



**HAL**  
open science

# Give MEANinGS to Robots with Kitchen Clash: A VR Human Computation Serious Game for World Knowledge Accumulation

Johannes Pfau, Robert Porzel, Mihai Pomarlan, Vanja Sophie Cangalovic, Supara Grudpan, Sebastian Höffner, John Bateman, Rainer Malaka

## ► To cite this version:

Johannes Pfau, Robert Porzel, Mihai Pomarlan, Vanja Sophie Cangalovic, Supara Grudpan, et al.. Give MEANinGS to Robots with Kitchen Clash: A VR Human Computation Serious Game for World Knowledge Accumulation. 1st Joint International Conference on Entertainment Computing and Serious Games (ICEC-JCSG), Nov 2019, Arequipa, Peru. pp.85-96, 10.1007/978-3-030-34644-7\_7. hal-03652050

**HAL Id: hal-03652050**

**<https://inria.hal.science/hal-03652050>**

Submitted on 26 Apr 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

# Give MEANinGS to Robots with Kitchen Clash: A VR Human Computation Serious Game for World Knowledge Accumulation

Johannes Pfau<sup>1</sup>, Robert Porzel<sup>1</sup>, Mihai Pomarlan<sup>1</sup>, Vanja Sophie Cangalovic<sup>1</sup>,  
Supara Grudpan<sup>1</sup>, Sebastian Höffner<sup>1</sup>, John Bateman<sup>1</sup>, and Rainer Malaka<sup>1</sup>

<sup>1</sup>University of Bremen, Bibliothekstraße 1, 28359 Bremen, Germany  
{jpfau,porzel,pomarlan,vanja,gud,shoeffner,bateman,malaka}@uni-bremen.de

**Abstract.** In this paper, we introduce the framework of MEANinGS for the semi-autonomous accumulation of world knowledge for robots. Where manual aggregation is inefficient and prone to incompleteness and autonomous approaches suffer from underspecified information, we deploy the human computation game *Kitchen Clash* and give evidence of its efficiency, completeness and motivation potential.

**Keywords:** Serious Game · Knowledge accumulation · Framework

## 1 Introduction

Robotic proficiency excels in well-defined tasks and environments [12, 1, 4], but fails in compensating for missing or too generic information. Human-level world knowledge has been shown to close the reasoning gap [4], yet teaching robots this kind of knowledge remains one of the most challenging tasks for robotic AI research, since autonomous approaches end up with underspecified information and manual accumulation results in incalculable effort. In this paper, we introduce *Kitchen Clash*, a VR human computation serious game for the extraction of human world knowledge in the context of everyday activities. Within the framework of MEANinGS (MALLEATING EVERYDAY ACTIVITY NARRATIVES IN GAMES AND SIMULATIONS), we integrate a combination of information-transforming modules that include finding a proper set of instructions for a given complex task, processing these syntactically as well as semantically to detect underspecified information, autonomously generating testbed scenarios including a variety of decision making affordances and finally solving world knowledge problems by human computation through a serious game aided by physical simulation. In an explorative pilot study, we assessed user experience, appraisal and the overall viability of the presented serious game to report on the findings and demonstrate the feasibility of the approach. To constitute a baseline condition, we evaluated these findings against a control group executing manual knowledge accumulation, resulting in higher efficiency, increased motivation and considerably higher information retrieval. This paper contributes to the community of serious game

research, presenting a successful application advantageous to a real-world problem solving field, as well as to the community of robotic research, exemplifying the practicability of a novel framework to overcome underspecified knowledge.

## 2 Related Work

One of the earliest research programs to study autonomous robots was the Shakey project [12]. Shakey was a mobile robot that used planning to reason about its actions and performed tasks that required planning of paths and actions as well as re-arranging of simple objects. This work was seminal for the fields of classical planning and computer vision. Nevertheless, even in Shakey’s simple environment, the limitations of the approach became clear, as the computational complexity of planning problems proved, in general, to be intractable.

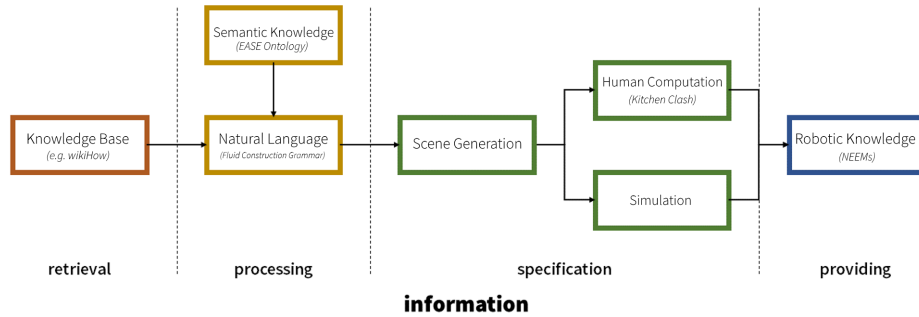
Many researchers [4, 12, 11, 13, 18] have worked on providing robotic systems with human-like common sense knowledge so that the robots could, hopefully, avoid costly planning from scratch or trial and error. Dang et al.[3] proposed a method to teach a robot to manipulate everyday objects through human demonstration. The authors asked participants to put on motion capture suits and perform tasks, such as opening a microwave or a slide door, and recorded 3D marker trajectories. These trajectories were used as the input for a chain learning algorithm. Parde et al.[13] developed a method to train robots to learn the world around them by using interactive dialogue and virtual games. The game asked a human player to put some objects in front of the robot and challenge it to guess which object the user has in mind. Through many gameplay sessions, the robot learns about objects and features which describe them and associates these with newly captured training images. Beetz et al.[2, 1] proposed the software toolbox for design, implementation, and the deployment of cognition-enabled autonomous robots to perform everyday manipulation activities. To teach the robot, they use a marker-less motion capture system to record human activity data, which is then stored as experience data for improving manipulation program parameters. Programs, object data, and experience logs are uploaded to the openEASE web server, from which they can be retrieved as needed to extend the task repertoires and objects that a robot can recognize.

The representations needed for action knowledge have also been a topic of research, because the symbolic, highly abstract, “actions as black boxes” representations of the Shakey era do not result in robust behaviors in a realistic environment. In general, action knowledge tends to be subsymbolic, and often takes the form of success/failure probability distributions over an action’s parameter space [17, 19]. Note that the experience the robot learns from doesn’t have to be from the real world. Simulated episodes, produced either by a human player of a game or a robot simulating itself, can be used for this purpose. Simulation will of course not provide a complete description of a realistic action, but even very coarse simulation can already be useful for a robot that needs to validate its plans and/or pick a better set of parameters [8].

In our work, we propose MEANinGS to use a VR human computation serious game to simulate real-world tasks in realistic environments and situations. Recorded trajectories can be translated to real-world robotic movements which are spatially less constrained than motion capturing approaches and accumulated world knowledge can help overcoming underspecified information, which has the potential to reduce planning computation considerably. Similar to the aforementioned approach, we contribute to the field of cumulative robotic knowledge by adding resulting symbolical and subsymbolical insights to the openEASE repertoire.

### 3 Implementation

#### 3.1 Framework



**Fig. 1.** Flowchart of the introduced MEANinGS framework.

Figure 1 demonstrates the flow of information, as well as the impact of and interaction between the particular modules. MEANinGS originated in the context of everyday household activities and focuses on knowledge accumulation in this area, while its functionality is not limited to the application field. Offering an interface to any natural language based instruction set (**retrieval**), the contained information is **processed** in order to represent subtasks as tuples of manipulative actions and objects acted on or with. When utilizing natural language, the ontological scope of these objects is often heavily underspecified, since humans are used to working with generalized information and specifying these in terms of individual choice, influenced by world knowledge, availability and

preference. Yet, this underspecification does not render all possible objects contained in a general term as viable (or usual). Thus, in the **specification** layer, human knowledge is added through *Kitchen Clash*, a serious game presenting a decision making paradigm within this set of objects. Parallel to this, complementing world knowledge is derived from physical properties of the objects and their surroundings via simulation. In both approaches, object choices can be quantified and thus ranked by efficiency and effectiveness. Within the simulation, this assessment can be realized in a fully autonomous manner, while the serious game offers further qualitative insights, since peer-rated quality measurements are included in the rating process, as well as preference and conventionality measures. Eventually, world knowledge is aggregated with trajectorial and contextual information and **provided** as *narrative-enabled episodic memories* (*NEEMs*) according to the KnowRob [1] paradigm.

### 3.2 Knowledge Base

In order to have a generic, comprehensive framework that is capable of adapting to human data input, our approach does not rely on a single knowledge base, but is designed to handle any set of natural language instructions that are goal-oriented and describe the most crucial subtasks sequentially or hierarchically. In this way, we introduce an interface that manages to grasp verbal commands equally effective as cooking recipes or tutorial websites (e.g. wikiHow). After retrieving the document encompassing the entire task completion, subgoals are derived from the contained sentences or steps and processed in the next module, independently of each other.

### 3.3 Natural Language Processing

In order to flexibly handle natural language input consisting of abstract and underspecified instructions, a deep semantic parser based on the Fluid Construction Grammar formalism [16] is used. Both the lexicon and the analysis itself make use of ontological knowledge described in Section 3.4, to guide the extensive search process, disambiguate otherwise unclear instructions and evoke unspecified parameters which need to be inferred by later processing steps of MEANinGS. In this way, natural language commands are transformed into a series of desired actions, accompanied by their parameters and respective pre- and post-conditions.

### 3.4 Ontology

The semantics of the actions and entities involved in the game are defined by a formal ontology. This ontology is designed to provide descriptions for everyday activities in terms of human physiology and human mental concepts, as well as enabling formal reasoning. The ontology supplying the labels for the objects has been designed using the principles proposed by Masolo et al. and is created

by using the DOLCE+DnS Ultralite ontology (DUL) as an overarching foundational framework [6, 5]. Specific branches of the KnowRob knowledge model pertaining to everyday activities [1], such as those involved in table setting and cooking, have consequently been aligned to the DUL framework. Additional axiomatization that is beyond the scope of description logics is integrated by means of the Distributed Ontology Language [9]. For the task at hand, however, only the taxonomic model is employed to classify events and objects.

### 3.5 Scene Generation

Within the scene generation module, we aim to provide a rich contextual world for the following **specification** methods by preparing a scene that contains sufficient interactable objects to ensure *completeness* (i.e. solvability of each contained subgoal) and to facilitate *variety of choices* (in order to retrieve actual world knowledge through humans’ decisive solutions or physical properties of the simulation). Since the **processing** layer results in rather generic, underspecified semantic descriptions of objects required to fulfill the task, this module tries to generate as many alternatives for the respective objects as possible. This can be realized either in a bottom-up (empty scene where only necessary objects and alternatives are generated) or top-down (fully fledged household scene where only objects missing for completion and/or their alternatives are generated) approach. Once a scene meets the conditions of the task, it can be used for both human computation as well as simulation.

Placement of objects in a scene is done in a generate and validate fashion. The qualitative constraints on object placements are first used to select and/or modify probability distributions for object positions. These probability distributions can be learned from a set of training scenes— e.g., what it means for a chair to be “near” a table can be represented as a distribution on relative locations of the chair to the table—, or sometimes inferred from an object’s shape; for example, the top of an object corresponds to the fragments of its surface with the highest z-coordinate. Probability distributions resulting from different constraints on the same object are combined via point-wise multiplication. Once constructed, a probability distribution accounting for all qualitative constraints on an object is sampled several times to produce candidate poses, and the first candidate that passes a list of tests— e.g. placing the object there would not result in collisions— is used.

### 3.6 Human Computation

As the primary gap filler for underspecified information, we introduce *Kitchen Clash*, a virtual reality-based, competitive household serious game. Players are challenged with the same set of instructions that stem from the original knowledge base within a virtual household produced by the scene generation module. Each instruction is realized as reaching a subgoal represented by the contained objects and the type of the manipulation (picking up/dropping objects, combining objects with other objects, making use of specific object properties, etc).

VR, compared to offline or non-natural interaction approaches, offers the great potential of tracking complete trajectories of hand, head and body movement, as well as the distinctly classified manipulation actions. Players are asked to execute these tasks with optimal efficiency and quality, which is measured by *time spent on a task*, the *number of recognizable actions* and the *number of undesired events* (e.g. breaking dishes or glitching through physical barriers). Additionally, these sessions are assessed qualitatively by peer-rating individual executions from other players, in an either absolute or relative measurement. Eventually, players are rewarded with a score representing their qualitative and quantitative success.

### 3.7 Simulation

Within MEANinGS, the simulation branch is employed to estimate concrete parameter setting for the ultimate robotic execution of the activities involved. For example, in the case of transporting liquids in various containers from a source to a target location, the game engine physics can be used to simulate different velocities and trajectories and measure the ensuing spill rate in order to find a suitable setting. Ultimately, we see this as a modern extension of the KARMA system [10], in which the complete understanding of an utterance entails a mental simulation thereof. It is also related to “projection” [8], which is light-weight simulation used by a cognitive robot to try combinations of program parameters and/or change sequences of actions quickly, in a simulated world, before attempting them in reality.

## 4 Exemplary case

To showcase the functional principle of the framework, we present one of the example tasks used in the **Evaluation**, i.e. *to prepare a portion of cucumber salad*.

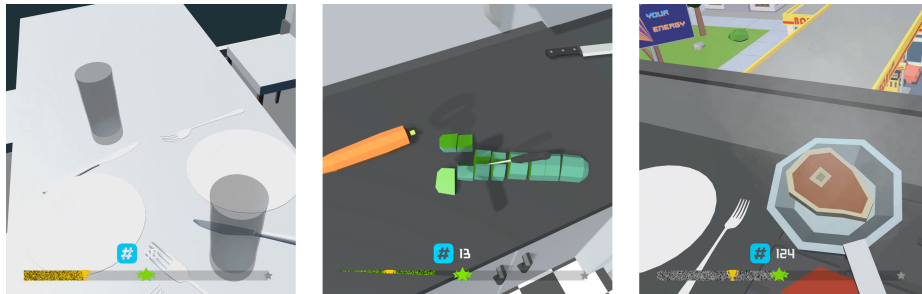
**Retrieval.** When querying wikiHow as a possible source for natural language instructions, *cucumber salad* will result in a multitude of cucumber salad variants, from which the most basic one will be chosen since no further specifications are asked for. Within this module, the overall task will be divided into subtasks (*Slice the cucumber into thin pieces*, *Place the slices into a bowl* and *Pour dressing over the cucumbers*), which will be forwarded to the **processing** layer.

**Processing.** The natural language parser extracts one action per subtask, each of which should be performed by the discourse addressee - in this case the human player. For the *slicing* action, the undergoer *cucumber* is identified while the obligatory instrument slot is left unspecified. Moreover, the action should result in a goal state that is defined by the changed consistency of the undergoing object. Also, the ontologically equivalent *cutting* action is extracted, to prepare

for the case in which only one of these actions is known by the following processing steps. The subsequent *placing* action describes the desired trajectory of the undergoing *slices* to their destination, an undetermined container of type *bowl*. For the final *pouring* action, the poured substance *dressing* and its destination, *the cucumbers*, are identified. Furthermore, the various referring expressions of the main ingredient all resolve to the initial *cucumber* object, in its different configurations.

**Specification.** In order to prepare a suitable testbed, the scene generation module spawns a *cucumber* (since it doesn't find more specific alternatives to the term) and different variants of *cutting objects* (scissors, a kitchen knife, a butter knife, a butcher's knife, etc).

Within *Kitchen Clash*, a new level is generated that constitutes the challenge and constraints of the overall task. Players entering this level have to find suitable solutions for the presented subtasks and execute these quickly and dexterously, since time, number of actions and the opinion of other players determine the final score. If e.g. a player executes a pickup action on the kitchen knife, triggers a collision between the knife and a cucumber (c.f. Figure 2), collects the resulting slices, causes them to fall into a bowl and initiates the final collision between dressing and cucumbers, all subgoal constraints have been fulfilled and the main task is completed.



**Fig. 2.** In-game representation of the three tasks. UI has been kept minimal to prevent distraction, action number is counted and required time outlined on a bar with respect to the best and average time targets. In the second screenshot segment, the *cucumber slicing* task is represented, where the required *cutting object* is specified by taking a *serrated utility knife*.

When it comes to simulating the physical properties, the same scene is populated by a robotic agent instead of a human performer, that evaluates the cutting action between all given alternatives and comes up with a quantitative result of the most appropriate parameters and choices.



**Providing.** In the end, trajectories and action choices from the **specification** layer are formulated into the standardized *NEEM* description to generalize and publish the insights to the open robotic community.

## 5 Evaluation

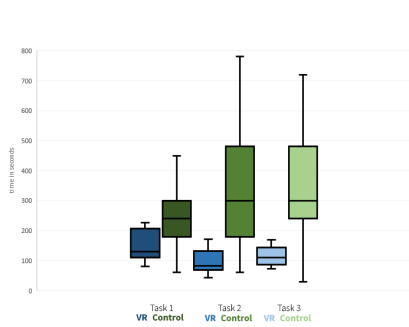
In order to assess the feasibility of the approach, the overall player experience and appraisal, as well as to generate a first data set for further analysis, we conducted an exploratory comparative user study in a laboratory setting. Data was gathered through game protocols, screen capture and a post-study questionnaire. The study was split into two groups in a between-subjects design, where the **VR** group was exposed to *Kitchen Clash* within the associated framework and the **control** group had to accumulate the desired world knowledge manually by depicting the respective tasks in written form.

**Measures.** In-game, we tracked movements from head and hands every second, as well as all of the players' actions, collision events, time measures and attained scores (quantitative and qualitative). The control group submitted instructional data textually. Through the questionnaire, demographics and prior experience in VR were recorded. Using seven-point Likert scales, we asked for players' motivation (using the Intrinsic Motivation Inventory (IMI) [7]), presence (using the igroup Presence Questionnaire (IPQ) [15]), comprehensibility and perceived usefulness of the game. Additionally, participants elaborated on their decision making processes with respect to world knowledge accumulation.

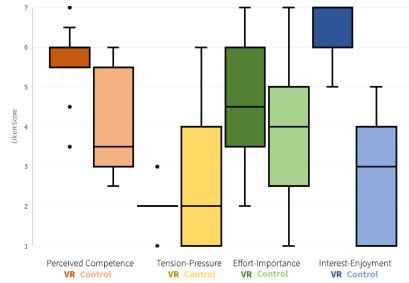
**Procedure.** Following informed consent and a temporally unlimited tutorial that explained the controls and interactions of the game, participants were asked to complete three levels containing complex tasks. In the first level, they had to set a table for two persons, deciding on the type of cutlery and tableware and arranging these in their usual composition. Level two consisted of the formerly explained task of turning cucumbers into a salad. Finally, they were asked to prepare a steak by heating the hotplate, choosing a pan, filling it with oil and cooking the steak until the desired degree of doneness was reached. The tasks did not differ between the VR and control group. They were specifically designed to extract world knowledge about solving underspecified information, providing preferred or conventional items, object target constellations and actual execution trajectories. After completing all levels, the subject was redirected to the final questionnaire.

**Participants.** ( $n = 26$ ) participants took part in the study. (46% male, 54% female, aged 22-58 ( $M = 29.9$ ,  $SD = 8.3$ )). 72.7% stated having prior experience in VR.

**Results.** On average, subjects of the VR group spent ( $M_1 = 150.9, SD_1 = 51.7; M_2 = 94.9, SD_2 = 42.3; M_3 = 114.9, SD_3 = 34.2$ ) seconds on the three respective tasks, whereas the control group required ( $M_1 = 244, SD_1 = 108.1; M_2 = 350, SD_2 = 196.3; M_3 = 336.3, SD_3 = 181.4$ ) seconds. Using a Welch’s t test, we found significant or even highly significant effects for required time between the groups in all tasks ( $p_1 < 0.05, d_1 = 1.1$ ), ( $p_2 < 0.01, d_2 = 1.8$ ), ( $p_3 < 0.01, d_3 = 1.7$ , cf. Figure 3).



**Fig. 3.** Time required to fulfill the three tasks between VR (blue) and control (green).



**Fig. 4.** Results of IMI categories Perceived Competence (red), Tension/Pressure (yellow), Effort-Importance (green) and Interest/Enjoyment (blue) between VR (left) and control (right).

Assessing the IMI, we found no difference for Effort-Importance or Tension-Pressure, but highly significant effects for Perceived Competence ( $p < 0.01, d = 1.26$ ) and Interest-Enjoyment ( $p < 0.01, d = 3.13$ ), showing VR drastically outperforming the control group in terms of motivation (cf. Figure 4). When asked how descriptive the execution in VR (or in written instructions) can be with respect to the real set of actions, 81.2% of the VR group stated that the execution comes close to the real actions, where from the control group only 40% were convinced that real tasks can be sufficiently expressed in written form. Participants had no trouble following the given instructions (indicated by ( $M = 6.27, SD = 0.62$ ) on a comprehensibility scale). According to the IPQ, VR participants reported a mediocre presence ( $M = 4.15, SD = 0.81$ ) for Spatial Presence, ( $M = 4.3, SD = 0.42$ ) for Involvement, ( $M = 3.3, SD = 0.54$ ) for Realness and ( $M = 5.63, SD = 1.15$ ) for General Presence). Regarding simulation sickness, most participants reported no discomfort at all ( $M = 2.1, SD = 1.73$ ). Most of the subjects stated that they would like to play similar games more often ( $M = 5.72, SD = 1.6$ ). Elaborating on the decision making strategy, 45.4% of the participants stated to select the necessary objects based on the respective task or prior experiences, where 54.6% tended to just take the first available thing.

For the qualitative measurements, subjects reported that VR is “*capable of capturing the most crucial aspects of the tasks*” and “*close to reality*”, despite “*lacking haptic feedback [that] decreases grasping accuracy*” and “*not [being] able to perform fine motor functions*”. Participants of the control group stated that it is “*impossible to find the right level of detail*”, “*implicit knowledge is easily overlooked*”, “*it takes way too long to describe all actions in detail*” and “*you cannot really describe cooking since you don’t think at details that will come up in the process*”.

We also assessed the amount of information retrievable from the sessions in both groups. Within VR, all executions managed to complete the tasks and filled all occurrences of underspecified information, since these were needed to finish the respective level. Yet, many unnecessary actions were tracked that trace back to the novel experience of the game, accustoming to VR and the controls and the very broad tracking scope. The amount of unnecessary information was significantly smaller in the control group, but in most of the cases they failed to solve the underspecification problem, even when going into detail. Above that, the textual descriptions deviated considerably in their semantics, due to different perceptions of the task, the projection to their individual environment or personal preferences.

## 6 Discussion and Future Work

Contrasting accumulation of world knowledge manually and in a gamified approach, we have given evidence that human computation can result in significantly higher efficiency, motivation and closeness to the actual execution. Above that, *Kitchen Clash* was able to track complete sequences of actions that describe the fulfillment of tasks both symbolically (registering required operations) as well as subsymbolically (tracking continuous trajectories and contact parameters). Participants enjoyed playing and competing with other players and were interested in continuing the game. Based on these results, we have demonstrated the opportunities and usefulness of human computation for world knowledge aggregation and the feasibility of the overall framework. Yet, this study illustrated that the current implementation suffers from over-collecting unnecessary information and undesirable player choices (e.g. players who take the first object available instead of making an informed decision). Regarding the first issue, we aim to compile large sets of similar task executions using Deep Player Behavior Models [14], offering an optimization paradigm across sessions to extract the necessary core actions needed to fulfill the task probabilistically. When it comes to undesirable player choices, we will evaluate a knockout system of object alternatives that constrains the *variety of choices* of the Scene Generation module in order to force players to overcome obstinate individual preferences and obvious decisions. Furthermore, we are aiming for a narrower interaction between the human computation and the simulation module to generate more elaborate

level constellations in *Kitchen Clash* and to make use of the accumulated sequential action knowledge while simulating. Eventually, we are going to open up the game to online multiplayer scenarios where players have to compete against other human players as well as agents representing the aggregated knowledge while learning continually.

## 7 Conclusion

Learning from natural language instructions is a desirable opportunity for robots, but ends up in underspecified information, even when accessing detailed directions. Introducing MEANinGS, we present a potent framework able to break down these instructions syntactically and semantically, before resolving missing or underspecified information with the aid of human computation. With this approach, we have shown to outperform manual accumulation in terms of efficiency, motivation and completeness. This work demonstrates a successful application of a human computation serious game to facilitate research in the context of robotic learning.

## 8 Acknowledgments

This work was funded by the German Research Foundation (DFG) as part of Collaborative Research Center (SFB) 1320 EASE - Everyday Activity Science and Engineering, University of Bremen (<http://www.easecrc.org/>), subprojects H2 and P1. We thank all participants.

## References

1. Beetz, M., Bessler, D., Haidu, A., Pomarlan, M., Bozcuoglu, A.K., Bartels, G.: Know Rob 2.0 - A 2nd Generation Knowledge Processing Framework for Cognition-Enabled Robotic Agents. Proceedings - IEEE International Conference on Robotics and Automation pp. 512–519 (2018). <https://doi.org/10.1109/ICRA.2018.8460964>
2. Beetz, M., Mösenlechner, L., Tenorth, M.: Cram—a cognitive robot abstract machine for everyday manipulation in human environments. In: 2010 IEEE/RSJ International Conference on Intelligent Robots and Systems. pp. 1012–1017. IEEE (2010)
3. Dang, H., Allen, P.K.: Robot learning of everyday object manipulations via human demonstration. In: 2010 IEEE/RSJ International Conference on Intelligent Robots and Systems. pp. 1284–1289. IEEE (2010)
4. Kunze, L., Tenorth, M., Beetz, M.: Putting people’s common sense into knowledge bases of household robots. In: Annual Conference on Artificial Intelligence. pp. 151–159. Springer (2010)
5. Mascardi, V., Cordì, V., Rosso, P.: A comparison of upper ontologies (technical report disi-tr-06-21). Dipartimento di Informatica e Scienze dell’Informazione (DISI), Università degli Studi di Genova, Via Dodecaneso **35**, 16146 (2008)
6. Masolo, C., Borgo, S., Gangemi, A., Guarino, N., Oltramari, A.: Wonderweb deliverable d18, ontology library (final). ICT project **33052**, 31 (2003)

7. McAuley, E., Duncan, T., Tammen, V.V.: Psychometric properties of the Intrinsic Motivation Inventory in a competitive sport setting: a confirmatory factor analysis. *Research quarterly for exercise and sport* **60**(1), 48–58 (1989)
8. Mösenlechner, L., Beetz, M.: Fast temporal projection using accurate physics-based geometric reasoning. In: *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*. pp. 1821–1827. Karlsruhe, Germany (May 6–10 2013)
9. Mossakowski, T.: The distributed ontology, model and specification language–dol. In: *International Workshop on Algebraic Development Techniques*. pp. 5–10. Springer (2016)
10. Narayanan, S.S.: Karma: Knowledge-based active representations for metaphor and aspect. (1999)
11. Nielsen, R.D., Voyles, R., Bolanos, D., Mahoor, M.H., Pace, W.D., Siek, K.A., Ward, W.H.: A platform for human-robot dialog systems research. In: *2010 AAAI Fall Symposium Series* (2010)
12. Nilsson, N.J.: Shakey the robot. Tech. rep., SRI INTERNATIONAL MENLO PARK CA (1984)
13. Parde, N.P., Papakostas, M., Tsiakas, K., Dagioglou, M., Karkaletsis, V., Nielsen, R.D.: I spy: An interactive game-based approach to multimodal robot learning. In: *Workshops at the Twenty-Ninth AAAI Conference on Artificial Intelligence* (2015)
14. Pfau, J., Smeddinck, J.D., Malaka, R.: Towards deep player behavior models in mmorpgs. In: *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play*. pp. 381–392. CHI PLAY '18, ACM, New York, NY, USA (2018). <https://doi.org/10.1145/3242671.3242706>
15. Schubert, T., Friedmann, F., Regenbrecht, H.: Embodied presence in virtual environments (1999)
16. Steels, L.: Basics of fluid construction grammar. *Constructions and Frames* **9**(2), 178–225 (2017). <https://doi.org/https://doi.org/10.1075/cf.00002.ste>
17. Stulp, F., Fedrizzi, A., Mösenlechner, L., Beetz, M.: Learning and Reasoning with Action-Related Places for Robust Mobile Manipulation. *Journal of Artificial Intelligence Research (JAIR)* **43**, 1–42 (2012)
18. Walther-Franks, B., Smeddinck, J., Szmids, P., Haidu, A., Beetz, M., Malaka, R.: Robots, pancakes, and computer games: designing serious games for robot imitation learning. In: *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. pp. 3623–3632. ACM (2015)
19. Winkler, J., Bozcuoğlu, A.K., Pomarlan, M., Beetz, M.: Task parametrization through multi-modal analysis of robot experiences. In: *Proceedings of the 16th Conference on Autonomous Agents and MultiAgent Systems*. pp. 1754–1756. AAMAS '17, International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC (2017), <http://dl.acm.org/citation.cfm?id=3091125.3091428>