities) [2].

Heuristics describe approaches to problem solving and decision making which are not perfect, but sufficient for reaching immediate goals [11].

Human-Computer Interaction (HCI) is a multi-disciplinary research field that deals with "*the design, implementation and evaluation of interactive systems in the context of the user's task and work*" [12, p. 4]. It can be located at the intersection of psychology and cognitive science, ergonomics, computer science and engineering, business, design, technical writing and other fields [12, p. 4].

Judgment and Decision Making (JDM) is a descriptive field of research which focuses on understanding decision processes on an individual and group level.

Machine Learning (ML) is a research field grounding in computer science that "*concentrates on induction algorithms and on other algorithms that can be said to 'learn'*" [13]. While in *automatic Machine Learning (aML)* representations of real-world objects and knowledge are automatically generated from data, *interactive Machine Learning (iML)* methods allow humans to interact with computers in some way to generate knowledge and find an optimal solution for a problem. More specifically, *collaborative interactive Machine Learning (ciML)* is a form of iML, where at least one human is integrated into the algorithm using a specific user interface that allows manipulating the algorithm and its intermediate steps to find a good solution in a short time.

Perception-Based Classification (PBC) is a classification of data done by humans based on their visual perception. In the context of ML, PBC has been introduced by Ankerst et al. [14] who enabled users to interactively create decision trees. PBC can be seen as one possible way of realizing iML.

Utility Theorem describes that a decision-maker faced with probabilistic (particularly when probabilities are distorted or unknown) outcomes of different choices will behave as if she/he is maximizing the expected value [15]; this is the basis for the expected utility theory.

3   State-of-the-Art

In this section, we will provide an overview of the current research regarding two fields: First, we will investigate machine learning (ML) and focus especially

on the advances in interactive machine learning (iML). Second, we will provide an overview of the research on JDM under uncertainty. We will further focus on bridging the research on human decision making and the research on iML. We will motivate, why the knowledge of and research on human decision-making is key for the development of future human-oriented ciML systems.

3.1   Machine Learning (ML)

ML is a very practical field with many application areas, though at the same time well grounded theories with many open research challenges exist. There are many various definitions, depending on whom to ask; a Bayesian will give a different answer than a Symbolist [16]; a classical definition is close to and grounding in computer science that "*concentrates on induction algorithms and on other algorithms that can be said to 'learn'*" [13]. This definition is at the same time the goal of ML which concentrates on the development of "*programs that automatically improve with experience*" [17]. Advances in ML have solved many practical problems, e.g., recognizing speech [18], giving movie recommendations based on personal references [19] or driving a vehicle autonomously [20].

In the following, we will differentiate between classical ML approaches, that we will call aML and the newer concepts of iML.

Automatic Machine Learning (aML): Methods and algorithms of machine learning are often categorized as follows (here the classification of Marsland [21]):

– With *supervised learning* methods, an algorithm creates a general model from a training set of examples containing input and output data (targets). With this model, the output of new unknown input can be predicted.
– Contrary, when using *unsupervised learning* methods, the output data are not provided to the algorithm. The algorithm focuses on finding similarities between a set of input data and classifies the data into categories.
– *Reinforcement learning* is somehow between supervised and unsupervised learning. It characterizes algorithms that receive feedback, in the case that their created output data are wrong. By this feedback the algorithm can explore possibilities and iteratively find better models, respectively outputs.
– Finally *evolutionary learning* methods develop models iteratively by receiving an assessment of the quality (fitness) of the current model. As the term depicts, this learning method is inspired by the biological evolution.

The mentioned methods and algorithms all have in common, that they – once started – run automatically. We therefore call those classical machine learning methods *automatic machine learning*. When using aML methods, human involvement is in general very limited and restricted to the following three aspects:

– Humans have to prepare the data and remove corrupt or wrong data sets from the input data (data cleansing).

– When using supervised learning methods, humans are responsible for providing the output data, e.g., for labeling data in classification tasks.
– Another user involvement is the assessment of a certain model and the evaluation. Humans can assess the generated model and its results, and decide, whether a certain model is able to produce good predictions or not.

The traditional approach does not put much emphasize on the human interactions with the ML system. Humans are somehow involved in providing the data as described above, but the early ML research mostly neglects the question, how humans can provide data and how they deal with an inaccurate model. From a practical perspective, this is a huge restriction in automatic machine learning (aML) systems. The main problems of practical ML applications are often not the implementation of the algorithm itself, but rather the data acquisition and cleansing. Often data are corrupt or of bad quality and in most cases data do not cover all required context information to solve a specific problem [3,4].

Interactive Machine Learning (iML): Compared to aML, iML is a relative new approach that also considers the human involvement and interactions in ML and aims at putting the human into the loop of machine learning. In this section, we will discuss the approaches and concepts that previously have been described under the term iML. We will distinguish in this section between three types of iML methods: First, early works in the iML research considered iML as an alternative way of ML where humans accomplish the model generation, which basically means that humans replace algorithms. Second, concepts have been proposed under the term iML that put a human into the training-evaluation loop, but still execute algorithms automatically. Contrary to aML in this type of iML algorithms have to be much faster to give rapid feedback to a user. Third, humans can work hand in hand with algorithms to create a certain model, which we consider as the most promising concept of iML with the best integration of users and algorithms.

Humans replacing algorithms: Early work in iML has been done by Ankerst et al. [14]. They implemented a system called perception-based classification (PBC) that provides users the means to interactively create decision trees by visualizing the training data in a suitable way. By interacting with the visualized training data, users select attributes and split points to construct the decision trees. The system cannot automatically generate the tree. Instead, the user of the system replaces the algorithm and creates the tree manually with the interactive application provided. According to their evaluation, the system reaches the same accuracy as algorithmic classifiers but the human-generated decision trees have a smaller tree size, which is beneficial in terms of understandability. Another advantage of the interactive and manual approach is the possibility of backtracking in case of a suboptimal subtree – a situation that humans can easily recognize [14]. A huge benefit of this human-centered approach is the integration of the users' domain knowledge into the decision tree construction [22].

Building on the work of Ankerst et al., Ware et al. [23] developed a similar system that replaces the algorithm with users. Their work focuses mainly on an empirical evaluation of the performance of humans compared to state-of-the-art algorithms. According to their study, novice users can build trees that are as accurate as the ones provided from algorithms, but similar to Ankerst et al. they found, that the tree size is decreased, when humans generate the decision trees. On the other hand, Ware et al. point out that this manual iML approach might not be suitable for large data sets and high-dimensional data. This early variant of interactive machine learning is shown in Fig. 1A.

Humans in the training-evaluation loop: Another variety of iML is the integration of humans into the training-evaluation loop, when using supervised learning methods. Fails and Olsen [24] were one of the first, who used the term iML and proposed this integration for the rapid development of models, if the feature selection cannot be done by domain-experts due to missing knowledge. They give an example of the use of iML for the rapid development of perceptual user interfaces (PUIs), that are developed by interaction designers who are usually not familiar with computer vision (CV) algorithms. For this purpose, they provide a tool that gives designers rapid visual feedback of the produced classifiers and the iterative changes of the selected features for the model generation. The tool masks the complexity of the feature selection and rather allows users to assess the output of the model generation and to drive the feature selection into the right direction. A similar concept has been described by Fiebrink et al. [25]. They developed Wekinator[1], a system that analyses human gestures in the context of music making. A graphical user interface supports users with the creation of appropriate training data, the configuration of various ML algorithms and parameters and allows a real-time evaluation of the trained model by giving visual or auditory feedback. This real-time evaluation allows a domain user to rapidly adapt the input data to improve the model. Fogarty et al. [26] presented CueFlik, a similar iML tool for generating models for image classification tasks. For the mentioned type of iML, it is essential to have algorithms that have a very short learning time to be able to give rapid feedback on the results [24]. Addressing this particular aspect in connection with big data, Simard et al. [27] described a system that is very generic in terms of the data types and tasks and interactive even when using big data. Their system called ICE (interactive classification and extraction) allows users to interactively build models consisting of several millions of items. In [28] they extend their approach and additionally deliver feedback about the performance of the generated model to the user. With this system they empower users to not only optimize the model in terms of accuracy, but to optimize in terms of performance as well. While the mentioned systems use only one model, in recent years model ensembles became the standard of ML [16]. Talbort et al. therefore provide a tool that deals with multiple models and allows users to interactively build combination models [29]. All mentioned publications in this section use the term iML to describe a concept,

---

[1] http://www.wekinator.org/.

visualization by Mucha 2016 | icons from the noun project: (1) DesignNex, (2) Chameleon Design, (3) Nono Martínez Alonso, (4) Wilson Joseph, (5) Lloyd Humphreys
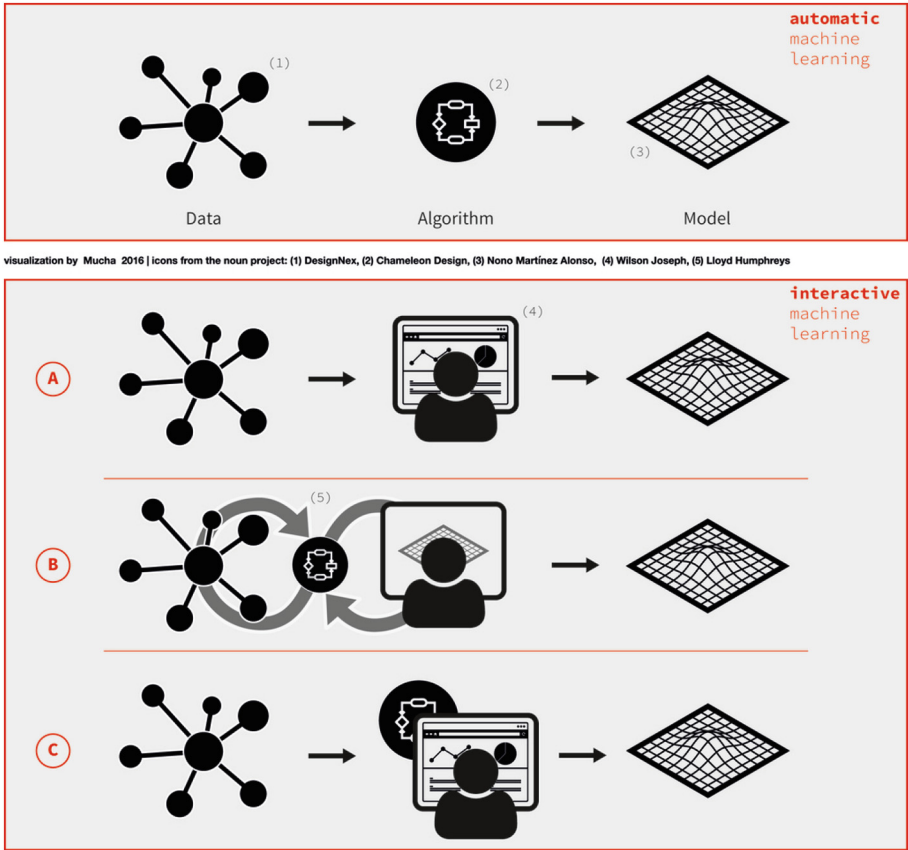
Fig. 1. Classification of interactive machine learning (iML). A: Early iML research aimed at replacing algorithms and using human pattern recognition capabilities instead. B: Later iML methods have been proposed that provide a rapid feedback cycle to users. Models are generated in a very short time and presented to users. Based on the presented model, users can adapt the input data and rerun the machine learning algorithm. With this approach the model is iteratively improved. C: Using collaborative interactive machine learning (ciML) humans can manipulate an algorithm during runtime and improve the model while it is generated. Human and computational agents work collaboratively on a specific problem.

where humans are in the training-evaluation loop, but cannot interfere with the algorithm itself – from a human perspective the algorithm is a black-box. The method of putting humans into the training-evaluation loop is shown in Fig. 1B.

Humans collaborating with algorithms: Sinard et al. define iML as a ML scenario, where "*the teacher can provide [. . .] information to the machine as the learning task progresses*" [27]. De facto, most systems presented in the past realized this iML by providing means to users to evaluate a certain model and by changing the training data to optimize the previously generated model. In this section,

we present work that goes even one step further and integrates humans into the process by providing a user interface that allows humans to manipulate the parameters of the algorithm during its execution. We will call this approach *collaborative interactive machine learning (ciML)*. Im this approach, humans can directly collaborate with an algorithm. With this deep integration, new possibilities of human-computer collaboration in ML might rise. One of the earliest works, that aimed at the collaboration between human and algorithm in a ML scenario has been presented by Ankerst et al. [30]. They built up on their earlier PBC system [14] and provide an iML system for building decision trees for a classification task. While their earlier PBC system only visualized data and left the decision tree building to the users, algorithms are now integrated into the system that might (but does not have to) be used. With the options provided, different types of cooperation can be realized: manual (equivalent to the earlier PBC), combined or completely automatic model generation. For the decision tree construction, the system supports with proposing splits, with visualizing hypothetical splits – up to a defined number of levels ("look-ahead function"), and with the feature of automatically expanding subtrees. One mentioned goal of their work is the use of human pattern recognition capabilities in the interactive decision tree construction by still using algorithmic operations to allow dealing with huge data sets [30]. Along these lines, Holzinger defines iML as "*algorithms that can interact with agents and can optimize their learning behavior through these interactions, where the agents can also be human*" [31], consequently, he considers iML as this deeply integrated type of a collaboration between algorithm and human. He discusses another issue that can be addressed with this deeply integrated form of iML: Sometimes ML needs to deal with rare events, like occurrences of rare diseases in health informatics, and consequently adequate training data are missing. He identifies new application areas for ciML within the health domain, e.g. for subspace clustering, protein folding, or k-anonymization of patient data and names challenges for the future ciML research. Holzinger also shows that the solution of complex problems is possible by using ciML. He presents the integration of users into an ant colony algorithm to solve a traveling salesman problem (TSP) [32]. A visualization shows the pheromone tracks of the ants in the TSP and the optimal round-trip found by the algorithm so far. Users can select edges and add or remove the current amount of pheromones on the edge between each of the iterations. First experiments show that the process is sped up in terms of required iterations to find the optimal solution [32]. The collaborative variant of interactive machine learning is shown in Fig. 1C. As the related work regarding the collaboration between humans and algorithms in iML shows, there has not been done a lot of research investigating the challenges and opportunities of a human-algorithm interaction. Application areas of this new iML approach need to be further identified and the implications of a human agent in the iML system need to be explored. While humans can bring tacit knowledge and context information into the process of building models, the question remains unclear how human decisions effect the output of the iML system. However, there has been a lot of research regarding human-decision making that we will introduce in the next section.

3.2   Judgement and Decision Research

Generally, the main focus of ML is on dealing with uncertainty and making predictions. In order to infer unknowns, data sets have to be learned and analysed. Therefore, most ML approaches are based on *normative* models such as formal probability theory and EU theory. EU theory accounts for decisions under uncertainty and is based on axioms of rational behavior, codified by von Neumann and Morgenstern [2]. It states that the overall utility of an option equals the expected utility, calculated by multiplying the utility and probability of each outcome [33, p. 24]. Probability theory in ML is most often used in terms of Bayesian decision theory [34–37], which is build on EU theory as a framework for solving problems under uncertainty [38, p. 140]. "Individuals who follow these theories are said to be rational" [39, p. 724].

The successful integration of knowledge of a domain expert in the black-box as discussed in the iML approach stands or falls with the careful consideration of people's actual decision making abilities. It is generally accepted that human reasoning and decision making abilities can exhibit various shortcomings when compared with mathematical logic [3]. Hence, the question that arises is, how to integrate human and computer input, accounting for the imperfections of both [40, p. 2122]. At this point *descriptive* decision theory can offer useful insights for the optimal integration of human judgement in iML approaches.

Descriptive decision theory deals with questions pertaining to how people behave in given choice situations and what we need to fully predict their behaviour in such situations [41, p. 2]. In many cases, this is a difficult task due to given inconsistencies in people's choices. These inconsistencies can often be attributed to irrational behaviour or accidental errors, which can also lead to deficient decisions [41, p. 6].

Within the last decades, a growing research community within the area of descriptive decision making is focusing on understanding individual and group judgement and decision making (JDM) [42,43].[2] Researchers from various fields are actively contributing to JDM, e.g. cognitive psychologists, social psychologists, statisticians and economists [42,45]. They have developed a detailed picture of the ways in which individuals judgement is bounded [46], e.g., people violate the axions of EU theory and do not always follow basic principles of calculus [47,48]. JDM tasks are characterized by uncertainty and/or by a concern for individual's preferences and will therefore apply to central aspects of human activities in iML [38, p. 140]. In detail, JDM research focuses on how different factors (e.g., information visualization) affect decision quality and how it can be improved [49,50]. In order to give any predictions about human judgement, JDM usually presupposes a definition of rationality that makes certain actions measurable. This instrumental view of rationality only accords with normative theory if keeping in line with it helps to attain satisfaction – measured in subjective utility [51]. A basic approach of JDM is to compare actual judgements to normative models and look for deviations. These so called *biases* are

---

[2] See also [44] for the chapter.

the starting point for building models that explain and predict human decision making behaviour. A fundamental outcome of early JDM research reveals that the typical model of a "rational man" as presumed by most normative theories – considering every possible action, every outcome in every possible state and calculating the choice that would lead to the best outcome – is unrealistic and does not exist [5]. Instead innumerous studies revealed that people cannot carry out the complex and time-consuming calculations necessary to determine the ideal choice out of possible actions [52, p. 7]. Instead people act as "satisficers" and make decisions on the basis of limited information, cognitive limitations and the time available. Simon's concept of *bounded rationality* describes how people actually reach a judgement or a decision and has become a widely used model for human decision behaviour [5].

Building on Simon's model, Tversky and Kahneman developed their heuristics and biases program that fundamentally shaped our understanding of judgment as we know it today [48]. According to their argumentation, coming to a decision requires a process of information search. Information can be retrieved from memory or other external sources. In any case, information has to be preprocessed for the particular problem and a final conclusion has so be drawn. Therefore, information processing is key for decision making and limited cognitive abilities, as stated in the model of bounded rationality, might essentially impact decision quality. The major reason for the huge impact of the heuristics and biases program in research is, that it is able to explain a wide variety of different decision situations without restricting it due to motivated irrationality [52, p. 1].

Tversky and Kahneman assume, that decisions under uncertainty are based on heuristics rather than complex algorithms [48]. Heuristics are defined as mental short-cuts or rules of thumb and require only limited amount of information and cognitive abilities. Generally, heuristics achieve results fast and depend on low effort. To do so, they neglect relevant information, which can lead to systematic predictable deviations from rationality. There is a huge amount of evidence that biases can lead to poor outcomes in important and novel decisions [42, 53]. This, together with the fact that biases are systematic, emphasises the importance of incorporating heuristics in modelling.

In their pioneering work, Tversky and Kahneman described three fundamental heuristics [48] which are relevant in countless practical situations. The representativeness heuristic is applied when people make judgements about the probability of an unknown event. To come up with a judgement, people tend to judge the probability of the unknown event by finding a comparable known event and assume that the probabilities will be similar. For illustration, Tversky and Kahneman developed the "Linda problem", where they describe the fictitious person Linda as "31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations" [54, p. 297]. Thereupon they asked subjects which is more probable, (a) Linda being a bank teller or (b) Linda being a bank teller and actively involved in

feminist movement. Results reveal, that in accordance with their hypothesis, a vast majority (80–90%) of subjects chose the conjunction (b) to be more likely than the single event (a). From a logical perspective, a conjunction of events (b) can never be more likely than any of its constituents (a) and therefore indicates a violation of rationality. Within the last decades, many different biases have been linked to the representativeness heuristic (e.g., conjunction fallacy, base rate neglect, insensitivity to sample size) [42].

The availability heuristic is the second of Tversky and Kahnemans heuristics and states, that people rely upon knowledge that is easily available and comes to mind rather than complete data [55]. By relying on the availability of a given event in someone's memory, the actual probability of the event can often be predicted quite good. Nevertheless, sometimes the availability of an event is influenced by other factors besides the probability or frequency of the occurrence and in this case the availability heuristic will lead to systematic deviations from rationality [55]. For example the chronological distance or conciseness are factors that can influence the availability of an event. The cause of death "firearm" is estimated as much higher compared to "tobacco", which can be attributed to the media coverage of violence [42,56]. Similar to this, subjects who were asked to estimate "If a random word is taken from an English text, is it more likely that the word starts with a K, or that K is the third letter?" [55, p. 1125]. Following Tversky and Kahneman's hypothesis, people easier recall words beginning with an K and therefore overestimate the number of words that begin with the letter K. Although experimental results support this hypothesis, a text typically contains twice as many words which have the letter K at the third, rather than first letter.

The so-called anchoring and adjustment heuristic describes a widely explored and robust phenomenon in human decision making [48]. The heuristic can be very useful when primary values of information do hint to a correct answer and are relevant to the underlying decision-problem – a situation found in many daily tasks. The anchor effect – as the central result of the anchoring and adjustment heuristic – can be found in situations, where a numerical starting point (the *anchor*) is processed to form a final estimation. In case the final estimation is biased towards the initial starting point, one talks about an anchoring effect. In a well-known demonstration, Tversky and Kahneman asked subjects to estimate the percentage of African countries that are in the United Nations (UN) [48, p. 1128]. Prior to this, for every subject of the experiment, a random number between one and one hundred after spinning a wheel of fortune was chosen. Subjects had to state if the random number is higher or lower compared to the true value. It was found, that people who had a lower number estimated fewer countries in the UN than people who had a higher number. Thereupon numerous experiments validated the robustness of the anchoring effect in varying fields of application, e.g. general knowledge [57], probability estimates [44,58] and negotiations [59,60]. Neither financial incentives nor explicit advices could effectivly mitigate the anchoring effect [61,62]. Moreover, the numerical starting point does not have to be relevant to the underlying decision-problem, even

unconsciously perceived or irrelevant values can distort the judgement [61, p. 123]. In general, there are two different approaches to explain the occurrence of the anchor effect. The original approach of Tversky and Kahneman states that individuals tend to anchor onto a numerical value and then gradually adjust away from that value until they reach a decision that seems reasonable [48]. This anchoring and adjustment process is usually insufficient and therefore biased. In contrast, the selective accessibility approach argues, that biased estimations are rooted in an early phase of information processing [57,63,64]. Following the approach, individuals, when given an anchor, will evaluate the hypothesis that the anchor is a suitable answer (confirmatory hypothesis testing) and therefore access all the relevant attributes of the anchor value. Thereon, the approach assumes that anchoring effects are mediated by the selectively increased accessibility of anchor-consistent knowledge and the final estimate is therefore biased towards the anchor. Overall, none of the mentioned approaches can fully explain empirical evidence and the origin of the anchoring effect is still highly debated within the research community [42,65].

In addition to the three fundamental heuristics and their resulting biases, there are further heuristics which try to explain decision making under specific situations. Despite the tremendous success of the heuristics and biases program, there are alternative approaches to explain actual decision making behaviour. For example the fast-and-frugal-approach – mostly based on Gigerenzers works – is also based on several simple heuristics, but in contrast to the classical heuristics, they are precisely defined and can be directly validated [66,67]. Moreover, the probabilistic mental model [68] and prospect theory [69] also build on limited cognitive abilities and are used in different areas to predict decision making behaviour.

3.3   Practical Aspects for the Integration of Interactive ML and Decision Theory

The importance of the integration of interactive ML and decision theory is evident. Given the massive consequences that can result from suboptimal decision making, it is critical to improve our knowledge about ways to yield better decision outcomes [46, p. 379]. In our knowledge-based economy, each decision is likely to have vast implications and will affect subsequent decisions on their own. Decision problems have to be analysed for their potential receptiveness to decision biases and in what ways they are likely to benefit from automatic processing.

On the one side, current technological and methodical advances enable us to cope with more complex decision tasks. But on the other side, in many practical situations decision making in terms of the interaction between human and computer input is still limited and does not tap the full potential. Moreover, new decision situations in many fields of application are characterized by the same underlying process and therefore share the common need for new ways of interaction.

For example, there are innumerable applications in the field of medical decision making and cyber-physical systems (e.g. "Industry 4.0") such as assistance or recommender systems that are based on the same abstract decision problem, combine similar approaches of computer algorithms with human input and therefore face similar challenges. For instance, the analysis of sensor data is pretty similar in many practical applications. On the one hand, data may describe body parameters such as temperature, heartbeat or blood plasma concentration in a medical context. On the other hand, data may provide information about the energy consumption of a power unit, the temperature of an engine or the status of a relay in an industrial context. Although there are many algorithms that can analyse the captured data in a purely unsupervised fashion, in order to achieve excellent and instant results, an interactive data analysis backed by human decision making skills can offer new possibilities and bring context information into the process. The same applies to the area of image exploitation. In many cases, it is about finding structural anomalies in data and learning from previous examples. With up-to-date methods of image exploitation, algorithms can detect, count and cluster different types of objects. These algorithms are in many cases only partially automatic and require human input. In medical image exploitation, doctors can help to provide diagnostic findings in the segmentation of skin cancer images [70]. In the industrial context, image exploitation is for example used to detect tool wear [71]. In both situations, wrong diagnoses and decisions potentially bear extensive risk and therefore the optimal integration of human and computer input is of great importance. A big issue is accordingly the integration process, because exactly here setting up a system between the expert and the algorithm requires a common ground between them and is crucial for total imaging. This common ground has to exploit computational power and integrate human intelligence to realise the best possible output.

4   Open Problems

The study of ML is primarily based on normative models. Most of these models are the result of centuries of refection and analysis and are widely accepted as the basis of logical reasoning. For the fact that human decision making skills are in certain settings superior to computer algorithms – e.g. many ML-methods perform very badly on extrapolation problems which would be very easy for humans [32, p. 4] – and major assumptions of normative models cannot be applied in reality, a conjoint approach of human and machine input could be key to enhanced decision quality. Therefore, the answer is to put humans in the loop [40]. However, using normative models to integrate human decision making in centrals parts of machine learning could lead to faulty predictions since the nature of actual decision making is of bounded rationality [5].

Based on the described approaches, today we know the specific ways in which decision makers are likely to be biased and we can describe how people make decisions with astonishing detail and reliability. In addition, with regards to normative models, we have a clear vision of how much better decision making

could be [46]. The most important step now is to integrate those two different approaches, correct biases and improve decision making. The prescriptions for such corrections are called prescriptive models [33, p. 19] and will decide about the success of human-in-the-loop approaches in ML. Altogether, not only do we need to know the nature of the specific problem, "but normative models must be understood in terms of their role in looking for biases, understanding these biases in terms of descriptive models and developing prescriptive models" [72, p. 20].

In consideration of this fact, interactive ML approaches are a promising candidate for further enhancing the knowledge discovery process. One important problem which we have to face in future research is which questions to pose to humans and how to ask those questions [40]. At this point, human machine-interaction could provide useful insights and offer guidelines for the design of interfaces and visualisations. Moreover, research in this area, i.e. at the intersection of cognitive science and computational science is fruitful for further improving ML thus improve performance on a wide range of tasks, including settings which are difficult for humans to process (e.g., big data and high dimensional problems) [32]. According to Lee and Holzinger [73], there is a very common misconception about high dimensionality, i.e. that ML would produce better outcomes with higher dimensional data. Increasing amounts of input features can build more accurate predictors as features are key to learning and understanding. However, such attempts need high computational power, and due to limitations in human perception, understanding structures in high dimensional spaces is practically impossible. Hence, the outcome must be shaped in a form perceivable for humans, which is a very difficult problem. Here graph-based representations in $\mathbb{R}^2$ are very helpful in that respect and open up a lot of future possibilities [74,75].

5   Future Challenges

The important role of iML for dealing with complexity is evident. However, future research has to be done in various areas.

First of all, only a few research projects have dealt with ciML. The development of new ciML approaches for different algorithms has to be expanded to be able to develop generic human-algorithm interfaces. Research has to focus on further algorithms beyond decision trees and ant colony algorithms that could benefit from the new approach of ciML to analyze its full potential.

Secondly, from the knowledge today it cannot be said, which problems ciML can address and which problems will not be addressable with ciML. Future research has to focus on the classification of problems in terms of the different aML, iML and ciML approaches. For some problems we do know that aML can provide very efficient algorithms, some problems are known to be unsolvable in polynomial time, but we currently do not have comprehensive knowledge about the opportunities of ciML.

Thirdly, the iML algorithms proposed so far address very specific problems. In general, the questions have been solved, how humans can be integrated into

the algorithm and understand both the underlying problem and the algorithm with its parameters. Therefore, the past and ongoing research on HCI will play a prominent role in the future of iML: It has to be further analyzed, how humans (not only computer scientists) can be empowered to better understand the specific ML algorithms. This involves adequate visualization techniques of the input data, as shown by past research projects as well as visualizations to support the understandability of complex algorithms. In this respect, new interaction technologies might come in handy. Large displays [76], room-spanning projections [77], gesture-based interactions and virtual and augmented reality (VR and AR) [78,79] are new interaction concepts and technologies that have been applied successfully in the medical [80] and industrial [81–83] domains and might be able to play a roll in the interaction with algorithms in the future.

6   Conclusion

In this paper, we presented the current state of research in two domains: JDM and ML. We presented a new classification of ML emphasizing on iML and – more specificly – on ciML. We bridged the two research domains and argued that future research will have to take both research domains into account, when dealing with highly complex problems. Both humans and computers have their specific strengths and weaknesses and putting humans into the loop of ML algorithms might be a very efficient way for solving specific problems. We identified two application areas, which provide complex problems that might benefit from the new approach of ciML: health informatics and cyber-physical systems. While these two domains seem to be different on the first sight, their problems often share the same characteristics: Often exceptional variances in data need to be found, e.g. a specific diseases based on physiological data in medicine or malfunctions of complex cyber-physical systems based on sensor data of machines. The classical approach of aML focuses on finding these patterns based on previous knowledge from data. However, aML struggles on function extrapolation problems which are trivial for human learners. Consequently, integrating a human-into-the-loop (e.g., a human kernel [84]) could make use of human cognitive abilities and will be a promising approach. While we outlined the potential of ciML there are multiple open questions to be tackled in the research community. The explorative development of new ciML approaches for different algorithms will help to analyze the full potential of ciML. Existing complex problems need to be classified and application areas for the different iML approaches need to be identified. And last but not least, the questions on how to support humans ideally when collaborating with algorithms and big data needs to be addressed. In this area the experts from both ML and HCI will have to work hand in hand in this new joint research endeavor that will greatly help in future problem solving.

References

1. Holzinger, A.: Trends in interactive knowledge discovery for personalized medicine: cognitive science meets machine learning. IEEE Intell. Inform. Bull. **15**, 6–14 (2014)
2. Von Neumann, J., Morgenstern, O.: Theory of Games and Economic Behavior. Princeton University Press, Princeton (1944)
3. Fox, J., Glasspool, D., Bury, J.: Quantitative and qualitative approaches to reasoning under uncertainty in medical decision making. In: Quaglini, S., Barahona, P., Andreassen, S. (eds.) AIME 2001. LNCS, vol. 2101, pp. 272–282. Springer, Heidelberg (2001). doi:10.1007/3-540-48229-6_39
4. Ma, W., Xiong, W., Luo, X.: A model for decision making with missing, imprecise, and uncertain evaluations of multiple criteria. Int. J. Intell. Syst. **28**, 152–184 (2013)
5. Simon, H.A.: A behavioral model of rational choice. Q. J. Econ. **69**, 99–118 (1955)
6. Dequech, D.: Bounded rationality, institutions, and uncertainty. J. Econ. Issues **35**, 911–929 (2001)
7. Holzinger, A.: Lecture 8 biomedical decision making: reasoning and decision support. In: Biomedical Informatics, pp. 345–377. Springer, Heidelberg (2014)
8. March, S.T., Hevner, A.R.: Integrated decision support systems: a data warehousing perspective. Decis. Support Syst. **43**, 1031–1043 (2007)
9. Hansson, S.O.: Decision theory: a brief introduction (2005)
10. Bell, D.E., Raiffa, H., Tversky, A.: Descriptive, normative, and prescriptive interactions in decision making. Decis. Making Descriptive Normative Prescriptive Interact. **1**, 9–32 (1988)
11. Pearl, J.: Heuristics: Intelligent Search Strategies for Computer Problem Solving. Addison-Wesley, Reading (1984)
12. Alan, D., Janet, F., Gregory, A., Russell, B.: Human-Computer Interaction. Pearson Education Limited, Harlow (2004)
13. Kohavi, R., Provost, F.: Glossary of terms. Mach. Learn. **30**, 271–274 (1998)
14. Ankerst, M., Elsen, C., Ester, M., Kriegel, H.P.: Visual classification: an interactive approach to decision tree construction. In: Proceedings of the Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 392. ACM (1999)
15. Wakker, P., Deneffe, D.: Eliciting von neumann-morgenstern utilities when probabilities are distorted or unknown. Manage. Sci. **42**, 1131–1150 (1996)
16. Domingos, P.: A few useful things to know about machine learning. Commun. ACM **55**, 78–87 (2012)
17. Mitchell, T.M.: Machine Learning. McGraw-Hill, Boston (1997)
18. Martin, J.H., Jurafsky, D.: Speech and language processing. In: International 710th edn. (2000)
19. Adomavicius, G., Tuzhilin, A.: Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions. IEEE Trans. Knowl. Data Eng. **17**, 734–749 (2005)
20. Li, Q., Zheng, N., Cheng, H.: Springrobot: a prototype autonomous vehicle and its algorithms for lane detection. IEEE Trans. Intell. Transp. Syst. **5**, 300–308 (2004)
21. Marsland, S.: Machine Learning: An Algorithmic Perspective. CRC Press, Boca Raton (2015)
22. Ankerst, M., Ester, M., Kriegel, H.P.: Towards an effective cooperation of the user and the computer for classification. In: Proceedings of the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 179–188. ACM (2000)

23. Ware, M., Frank, E., Holmes, G., Hall, M., Witten, I.H.: Interactive machine learning: letting users build classifiers. Int. J. Hum. Comput. Stud. **55**, 281–292 (2001)
24. Fails, J.A., Olsen Jr., D.R.: Interactive machine learning. In: Proceedings of the 8th International Conference on Intelligent User Interfaces, pp. 39–45. ACM (2003)
25. Fiebrink, R., Cook, P.R., Trueman, D.: Human model evaluation in interactive supervised learning. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI 2011, pp. 147–156. ACM, New York (2011)
26. Fogarty, J., Tan, D., Kapoor, A., Winder, S.: Cueflik: interactive concept learning in image search. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI 2008, pp. 29–38. ACM, New York (2008)
27. Simard, P., Chickering, D., Lakshmiratan, A., Charles, D., Bottou, L., Suarez, C.G.J., Grangier, D., Amershi, S., Verwey, J., Suh, J.: Ice: enabling non-experts to build models interactively for large-scale lopsided problems. arXiv preprint arXiv:1409.4814 (2014)
28. Amershi, S., Chickering, M., Drucker, S.M., Lee, B., Simard, P., Suh, J.: Modeltracker: redesigning performance analysis tools for machine learning. In: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, CHI 2015, pp. 337–346. ACM, New York (2015)
29. Talbot, J., Lee, B., Kapoor, A., Tan, D.S.: Ensemblematrix: interactive visualization to support machine learning with multiple classifiers. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI 2009, pp. 1283–1292. ACM, New York (2009)
30. Ankerst, M., Ester, M., Kriegel, H.P.: Towards an effective cooperation of the user and the computer for classification. In: Proceedings of the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD 2000, pp. 179–188. ACM, New York (2000)
31. Holzinger, A.: Interactive machine learning for health informatics: when do we need the human-in-the-loop? Brain Inform. **3**, 119–131 (2016)
32. Holzinger, A., Plass, M., Holzinger, K., Crişan, G.C., Pintea, C.-M., Palade, V.: Towards interactive Machine Learning (iML): applying ant colony algorithms to solve the traveling salesman problem with the human-in-the-loop approach. In: Buccafurri, F., Holzinger, A., Kieseberg, P., Tjoa, A.M., Weippl, E. (eds.) CD-ARES 2016. LNCS, vol. 9817, pp. 81–95. Springer, Heidelberg (2016). doi:10.1007/978-3-319-45507-5_6
33. Baron, J.: Normative Models of Judgment and Decision Making. Wiley, New York (2004)
34. Raiffa, H.: Applied statistical decision theory (1974)
35. Murphy, K.P.: Machine Learning: A Probabilistic Perspective. MIT Press, Cambridge (2012)
36. Friedman, J., Hastie, T., Tibshirani, R.: The Elements of Statistical Learning. Springer Series in Statistics, vol. 1. Springer, New York (2001)
37. Tulabandhula, T., Rudin, C.: Machine learning with operational costs. J. Mach. Learn. Res. **14**, 1989–2028 (2013)
38. Pitz, G.F., Sachs, N.J.: Judgment and decision: theory and application. Annu. Rev. Psychol. **35**, 139–164 (1984)
39. Fischhoff, B.: Judgment and decision making. Wiley Interdisc. Rev. Cogn. Sci. **1**, 724–735 (2010)
40. Russakovsky, O., Li, L.J., Fei-Fei, L.: Best of both worlds: human-machine collaboration for object annotation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2121–2131 (2015)

41. Rapoport, A.: Decision Theory and Decision Behaviour: Normative and Descriptive Approaches, vol. 15. Springer, Amsterdam (2013)
42. Bazerman, M.H., Moore, D.A.: Judgment in managerial decision making (2013)
43. Bonner, S.E.: Judgment and Decision Making in Accounting. Prentice Hall, Upper Saddle River (2008)
44. Robert, S.: Informationsverarbeitung in Prognosen: Experimentelle Evidenz. dissertation, University of Osnabrueck (2016)
45. Goldstein, W.M., Hogarth, R.M.: Research on Judgment and Decision Making: Currents, Connections, and Controversies. Cambridge University Press, Cambridge (1997)
46. Milkman, K.L., Chugh, D., Bazerman, M.H.: How can decision making be improved? Perspect. Psychol. Sci. **4**, 379–383 (2009)
47. Baron, J.: Thinking and Deciding. Cambridge University Press, Cambridge (2000)
48. Tversky, A., Kahneman, D.: Judgment under uncertainty: heuristics and biases. In: Wendt, D., Vlek, C. (eds.) Utility, Probability, and Human Decision Making, pp. 1124–1131. Springer, Amsterdam (1974)
49. Libby, R.: Accounting and Human Information Processing: Theory and Applications. Prentice Hall, Englewood Cliffs (1981)
50. Ashton, R.H.: Human Information Processing in Accounting. American Accounting Association, Sarasota (1982)
51. Over, D.: Rationality and the normative/descriptive distinction. In: Blackwell Handbook of Judgment and Decision Making, London, pp. 3–18 (2004)
52. Gilovich, T., Griffin, D., Kahneman, D.: Heuristics and Biases: The Psychology of Intuitive Judgment. Cambridge University Press, New York (2002)
53. Newell, B.R.: Judgment under uncertainty (2013)
54. Tversky, A., Kahneman, D.: Extensional versus intuitive reasoning: the conjunction fallacy in probability judgment. Psychol. Rev. **90**, 293 (1983)
55. Tversky, A., Kahneman, D.: Availability: a heuristic for judging frequency and probability. Cogn. Psychol. **5**, 207–232 (1973)
56. Mokdad, A.H., Marks, J.S., Stroup, D.F., Gerberding, J.L.: Actual causes of death in the United States, 2000. JAMA **291**, 1238–1245 (2004)
57. Strack, F., Mussweiler, T.: Explaining the enigmatic anchoring effect: mechanisms of selective accessibility. J. Pers. Soc. Psychol. **73**, 437 (1997)
58. Plous, S.: Thinking the unthinkable: the effects of anchoring on likelihood estimates of nuclear war1. J. Appl. Soc. Psychol. **19**, 67–91 (1989)
59. Ritov, I.: Anchoring in simulated competitive market negotiation. Organ. Behav. Hum. Decis. Process. **67**, 16–25 (1996)
60. Galinsky, A.D., Mussweiler, T.: First offers as anchors: the role of perspective-taking and negotiator focus. J. Pers. Soc. Psychol. **81**, 657 (2001)
61. Chapman, G.B., Johnson, E.J.: Incorporating the irrelevant: anchors in judgments of belief and value. In: The Psychology of Intuitive Judgment, Heuristics and Biases, pp. 120–138 (2002)
62. Wilson, T.D., Houston, C.E., Etling, K.M., Brekke, N.: A new look at anchoring effects: basic anchoring and its antecedents. J. Exp. Psychol. Gen. **125**, 387 (1996)
63. Mussweiler, T., Strack, F.: Comparing is believing: a selective accessibility model of judgmental anchoring. Eur. Rev. Soc. Psychol. **10**, 135–167 (1999)
64. Chapman, G.B., Johnson, E.J.: Anchoring, activation, and the construction of values. Organ. Behav. Hum. Decis. Process. **79**, 115–153 (1999)
65. Furnham, A., Boo, H.C.: A literature review of the anchoring effect. J. Socio-Econ. **40**, 35–42 (2011)

66. Gigerenzer, G.: Why the distinction between single-event probabilities and frequencies is important for psychology (and vice versa). In: Subjective Probability, pp. 129–161 (1994)
67. Gigerenzer, G., Czerlinski, J., Martignon, L.: How good are fast and frugal heuristics? In: Shanteau, J., Mellers, B.A., Schum, D.A. (eds.) Decision Science and Technology, pp. 81–103. Springer, New York (1999)
68. Gigerenzer, G., Hoffrage, U., Kleinbölting, H.: Probabilistic mental models: a brunswikian theory of confidence. Psychol. Rev. **98**, 506 (1991)
69. Kahneman, D., Tversky, A.: Prospect theory: an analysis of decision under risk. Econometrica: J. Econometric Soc. **47**(2), 263–291 (1979)
70. Xu, L., Jackowski, M., Goshtasby, A., Roseman, D., Bines, S., Yu, C., Dhawan, A., Huntley, A.: Segmentation of skin cancer images. Image Vis. Comput. **17**, 65–74 (1999)
71. Królczyk, G., Legutko, S., Raos, P.: Cutting wedge wear examination during turning of duplex stainless steel. Tehnički Vjesnik-Technical Gazette **20**, 413–418 (2013)
72. Baron, J.: Rationality and Intelligence. Cambridge University Press, New York (2005)
73. Lee, S., Holzinger, A.: Knowledge discovery from complex high dimensional data. In: Michaelis, S., Piatkowski, N., Stolpe, M. (eds.) Solving Large Scale Learning Tasks. Challenges and Algorithms. LNCS (LNAI), vol. 9580, pp. 148–167. Springer, Heidelberg (2016). doi:10.1007/978-3-319-41706-6_7
74. Holzinger, A., Malle, B., Giuliani, N.: On graph extraction from image data. In: Slezak, D., Peters, J.F., Tan, A.H., Schwabe, L. (eds.) Brain Informatics and Health, BIH 2014. LNAI, vol. 8609, pp. 552–563. Springer, Heidelberg (2014)
75. Valdez, A.C., Dehmer, M., Holzinger, A.: Application of graph entropy for knowledge discovery and data mining in bibliometric data. In: Dehmer, M., Emmert-Streib, F., Chen, Z., Li, X., Shi, Y. (eds.) Mathematical Foundations and Applications of Graph Entropy, pp. 259–272. Wiley, New York (2016)
76. Cao, X., Balakrishnan, R.: Visionwand: interaction techniques for large displays using a passive wand tracked in 3d. In: Proceedings of the 16th Annual ACM Symposium on User Interface Software and Technology, UIST 2003, pp. 173–182. ACM, New York (2003)
77. Jones, B.R., Benko, H., Ofek, E., Wilson, A.D.: Illumiroom: peripheral projected illusions for interactive experiences. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI 2013, pp. 869–878. ACM, New York (2013)
78. Milgram, P., Takemura, H., Utsumi, A., Kishino, F.: Augmented reality: a class of displays on the reality-virtuality continuum. In: Photonics for industrial applications, International Society for Optics and Photonics, pp. 282–292 (1995)
79. Azuma, R.T.: A survey of augmented reality. Presence: Teleoperators Virtual Environ. **6**, 355–385 (1997)
80. Fuchs, H., et al.: Augmented reality visualization for laparoscopic surgery. In: Wells, W.M., Colchester, A., Delp, S. (eds.) MICCAI 1998. LNCS, vol. 1496, pp. 934–943. Springer, Heidelberg (1998). doi:10.1007/BFb0056282
81. Paelke, V., Röcker, C., Koch, N., Flatt, H., Büttner, S.: User interfaces for cyberphysical systems. at-Automatisierungstechnik **63**, 833–843 (2015)
82. Büttner, S., Sand, O., Röcker, C.: Extending the design space in industrial manufacturing through mobile projection. In: Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct, MobileHCI 2015, pp. 1130–1133. ACM, New York (2015)

83. Büttner, S., Funk, M., Sand, O., Röcker, C.: Using head-mounted displays and in-situ projection for assistive systems - a comparison. In: Proceedings of the 9th ACM International Conference on PErvasive Technologies Related to Assistive Environments, vol. 8. ACM (2016)

84. Wilson, A.G., Dann, C., Lucas, C.G., Xing, E.P.: The human kernel. arXiv preprint arXiv:1510.07389 (2015)