On Risk Formalization of On-line Risk Assessment for Safe Decision Making in Robotics.

Philipp Ertle, Holger Voos, Dirk Söffker

Abstract— **In the near future service robots may fulfill demanding tasks in unstructured and complex human-like environments. A central challenge is to prevent such autonomous systems from provoking hazards by interaction with their environment, especially when it is partially unknown. Real-time risk information is seen as essential basis for safe decision making processes. Therefore, generalized and conservative safety principles are used to determine the interaction risks. The solution for formalizing safety principles and for quantifying necessary risk values is shown.**

I. Introduction

A. The Field of Service Robots

The International Federation of Robotics (IFR) [1] has defined a Service Robot as *'a robot which operates semi or fully autonomously to perform services useful to the well-being of humans and equipment, excluding manufacturing operations.'* It is assumed that a service robot is intended to move freely in a dynamic environment and to interact with objects and humans over a longer period of time in order to solve given tasks. Hence, the so-called mobile service robots (MSR) are additionally expected to provide services in various domains of life. Herein, the main challenge for the robot is the execution of complex tasks within an unstructured dynamic environment while collaborating with human users in a natural and intuitive way. Classical examples are, for example, a robot butler [15], able to grasp recognized objects and open doors, a library robot [10], able to fetch ordered books or the 'RoboCup@Home' competition in which interaction tasks are integrated. Similarly, a service robot is addressed in this contribution which performs tasks in a domestic environment, for example in a usual flat. Such tasks could be fetching a cup of coffee, watering plants or fetch the user's medicine at the correct time, for instance.

B. Behavior-based Safety Concerns

During performing of tasks, several hazards can occur depending on kind and size of the robot. In the field of robotic, formerly strongly motivated by industrial applications, dangerous physical energies were mainly focused when taking safety concerns into account. Therefore, one of the main goals was to avoid collisions with humans, for example by separating stationary robots from humans by a safety cage. In the field of service robots, robot's mobility is a central aspect in order to share the typical human living space. Furthermore, touching humans might be even necessary. The physical energies remain as a safety critical aspect, whereupon that strongly depends on weight and power of the robot. Failures and defects of components may lead to uncontrolled release of the inherent energy; hence these are often focused with respect to safety aspects. Not so in this contribution, hence, a typical service robot without need for huge physical force or speed is assumed. Even so, the service robot cannot be regarded as safe even if there is no chance to directly injure a human. Depending on the task, the robot manipulates objects and therefore may produce dangers which are connected to attributes of these objects. For example, a manipulator with harmless physical power may become dangerous when it has gripped a knife. Therefore, the manipulated object also has to be taken into account with regard to overall system's safety investigations. In the case that a service robot uses potential dangerous objects like a knife, for example, the moving speed and force have to be reduced in dependence of distance to humans. Accordingly, the safety assessment process for service robots is extended to objects which are intended to be manipulated in tasks later on.

Assuming the robot now is aware of objects or tools it uses and thus, it adapts its coordination and behavior in order to avoid dangers. Nevertheless, there remains a further class of dangers: The service robot manipulates objects that are not dangerous directly, but the interaction of manipulated objects with environmental objects can cause hazards, for example, when the service robot fulfills the task 'bring dishes to the kitchen sink'. Any perception module detects the kitchen stove as suitable surface to place the dishes. Hence, it deposits the dishes which could be made of wooden and plastic, on the kitchen stove which in turn is still hot. The interaction between dishes and stove may now provoke a further kind of dangers that has to be taken into account. Finally, for safety concerns of service robots it is important that

- the direct affecting of the environment alias physical energies,
- the indirect affecting of the environment by manipulated or carried objects or used objects (alias tools), and
- the indirect provoking of dangers by interaction of

P. Ertle and H. Voos are with ZAFH - Center of Collaborate Research at University of Applied Sciences Ravensburg-Weingarten, 88241 Weingarten, Germany, {ertle, voos}@hs-weingarten.de

P. Ertle and D. Söffker are with the Chair of Dynamics and Control, University of Duisburg-Essen, 47057 Duisburg, Germany, soeffker@uni-due.de

manipulated objects with surrounding objects is taken into account.

The so-called robot's awareness of its environment [9] can also be connected to safety aspects [16]. A disadvantage of the awareness is that it relies on the system's perception. The correctness and completeness of the environmental perception is a central aspect when intended to create situational awareness. Hence, the perception of the environment is based on probabilistic assumptions including uncertainties.

C. Need for Cognitive Technical Systems (CTS)

Amongst other aspects, the objectives of the AI community are to enable service robots, robots or other systems to learn skills or tasks by environmental or human feedback or by demonstration. Therefore, different systemic functions like perception, vision or learning are needed in order to realize higher functionality by a well-coordinated interplay of these components. The components and their interplay have to deal with a huge variety of information with respect to the environment (typical human/domestic environment assumed) and depending on used sensors. As long as processing power of the controlling computers is not able to process the full variety of information (if this is ever possible), only a selection of information can be processed. These are the so-called 'relevant aspects' of the full range of information. As cognitive systems are basically characterized by the capability to represent system-relevant aspects of the environment internally [5] (in order to process them), it is obvious that robotic systems which are intended to operate in human like (complex) environments are cognitive oriented systems somehow. Therefore, it is intended to investigate the safety aspects of CTS in this contribution.

There are several approaches and architecture for CTS. The chosen architecture design approach, the 'Situation Operator Modeling' (SOM) approach, is though as meta model [12] on the one hand and on the other hand it is successfully applied to robotics [2],[14]. Furthermore, learning capabilities are investigated [5],[6] and aspects to reduce complexity are developed [4].

D. Intrinsic Safety Knowledge

Risk can be understand as ratio of hazard and safeguards, and *'"safeguards" is the idea of simple awareness. That is, awareness of risk reduces risk. Thus, if we know there is a hole in the road around the corner, it poses less risk to us than if we zip around not knowing about it'*[8].

It is assumed that awareness of hazards is generated with the help of knowledge about hazards. The question is how to transfer such knowledge to an autonomous system. It may be answered by a computer scientist in introducing learning algorithms. With introducing safety-critical systems with learning capabilities, a 'chicken-and-egg' dilemma arises. If a system is assumed as 'tabula rasa' at beginning of its operating time, the system is not able to avoid dangerous situations intrinsically in order to keep in a safe state. It may learn after more or less trials to avoid such dangerous states only by taking dangerous actions. Finally, it will 'converge' to take safer actions in the average by learning more or less abstract concepts of safety. But even if a system would have learned an adequate safety concept there is no comprehensible insight into the system in order to check how the concept looks like or how it is realized. According to standards and directives [IEC61507, DO-178B], at least documented safety assuring procedure is required. Therefore, a 'black-box' learning approach would be not acceptable.

On the other hand, safety assuring procedures during development phase of a system requires the complete definition of occurring hazards in form of a hazard analysis. Typically, the environment of a service robot, which could or should possibly be delivered to any arbitrary household, is complex and not completely known; these are two challenges that have to be taken into account when assuring safety for mobile service robots.

The complexity problem can be possibly circumvented when a safety assuring module is assumed which basically enables the refinement of the knowledge about dangerous situations, in the sense of above-mentioned learning approach. The resulting process could be seen as one that is converging to safety.

As long as a 'tabula rasa' solution is not satisfying, some initial definitions of dangerous situations (alias safety knowledge) have to be included a-priori. This initial safety knowledge has at least three additional functions (in addition to traditional safety assuring procedure):

- to guarantee a certain safety during early operation phases by being realized in a general and conservative manner,
- to develop and introduce certain measures to describe hazards and risks, and
- to construct and maintain a comprehensible symbolic representation of the safety knowledge to enable documentation, debugging, verification and transferring to other systems.

II. Formalization of Safety Knowledge

Basically, the presented real-time safety assurance method requires a system that is capable to perceive its environment in order to generate an internal representation, store and manage experiences of environmental interactions in a knowledge representation. The SOM-approach is used for this purpose and therefore, it is detailed in the sequel.

A. Knowledge Representation in SOM

The Situation-Operator-Modeling (SOM) approach describes the real world using the introduced termini 'situations' and 'operators', modeling scenes and actions of the real world. The situations are time-fixed and event discrete descriptions of 'moments' of real world. Changes in the considered systems are denoted with so-called operators. A changing world results in situation operator chains. The situation itself consists of characteristics c_i and relations r_i.

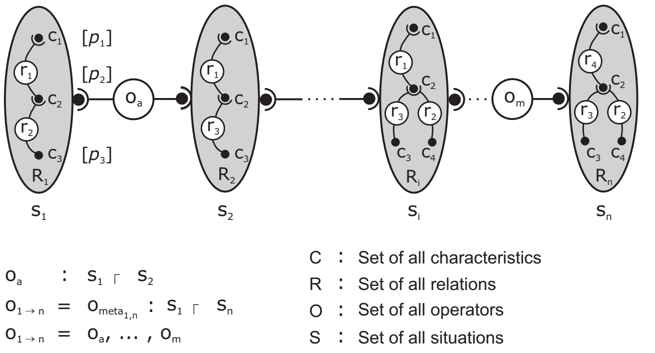


Fig. 1. Graphical representation of a situation operator sequence denoting the modeling of changes within the real world [5].

The relations represent an inner structure of the situation, which allows the linking of characteristics to each other through arbitrary functions [5]. The relations are from the same quality as the operators. Whereby the operators transfer situations to sequel situations and relations are used to abstract further characteristics. Characteristics can basically be measured physical values or higher abstracted information. The SOM approach comes with a graphical representation (see Fig. 1). The situation is denoted with gray ellipses, the black dots represent the characteristics. The white circles are relations when located in the gray ellipse of a situation or operators when connecting situations.

In simplified terms, the SOM technique can be used to construct a cognitive architecture that is capable to describe the perception, abstraction, storage of experiences and learning with a unique homogeneous SOM-based notation. Furthermore, the whole information processing is based on the same methodical background [5].

The stored experiences, in the so-called experience database consists of sets of initial situation, operator and final situation 'prototypes'. These can be used to generate a set of future reachable situations, the so-called action space. The action space can be uses for planning purposes [4].

A detailed description about the mentioned approach is given in [12],[13].

B. Risk-sensitive Planning and Risk Perception

The SOM-approach is used as the overall concept, used to create an architecture (see [5]) for an autonomous system (service robot). It provides functions like perception, learning, reduction of complexity, planning, etc. Additional components have to be embedded in the architecture, which are responsible to assure the safety of the overall system. The focus of the addressed safety concerns are hazards caused by manipulating objects as described in I-B.

Hence, two functions seem to be central in this context. The first important function is the planning capability. The planning component defines actions that are performed in order to reach a given goal. With the help of prediction capabilities of a CTS, actions with low or no risk are preferred in order to keep the system in a safe state. The perception is seen as a second important component because the perception is connected to the information selection (for example with respect to current goal or sub goals). Is has to be assured that present risks are perceived.

An advantage of the SOM-based architecture is the use of homogeneous notations for all cognitive functions. If the risk information extraction process is realized in SOM manner it can be used to assess risks for planning as well as for perception. The 'relation' is the notation to describe internal structures (abstraction) of situations. Hence, the notion 'relation' can be used to integrate risk assessment in the cognitive architecture.

C. Relations

In this presentation risks are focused, that can arise through interaction of environmental objects. Risks to be distinguished are assumed to be situational inherent and therefore, detectable by analyzing the inner structure of a situation. Relations are used in the SOM-technique to describe inner structures of situations. Hence, the core idea of this approach is to use a-priori 'artificially' generated relations in order to derive risk information. This method ensures that safety knowledge is consistent to the overall CTS. This basically enables the integration of the safety knowledge into the experience database. That in turn generally enables the refinement of the safety knowledge by the cognitive system itself what is essential because the initial safety knowledge can be hardly described completely during the development phase of the robotic system. The refinement of safety knowledge is not addressed in this contribution, but it should be structurally taken into account.

The 'relation' in SOM notation has the same structure as an 'operator' and is also called 'passive operator' (internal causal relation between characteristics: in terms of 'because' [13]). A relation is applicable to a situation if its so-called explicit assumptions eA_x are fulfilled. A certain relation requires the presents of certain characteristics. The required characteristic can be seen as a condition for its application and as inputs of the relation function. On the basis of the inputs (and parameters as implicit assumptions iA_x) the relation generates new (abstracted) characteristics c, for example like a mathematical function, $c = f(eA_1 \ldots eA_i, iA_1 \ldots iA_j)$, so the notion can be used to formalize the safety knowledge. Hence, the a-priori risk analysis results by risk computational rules or principles. These rules or principles in turn are formalized in relation notion. The set of produced risk assessment relations is called safety knowledge.

In a first step the formalization of hazard causes (if-then constructs) are realized with the help of the conditional applicability of relations. For example, the appearance of characteristic 'A' and 'B' indicates risk 'C' then the relation generating characteristic 'C' awaits the presence of characteristic 'A' and 'B'. A further problem is to express the risk in comparable measures. Therefore, the risk quantification problem is discussed in the sequel.

III. Quantification

The objective of integrating risk assessment capabilities in robotic systems is to realize safe autonomy. This requires deliberative decision-making capabilities at least at a higher systemic level.

'Rational decision-making requires, therefore, a clear and quantitative way of expressing risk so that it can be properly weighed, along with all other costs and benefits, in the decision process'[8].

The term 'properly weighed' in relation to risks, costs and benefits requires for comparable measures. A set of decision alternatives, equipped with such comparable parameters can be dealt with the help of decision theory. To reach comparability it is assumed that a standardization to absolute values in the range of

$0 \leq risks, costs, benefits, severity, probability, \ldots \leq 1$

is appropriate.

A. Quantification of Risks

Within the universal definition of risk the three questions 'what can happen', 'how likely is that' and 'what are the consequences' [8] are adopted for presented robotic risk assessment approach. Each answer results to a triplet $< S_i, p_i(\phi_i), p_i(X_i) >$, which describes the likelihood $p_i(\phi_i)$ and the consequence $p_i(X_i)$ of the scenario S_i [8]. Precise numbers for consequences and likelihoods are very are difficult to derive, thus, in [7],[8] is introduced the 'Evidence-based Approach', which is shortly described in the sequel.

B. Modeling of Consequences and Likelihoods

A key problem of describing risks numerically is that these most often cannot be determined with absolute precision because it is a subjective thing and relative to the observer [8]. Therefore, it is suggested to express such 'vague' issues with probability distributions. These are generated with the help of the 'Evidence-based Approach' [7]. Hence, the probability density of the numerical expression changes in accordance to the amount of considered evidence. The more evidence is available the more precise numerical expression can be processed with the help of Bayes' theorem. Hence, with [8],[7] an approach is given that enables the quantification of risks methodically. The overall process is shown in Fig. 2 which is strongly oriented on the graph in [7]. Initially, example scenarios are defined; causal conditions are extracted and supplemented with quantified consequences and likelihoods. This information is packaged in the 'safety principles' in an adequate structure as mentioned in II-C. This set of 'safety principles' (safety knowledge) is the starting point of the system's operation time. The current interpretation of the internal representation of the outside world in form of situations (see II-A) is presented to the risk assessment module. This module in turn evaluates situation on the basis of the safety principles, respectively. The outcome of the risk assessment module is the derived risk information. That in turn is an important base for decision-making.

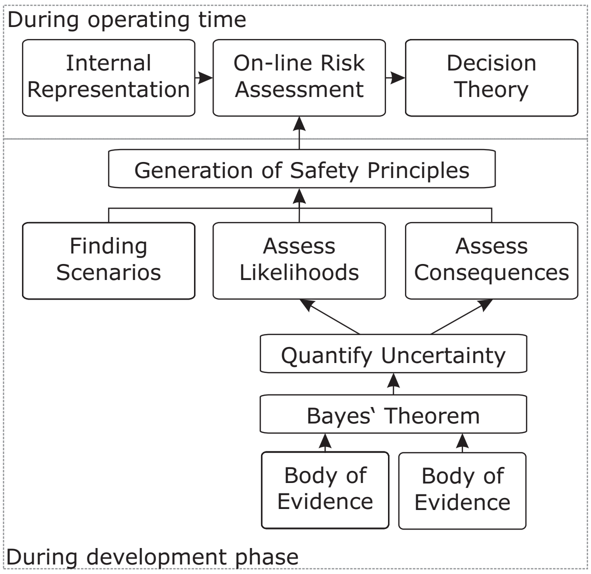


Fig. 2. Variation of the risk quantification process for the on-line risk assessment in the style of [7]

IV. Experiment

A. Thought Experiments

In the sequel are depicted three thought experiments (TE) in order to state risks that could possibly occur with respect to interaction of objects in the domain of service robots more precisely. Basically, causal and temporal relations of the described examples can be decoded very precisely, but the following experiments are kept simple with the intent to show a conceptual solution on the one hand and on the other hand to realize generalized rules (or due to this reason so-called principles). The 'conservative concept' of these principles is to over-generalize and simplify hazard causes with the objective to accept false alarms (false risk detections) in principle and to avoid missed alarms.

An object recognition system is assumed to be available as an under-laying system unit. It is assumed that the object recognition is powerful enough to recognize all objects in accurate time and no information filtering takes place. Furthermore, the output of the object recognition module is assumed be to a list. The list contains recognized objects and their position or even size and pose. For each recognized object there exists a sub-list which contains possible object identifiers and the identification probability, respectively. The object identifiers are assumed to be known natural language descriptions. The summed up probability P_i of each object's identification probability P_{ij} is assumed to be $P_i = \sum P_{ij} \leq 1$. The remaining difference $P_{uncertainty,i} = 1 - P_i$ is assumed to decode uncertainty.

1) TE1: Kitchen Stove: A service robot is instructed to bring the dished to the kitchen sink. In order to deposit the dished close to the sink, any module recognizes the modern ceramic stove top as preferable surface. If a cooking plate is still hot and there is some plastic bowl or other plastic dishes, the risk of toxic vapor or fire by inflamed plastic arises. In this case, the robot's acting provokes unacceptable risks and therefore it is not safe anymore [3]. To avoid such situations (in order to make the robot functional safe), a rule (principle) has to be implemented which states that a plastic bowl (plastic (or wooden...) object) is not allowed to be placed on top of (close to) a stove (heat source). There are several possibilities, for example the position of the cooking plate knobs, the measured plate temperature or even the daytime can be taken into account. These possibilities are not wrong, but nevertheless difficult to be realized robustly. The assumption, the cooking plate is potentially hot is much more easy, not fully correct, but safe.

The involved factors on which the situational risk depends are described as follows: The object recognition module recognizes the 'cooking plate' with a probability P_{CP} and as mentioned above, the cooking plate is assumed to be hot, so $P_{CP} = P_{HeatSouce}$. The 'bowl' is recognized with the probability P_{bowl} and it is assumed that all bowls are made of plastic, so $P_{Bowl} = P_{PlasticObject}$. The worst case accident of such situations is assumed to be a residential fire with the consequence of human injuries. The severity value is taken from the 'Hicks Scale' (see e.g. [11]), for example '0.15' ('Minor to major injuries, medical aid and low severity impairment'). The severity value is assumed to range between '0...1' - therefore, the values of the 'Hicks Scale' ('0...100') are normalized. Besides the pure availability of a plastic object and a heat source in a situation, further dependencies for determining an accidental risk are important. For example, the distance between both objects plays a central role. Therefore, it is assumed that there is a dependence between the spatial distance and the probability of a heat accident in terms of $P_{distance} = f_P(distance)$. This function can be approximated, for example, by linear interpolation of some known (extreme) values.

The description of the risk, in accordance to the triplet in [8] is $< S_i, p_i(\phi_i), p_i(X_i) >$ in which $p_i(\phi_i)$ describes the likelihood and $p_i(X_i)$ the consequence of the scenario S_i. In this example this is $< S_{heat+plastic}, P = P_{HeatSouce} \cdot P_{PlasticObject} \cdot P(distance), 0.15 >$.

The remaining probability difference, which takes unidentified objects into account, is treated separately. With regard to safety aspects, it is reasonable to assume them as dangerous objects with a very high accident severity, for example resulting in the following risk triplet notation $< S_{unknown}, P_{unknown} = 1 - (P_{HeatSouce} \cdot P_{PlasticObject}), 1 >$.

2) TE2: Watering the Power Plug: The second experimental scenario is derived from a task 'watering the plants'. Here, it is assumed that a power plug fell into a plant pot. If the robot is watering the plant, the risk of electrical shock arises, both, for human and robot. These risk factors can be regarded as follows: The object recognition recognized the power plug with a probability of P_{Plug} and knows with the probability of P_{Water} that there is water in the watering can. Finally, it is not really true, but safe, if the approaching of the robot with water ('liquid container') to a power plug ('dangerous electric device') would be forbidden by a corresponding rule (principle). The modeling of the accident severity could take place by consideration of electrical accident statistics or again by using the 'Hicks Scale' (Severity '0.6' according to 'Single fatality, permanent total disability'). The observing of the water jet striking the power plug is difficult to realize. Therefore, it would be a conservative approximation to avoid the approaching of water to dangerous electricity in principle. The risk triplet is $< S_{water+electr.}, P = P_{Plug} \cdot P_{water} \cdot P(Distance), 1 >$.

3) TE3: Gripping a Gun: The third experiment deals with a situation in which the robot is instructed to grip a gun. In the case of usual domestic service robots, the estimation of mentioned risk is very simple: The possible accident outcome is assumed to be at least the death of one human. Therefore, the severity is assumed to be 1' according to 'Hicks Scale' ('Multiple fatalities and injuries'). The probability of such accident can be taken from the statistic about average death rate caused by guns, for example $P_{Gun} = 0.001$[1] or it can be set to $P_{Gun} = 1$ with the intent to emphasize the uselessness of guns. Therefore, the single gripping results in a risk triple $< S_{grippedGun}, 1, 1 >$.

4) Conclusions: The quantification of accident probabilities and severities remains difficult. A conservative selection of parameters in problematic cases may cause standstill of the robot. Under assumption that a service robot in a domestic environment is not entrusted with tasks, whose failing are critical, it is acceptable that the robot's control refuses performing of tasks when insufficient knowledge (about objects) is available. Thus, a required refinement or extension of the safety knowledge could possibly take place under human supervision.

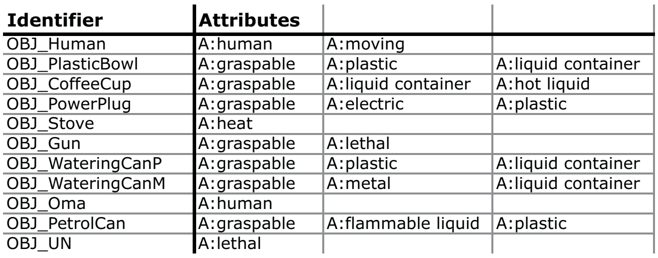
B. Experimental Simulation

For the proof of presented concept a small simulation environment was created. The simulator generates an environment in order to investigate the results of the described risk assessment module. Therefore, a scene in form of 2D-wold is made available, containing a robot and several environmental objects. The robot can be moved, the robot is able to grip objects. The scene can be changed by moving the robot with a gripped object. The object recognition module is 'simulated' by manually assigning several identities with different probabilities to the objects, respectively. This object recognition module response can be manually changed by the user. Furthermore, the object recognition

[1] Average death rate concerning accident with guns in Germany, see `http://www.fhvr-berlin.de/fhvr/fileadmin/content/publikation/heft48.pdf`, May 2010.

TABLE I

LIST OF OBJECTS IN THE OBJECT DATABASE. EACH OBJECT HAS SEVERAL ATTRIBUTES.



Identifier	Attributes		
OBJ_Human	A:human	A:moving	
OBJ_PlasticBowl	A:graspable	A:plastic	A:liquid container
OBJ_CoffeeCup	A:graspable	A:liquid container	A:hot liquid
OBJ_PowerPlug	A:graspable	A:electric	A:plastic
OBJ_Stove	A:heat		
OBJ_Gun	A:graspable	A:lethal	
OBJ_WateringCanP	A:graspable	A:plastic	A:liquid container
OBJ_WateringCanM	A:graspable	A:metal	A:liquid container
OBJ_Oma	A:human		
OBJ_PetrolCan	A:graspable	A:flammable liquid	A:plastic
OBJ_UN	A:lethal		

uses natural language identifiers for the recognized objects. Further information from the object recognition, like position or size of objects and information about the internal state of the robot, for example position and speed, are summarized in a time-fixed and event discrete situational description. Hence, the current situation is examined with respect to risks. Furthermore, the objects are classified with respect to attributes in an object database. The object 'watering can' can have associated attributes like 'liquid container', 'plastic' or 'metal'. For simplification, all attributes are connected with respective objects with probability '1'.

The third required input of the risk assessment module is the set of safety principles. The principles are divided into three parts. The first is the conditional part (explicit assumption eA_x, see II-C). With this conditional part is defined when the respective principle is applicable. For this example that is the case when respective objects are present in the current situation. In order to simplify the conditional part, it is limited to two objects - one environmental and one gripped object. Generally, more complex logical connections are possible. The condition can be formulated either with object names or with object attributes. The second part of the principle is the severity estimation instruction. In this instruction the severity value is defined. Therefore, the severity can be either computed with the help of additional situational information (relation) or it can be parameterized with values. In this experiment only fixed parameters are used. The third part of the principle is the determination instruction for the probability of accident occurrence. The concept is similar to the severity determination: the probability can be either a static parameter or it can be derived from the situational context.

C. Experiment and Results

The experiment was made in the mentioned simulation environment with a selection of objects. A set of attributes may be connected to each object (see Table I). The objects are placed in the 2D world (see Fig. 3), the robot can be moved and one object can be moved with the help of the robot by gripping it. While the user changes the scene, the risk assessment determines the current risk. The output of the risk assessment module is a risk matrix. In accordance to [8] it consists of a list of risk triplets. Each triplet represents a hazardous scenario. In order to generate a risk curve as an adequate description of risk, the scenarios are ordered starting with the scenario with the highest severity. The probability values are summed up [8].

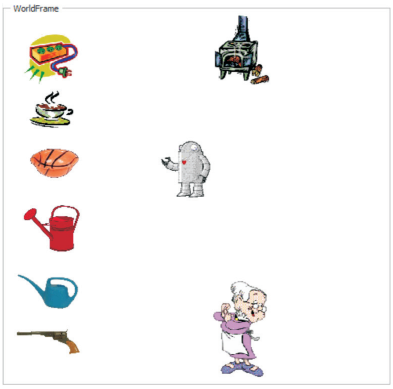


Fig. 3. The 2D world, including typical objects and the robot.

The implemented safety principles are shown in Table II. In the first experiment the robot was moved near the stove. According to principle '7', there is a risk in dependence of the robot's distance ('@' represents the robot) to the object (stove) with the attribute 'heat'. The resulting risk curve is shown in diagram a) in Fig. 4. A risk with the severity of 0.02 is present as parameterized in the receptive principle. Furthermore, it has the probability of occurrence 0.35 in accordance to the linear function that specified in the principle.

If the robot remains at the same position, but has gripped the bowl with the attribute 'plastic' then principle '2' generates a additional risk contribution to the graph; the peak at severity of 0.15 shown in diagram b) in Fig. 4.

The diagram c) shows the risk curve when the robot has gripped the red watering can. The risk curve contains additional risks because the object recognition identification of the red watering can is 'WateringCanM' with $P_1 = 90\%$ and therefore the attributes 'metal', 'liquid container' (according to table I). Furthermore, the identified object could be 'WateringCanP' with $P_2 = 7\%$ with the attributes 'plastic', 'liquid container' or it could be a 'PetrolCan' with probability $P_3 = 2\%$ and the attributes 'flammable liquid', 'plastic', 'liquid container'. Finally, the remaining probability of $P_4 = 1\%$ is uncertainty and therefore treated as unknown object with the attribute 'lethal'. The risk contribution of the unknown object is the risk probability with severity '1' coming from principle '15'. The principle '11' is responsible for the risk contribution at point '0.7' of the severity axis in the case when 'flammable liquids' are approached to 'heat sources'. The remaining risks result from the above-mentioned principles (robot or a 'plastic object' is approached to a 'heat source').

In diagram e) in Fig. 4 the robot is shown again at the same position with the blue watering can. The object recognition delivers the output 'WateringCanP', $P_1 = 90\%$ or 'WateringCanM', $P_2 = 7\%$ or 'PetrolCan', $P_3 = 2\%$ and again 'unknown' with $P_4 = 1\%$. The safety principles that are active are the same as in the latter experiment. The safety principle '2' for attribute 'plastic' and 'heat' generates a high risk contribution, similar to diagram b).

The last diagram f) shows the risk curve when the robot grips a gun in presence of a human and in accordance to principle '16'. This principle is designed to inhibit the robot to grip a gun in presence of a human and therefore, it generates high severity and probability values.

D. Discussion

The values for severity and probability of hazards and the conditions are simplified. Nevertheless, such exemplary values enable a risk assessment, which is plausible. The proposed strategy especially for complex cases is to inhibit dangerous situations in a very general manner. The refinement and extension of the principles can take place in further steps, for example by learning, by adding tasks skills, by exchange of knowledge between robots or by debugging the safety knowledge manually. The possibility for traceable debugging seems to be very important, especially when refinement of the safety knowledge by learning is allowed. The comprehensible a-priori safety knowledge-base provides a-priori safety. On the other hand, this knowledge is stored in a dynamic manner, so that adoption due to learning capabilities is not excluded. The natural language basis of the a-priori safety knowledge assures that further knowledge is maintained in a comprehensible way.

A further challenge which is related to complex safety-critical systems is the certification process. The efforts that have to spent in safety assuring mechanism depend on required safety integrity levels (SIL). Proposed approach will not be appropriate for highly safety-critical systems and of course the traditional safety analysis cannot be replaced. Nevertheless, due to the modularity , intensive testing and evaluation of the risk assessment module software is possible. For generating and debugging of the safety knowledge-base additional tools have to be developed. The verification of the safety aspects possibly takes place by examining or inspecting the current state of the safety knowledge-base. The critical information processing paths (information source, sensors) can be derived from the safety knowledge by checking the principle's input, for example.

A safety guarantee in form of a formal proof should be hardly possible due to the high complexity. Finally, this safety assuring approach is based on a (well documented) carefully developed mechatronic system. It is assumed that debug-able and maintainable safety knowledge which causes safety limitations in additional safety assuring modules (active safety measures) plays a key role on the way to safe service robots.

V. Conclusion

In this contribution an approach for realization of safe autonomous robot interaction is presented. The central idea is to embed dynamic risk assessment into a cognitive technical system (CTS) in order to assess the CTS's internal representation of the real world. It is shown how a-priori safety knowledge can be integrated dynamically, permitting modifications in principle in order to refine and extend it. The formalization of safety knowledge with the help of the Situation-Operator-Modeling (SOM) approach is pointed out. A simulation experiment shows the output of a risk assessment module, which can be directly used for a subsequent decision making processes.

References

[1] "IFR international federation of robotics, a definition of service robot," http://www.ifr.org/service-robots/, May 2010.

[2] E. Ahle and D. Söffker, "A Cognitive-Oriented architecture to realize autonomous behavior - part II: application to mobile robotics," in *Systems, Man and Cybernetics, 2006. SMC '06. IEEE International Conference on*, vol. 3, Oct. 2006, pp. 2221–2227.

[3] J. Börcsök, *Functional Safety: Basic Principles of Safety-related Systems*. Heidelberg: Hüthig, 2007.

[4] D. Gamrad, H. Oberheid, and D. Söffker, "Automated detection of human errors based on multiple partial state spaces," in *6th Vienna Conference on Mathematical Modeling on Dynamical Systems MATHMOD*, Vienna, Austrialia, 2009.

[5] D. Gamrad and D. Söffker, "Architecture for cognitive technical systems allowing learning from interaction with unknown environments." in *7th Workshop on Advanced Control and Diagnosis*, Zielona Góra, PL, 2009.

[6] ——, "Simulation of learning and planning by a novel architecture for cognitive technical systems," San Antonio, Texas, USA, 2009, pp. 2302–2307.

[7] S. Kaplan, "The words of risk analysis," *Risk Analysis*, vol. 17, no. 4, pp. 407–417, 1997.

[8] S. Kaplan and Garrick, "On the quantitative definition of risk," *Risk Analysis*, vol. 1, no. 1, pp. 27–37, 1981.

[9] C. Matheus, M. Kokar, and K. Baclawski, "A core ontology for situation awareness," in *Proceedings of the Sixth International Conference ofInformation Fusion*, vol. 1, 2003, pp. 545–552.

[10] M. Prats, E. Martínez, P. J. Sanz, and A. P. del Pobil, "The UJI librarian robot," in *Intelligent Service Robotics*, vol. 1, 2008, pp. 321–335.

[11] D. Proske, *Catalogue of risks natural, technical, social and health risks*, 16th ed. Berlin,London: Springer, 2008.

[12] D. Söffker, *Systemtheoretic Modeling of the knowledge-guided Human-Machine-Interaction (In German)*, ser. Habilitation Thesis, University of Wuppertal, 2001. Berlin: Logos Wissenschaftsverlag, 2003.

[13] ——, "Interaction of intelligent and autonomous systems - part 1: Qualitative structuring of interaction," *Mathematical and Computer Modelling of Dynamical Systems*, vol. 14, pp. 303–318, 2008.

[14] D. Söffker and E. Ahle, "Idea, conception, and realization of learning abilities for robot control using a Situation-Operator- model," in *International Workshop on Automatic Learning and Real-Time, ALaRT 2005, September 7-8, Siegen, Germany, Proceedings*, K. Kuhnert and M. Stommel, Eds. University Siegen, Germany, 2005, pp. 131–141.

[15] S. S. Srinivasa, D. Ferguson, C. J. Helfrich, D. Berenson, A. Collet, R. Diankov, G. Gallagher, G. Hollinger, J. Kuffner, and M. V. Weghe, "HERB: a home exploring robotic butler," *Autonomous Robots*, vol. 28, no. 1, pp. 5–20, 2009.

[16] A. Wardziński, "The role of situation awareness in assuring safety of autonomous vehicles," in *Computer Safety, Reliability, and Security*, 2006, pp. 205–218.

TABLE II

LIST OF APPLIED SAFETY PRINCIPLES, CONSISTING OF THREE PARTS, THE CONDITIONAL, SEVERITY ESTIMATE AND PROBABILITY ESTIMATE PART, SEVERITY PART AND PROBABILITY PART CAN BE DEFINED IN DEPENDENCE OF ANY INPUT CHANNEL WITH THE HELP OF INCLUDED PARAMETRIZABLE STANDARD FUNCTIONS.

	Conditional Part		Severity Part				Probability Part					
Nr	Object 1 (Robot/carried)	Object 2 (in environment)	Principle	Input	Sev. 1 (0...1)	@Val1 of Inp.	Principle	Input	Prob. 1 (0...1)	@Val1 of Inp.	Prob. 2 (0...1)	@Val2 of Inp.
1	A:hot liquid	A:human	NONE	NONE	0.1		LinearFunction	distance	1	20	0	70
1	*		NONE	NONE			LinearFunction	speed	1	2	0	0.3
2	A:plastic	A:heat	NONE	NONE	0.15		LinearFunction	distance	1	10	0	50
3	A:chemical	A:human	NONE	NONE	0.6		StepMFunction	distance	1	30		
4	@	A:human	NONE	NONE	0.02		LinearFunction	distance	1	0	0.1	70
4	*		NONE	NONE			LinearFunction	speed	1	2	0	0
5	A:liquid container	A:electric	NONE	NONE	0.6		LinearFunction	distance	1	0	0	50
6	A:electric	A:liquid container	NONE	NONE	0.6		LinearFunction	distance	1	10	0	100
6	*		NONE	NONE			LinearFunction	speed	1	1	0	0.3
7	@	A:heat	NONE	NONE	0.02		LinearFunction	distance	1	10	0	60
8	A:heat	A:plastic	NONE	NONE	0.15		LinearFunction	distance	1	10	0	30
9	A:electric	A:human	NONE	NONE	0.6		LinearFunction	distance	1	50	0	100
10	@	_	NONE	NONE	0.01		LinearFunction	distance	1	10	0	40
10	*		NONE	NONE			LinearFunction	speed	1	2	0	0
11	A:flammable liquid	A:heat	NONE	NONE	0.7		LinearFunction	distance	1	20	0	100
12	A:heat	A:flammable liquid	NONE	NONE	0.7		LinearFunction	distance	1	20	0	100
13	A:flammable liquid	A:electric	NONE	NONE	0.7		LinearFunction	distance	1	0	0	50
14	A:electric	A:flammable liquid	NONE	NONE	0.7		LinearFunction	distance	1	0	0	50
15	OBJ_UN	_	NONE	NONE	1		StepMFunction	NONE	1	1		
16	OBJ_Gun	A:human	NONE	NONE	1		StepMFunction	NONE	1	1		

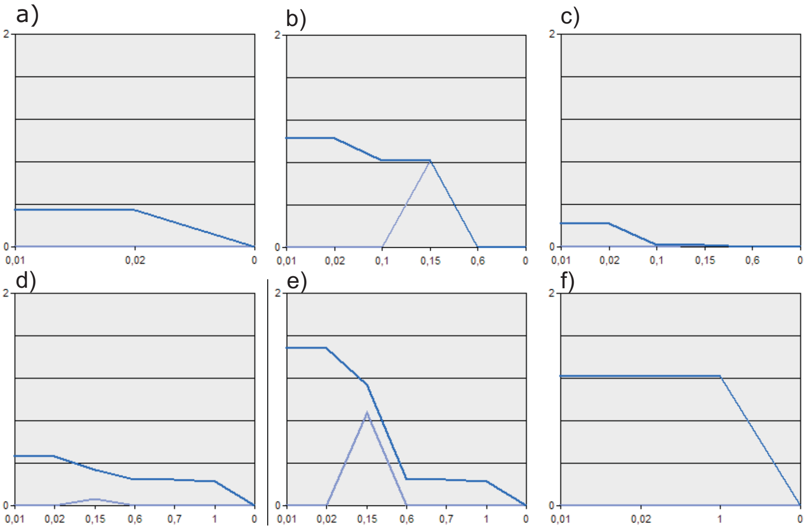


Fig. 4. Risk curve of several experiments. a) Robot close to heat source b) same position with a plastic bowl gripped c) with a coffee cup gripped d) with a metal watering can e) with a plastic watering can f) with a gun in the gripper. The danger scenarios are plotted on the x-axis, ordered by respective severity while the respective probabilities are accumulated on the y-axis.