Rock, Paper, Scissor: What's a Substitute For Hammer?
An Approach to Substitute Selection for Missing Tool Using Robot-Centric Conceptual Knowledge About Objects

Dissertation
zur Erlangung des akademischen Grades
Doktoringenieur (Dr.-Ing.)

Angenommen durch die Fakultät für Informatik
der Otto-von-Guericke-Universität Magdeburg

von M.Sc. **Madhura Thosar**
geb. am 18.12.1980 in Pune, Indien

Gutachterinnen/Gutachter:
Prof. Dr. Till Mossakowski
Prof. Dr. Sebastian Zug
Prof. Dr. Michael Beetz

Datum der Einreichung: 14 January, 2022
Verteidigungsdatum: 9 June 2022

Thosar, Madhura Dilip:
Rock, Paper, Scissor: What's a Substitute For Hammer?
An Approach to Substitute Selection for Missing Tool Using Robot-Centric Conceptual Knowledge About Objects
Doctoral Thesis, Otto-von-Guericke University
Magdeburg, 2022.

Contents

Abstract

When a robot is operating in a dynamic environment, it cannot be assumed that a tool required to solve a given task will always be available. Consider, for instance, a scenario where a robot is asked to serve drinks where the robot has to use a tray to serve the drinks. If the robot finds the tray, the task will progress, but what if the tray is unavailable. Such situations are common occurrences in our daily lives. In situations like these a robot would be expected to improvise like humans from the other available objects, for example, by using an eating plate for serving. This skill is significant when operating in a dynamic, uncertain environment because it allows a robot to adapt to unforeseen situations. The question is: how can a robot determine which object in the environment is a suitable substitute for a missing tool?

For substitute selection, we took inspiration from the way humans select a substitute from the existing objects for a missing tool wherein humans take into consideration conceptual knowledge about object's physical and functional properties. For instance, consider a scenario in which one has to choose between a plate and a mouse pad as an alternative for a tray. For a tray whose designated purpose is *to carry*, rigid and flat are more relevant to *carry* than a material or a color of a tray. In order to find a suitable substitute, the relevant properties of the missing tool need to correspond, as large a degree as possible, to the properties of the available objects. In this work, we have proposed an approach to substitute selection where a conceptual knowledge-driven computation is performed to identify the relevant properties of the missing tool and determine a substitute on the basis of shared relevant properties. The question is, how to acquire such knowledge about the properties.

An argument has been put forth in cognitive science for bottom-up generation of knowledge in which humans and animals alike develop conceptual understanding of objects based on their own perceptual experiences with objects. We have followed suit and propose that knowledge about properties should be generated from the sensory measurements of the properties. We have termed such bottom-up generation of knowledge from the sensory measurements as *robot-centric* knowledge. We propose an extensible property estimation framework which consists of estimations methods to obtain the sensory measurements of physical properties (rigidity, weight, etc.) and functional properties (containment, support, etc.) from household objects. In our second contribution we employ 1) unsupervised clustering methods to transform the sensory measurements of the properties into symbols, and then 2) bivariate joint frequency distributions and sample proportion to generate conceptual knowledge about objects.

In this work, we have presented a proof of concept of the proposed approaches. We acquired a dataset comprising six *physical* and four *functional* properties of 110 household objects using the proposed property estimation methods. This

dataset was used to evaluate the property estimation methods and the semantics of the considered properties within the dataset. Furthermore, the dataset is used to generate the proposed robot-centric conceptual knowledge which is then used by our proposed substitute selection system to identify a substitute from the available objects in different missing tool scenarios.

Zusammenfassung

Wenn ein Roboter in einer dynamischen Umgebung agiert, kann nicht davon aus-gegangen werden, dass ein Werkzeug, das zur Lösung einer bestimmten Aufgabebenötigt wird, immer verfügbar ist. Nehmen wir zum Beispiel ein Szenario, indem ein Roboter aufgefordert wird, Getränke zu servieren, wobei der Roboter ein Tablett benutzen muss, um die Getränke zu servieren. Wenn der Roboter das Tablett findet, wird die Aufgabe erfüllt, aber was ist, wenn das Tablett nicht verfügbar ist? Solche Situationen kommen in unserem täglichen Leben häufig vor. Insolchen Situationen wird von einem Roboter erwartet, dass er wie ein Mensch ausanderen verfügbaren Gegenständen improvisiert, indem er zum Beispiel einen Essteller zum Servieren verwendet. Diese Fähigkeit zur Improvisation ist von großer Bedeutung, wenn er in einer dynamischen, unbekannten Umgebung eingesetzt wird, denn sie ermöglicht es einem Roboter, sich an unvorhergesehene Situationen anzupassen. Die Frage ist: Wie kann ein Roboter feststellen, welches Objekt in der Umgebung ein geeigneter Ersatz für ein fehlendes Werkzeug darstellt

Für die Auswahl eines Ersatzes haben wir uns von der Art und Weise inspirieren lassen, wie Menschen einen solche Aufgabe aus den bereits vorhandenen Objekten lösen. Dazu berücksichtigen Menschen insbesondere das konzeptionelle Wissen über die physischen und funktionalen Eigenschaften des Objekts. Betrachten wir zum Beispiel ein Szenario, in dem man zwischen einem Teller und einem Mauspad als Alternative für ein Tablett wählen muss. Bei einem Tablett, dessen Hauptzweck das Tragen ist, sind die Eigenschaften wie starr und flach wichtiger als das Material oder die Farbe eines Tabletts. Um einen geeigneten Ersatz zu finden, müssen die relevanten Eigenschaften des fehlenden Werkzeugs so weit wie möglich mit den Eigenschaften der verfügbaren Objekte übereinstimmen.In dieser Forschungsarbeit haben wir einen Ansatz für die Auswahl eines Ersatz es vorgeschlagen, bei dem eine konzeptionelle, wissensbasierte Berechnung durchgeführt wird, um die relevanten Eigenschaften des fehlenden Werkzeugs zu ermitteln und einen Ersatz auf der Grundlage der gemeinsamen relevanten Eigenschaften zu bestimmen. Die Frage ist, wie man sich dieses Wissen über die Eigenschaften aneignet.

In der Kognitionswissenschaft wurde ein Argument für die Bottom-up-Erzeugung von Wissen vorgebracht, wonach Menschen und Tiere gleichermaßen ein begriffliches Verständnis von Objekten auf der Grundlage ihrer eigenen Wahrnehmungserfahrungen mit Objekten entwickeln. Wir sind diesem Beispiel gefolgt und schlagen vor, dass das Wissen über Eigenschaften aus den sensorischen Messungen der Eigenschaften gewonnen werden sollte. Wir haben eine solche Bottom-up-Generierung von Wissen aus den sensorischen Messungen als roboterzentriertes Wissen bezeichnet. Wir schlagen einen erweiterbaren Rahmen für die Schätzung von Eigenschaften vor, der aus Schätzungs methoden besteht, um die sensorischen Messungen der physikalischen Eigenschaften

(Steifigkeit,Gewicht usw.) und der funktionalen Eigenschaften (Behältnis, Halt usw.) von Haushaltsgegenständen zu erhalten. In unserem zweiten Beitrag schlagen wir 1) unüberwachte Clustering-Methoden vor, um die sensorischen Messungen der Eigenschaften in Symbole umzuwandeln, und dann 2) bivariate gemeinsame Häufigkeitsverteilungen und Stichprobenanteile, um konzeptionelles Wissenüber Objekte zu generieren.

In dieser Forschungsarbeit haben wir ein Proof of Concept der vorgeschlagenen Ansätze vorgestellt. Wir haben einen Datensatz mit sechs physikalischen und vier funktionalen Eigenschaften von 110 Haushaltsgegenständen erstellt und dabei die vorgeschlagenen Methoden zur Eigenschaftsschätzung verwendet. Dieser Datensatz wurde verwendet, um die Methoden zur Schätzung der Eigenschaftenund die Semantik der betrachteten Eigenschaften innerhalb des Datensatzes zubewerten. Darüber hinaus wird der Datensatz verwendet, um das vorgeschlagene roboterzentrierte konzeptionelle Wissen zu generieren, das dann von unserem vorgeschlagenen Ersatzauswahlsystem verwendet wird, um einen Ersatz aus den verfügbaren Objekten in verschiedenen Szenarien mit fehlendem Werkzeug zu identifizieren.

Star Trek: Deep Space Nine
Season 4, Episode 13: Return to Grace

1

Introduction

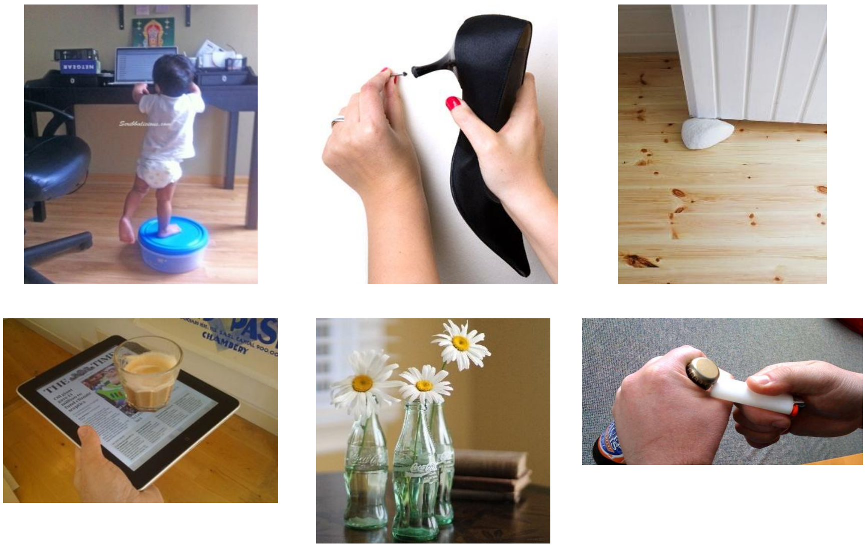


Figure 1.1: Looks familiar? (image credits are cited in the text)

1.1 Motivation

Does the scenarios in the figure 1.1 seem familiar? We, humans, have been in these situations from time-to-time: a plastic box is being *used as* a stool[1]; a heeled shoe is being *used as* a hammer[2]; a rock is being *used as* a door stopper[3], a tablet is being *used as* a tray[4]; coke bottles are being *used as* a vase[5] and a lighter being *used as* a bottle opener[6]. The common thread in these examples is that an object is not being used for its intended purpose, that is a heeled shoe is not intended to be used as a hammer. It just seems a coincidence, that it can also be used as a hammer or does it?

The sophistication pertaining to tool-use in humans involves not just the dexterity in manipulating a tool, but also the diversity in tool exploitation. The ability to exploit the tools has enabled humans to adapt and thus exert control over an uncertain environment, especially when they are faced with unfavorable situations.

Now, consider a scenario where a robot is asked to serve a coffee (see Fig 1.2[7]). Let us assume that a robot knows how to perform the task. It goes to a kitchen, locates a coffee pot and a cup, pours coffee in the cup. The robot then locates a tray and places the cup on the tray and delivers it. However, what if the tray is unavailable! In this case, what should a robot do when it is unable to locate the tray. We do not want the robot to quit or wait until the tray becomes available.

What do we humans do, in situations like this? These are not uncommon scenes for humans (See Fig. 1.1). In situations like this, humans typically respond by improvising the usability of available objects in an environment.

[1] `Scribbalicious.com`, (*the website does not exist anymore*)
[2] `https://www.omniagroup.com/dont-hire-a-shoe/`
[3] https://www.gardenista.com/posts/diy-idea-a-no-cost-painted-stone-door-stopper-koizumi-studio-tetu-iron-door-stopper/
[4] `https://www.businessinsider.com/hot-ipads-can-heat-up-coffee-2012-3`
[5] `http://www.lesliereese.com/leslie-reese/tag/flower+arranging`
[6] `https://www.pinterest.com/pin/132645151500976736/`
[7] Wall-e thinking: `https://www.pinterest.com/pin/182536591132680554/`
Coffee cups: `https://www.planspin.com/serving-trays.html`
Wall-e quitting: `https://stickers.cloud/fr/pack/wall-e`

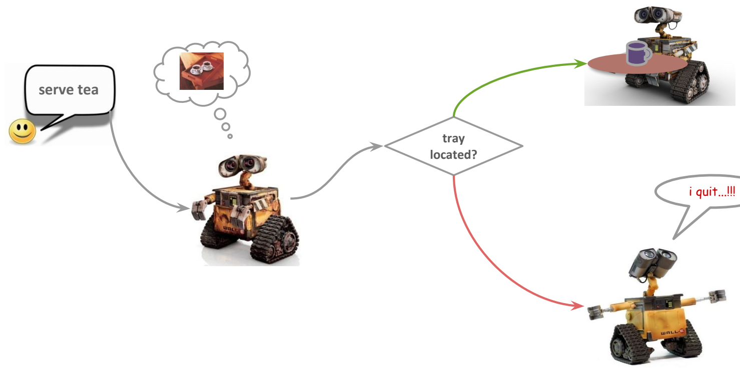


Figure 1.2: Consider a scenario... (image credits are cited in the text)

In other words, we find a replacement or a *substitute* that *can be used as* a missing tool. In situations like these, humans seem to know - either from the past experience or from observations or from the "necessity-is-the-mother-of-improvisation(invention)" type approach - what kind of object is needed as a substitute. This skill is significant when operating in a dynamic, uncertain environment because it will allow a robot to adapt to unforeseen situations to a degree. The question is how can a robot determine which object in the environment is a suitable candidate for a substitute?

One possible approach to determine a suitability would be by maneuvering an object in the same manner as a missing tool. However it would be time consuming if the robot interacts with every single object in the environment to determine the suitability of a substitute which makes this approach less practical. The question is how to identify a suitable candidate for a substitute for a missing tool without interacting with every single object in the environment. My doctoral research investigates this very problem.

1.2 Inspiration - how do humans select a substitute?

A typical tool substitution task involves selecting a suitable substitute for a missing tool and using it in an ongoing task. Like tool use, tool substitution is an elaborate endeavor which involves bio-mechanical and cognitive aspects of problem solving [1]. Consider, for example a heel of a shoe. Though its primary function is to extend the height of a person, it can also be used for hammering a nail (see Fig 1.1). One can observe cognitive reasoning involving analogous thinking about a shoe as a hammer while taking into account the bio-mechanics of manipulating the shoe as a hammer. The inquiry worth investigating, however, is, why was shoe selected.

In order to select a plausible substitute for a missing tool, the substitute needs to be recognised to be similar to the missing tool in some way without having to interact with it. The question is what is needed to determine the similarity. Consider, for instance, a scenario in which a substitute for a small stool to be used for sitting is a decision between two available objects: a plastic container and a cardboard box (figure 1.3). Baber in [1] pointed out that humans use *conceptual knowledge* about tools to reason about a substitute on the basis of suitability. For instance, we consider the stool as a rigid, medium sized object with a flat surface (ref table 1.1). Thus, to find the most appropriate substitute, the properties of the possible choices need to correspond to as large a degree as possible to the properties of the original object. The problems with this direct matching of properties is that some properties matter more than the others, i.e. that some properties are more relevant than others which enables a designated purpose of a tool [2]. Take, for instance the objects, a stool, a plastic container, and a cardboard box. For a stool, whose designated purpose is to be 'sit-able', we know that *rigidity* and *flat surface* are essential, *size* is less relevant and *color* is completely irrelevant. Similarly, for a plastic container and a cardboard box, whose designated purpose is to 'contain' something, *hollowness* is important, *size* is less relevant and *color* is irrelevant. In order to find the most appropriate substitute, the relevant properties of a missing tool need to correspond to as large a degree as possible to the properties of the possible choices for a substitute [3]. As a result, a plastic container seems to be an appropriate choice as a substitute for a stool as it shares most of the relevant properties of a stool than a cardboard box.

Figure 1.3: Possible replacements for a stool: a plastic container and a cardboard box

Property	**Stool**	**Plastic container**	**Cardboard box**
Rigidity	High	High	Low
Flatness of surface	High	High	High
Size	Medium	Small	Small
Color	Orange	Grey	Brown

Table 1.1: Property table for stool: tool substitution.

1.2.1 Objectives of this dissertation

In this research work, we have adopted a similar approach. Accordingly, in order to select a substitute, a substitute selection system needs: conceptual knowledge about objects, relevant properties of a missing tool and a mechanism to determine whether the relevant properties are present in the available objects in the environment. On this basis, we can now lay down the primary objectives that need to be carried out when selecting a substitute

Objective 1: Acquire conceptual knowledge about objects which contains knowledge about properties of objects

Objective 2: Identify relevant properties with respect to the primary purpose of a missing tool

Objective 3: Determine a substitute on the basis of relevant properties of a missing tool

1.3 Research Questions

Now that we have laid down the objectives, we are in a position to formulate research questions that are addressed in this research work. We first note down below the research questions related to the Objective 1 concerning the conceptual knowledge, followed by the research questions related to the Objectives 2 and 3 which focus on substitute selection. The research questions are divided into conceptual and technical research questions. We deemed it necessary to formulate the conceptual research questions as they allow us to define and understand the scope and the complexity of a problem being addressed. Accordingly, the conceptual research questions, in our work, form the foundations for the subsequent technical research questions. As a result, we pose a conceptual question first by asking *What* followed by a technical question formulated using *How*.

1) What is the *nature* of the conceptual knowledge about objects *desired* in substitute selection?

Before we approach the *Objective 1*, it is necessary to orient the conceptual knowledge and lay down its scope with respect to its contents such that it can be used in substitute selection scenarios. Such discussion will form the foundation for designing the approach to acquire the desired knowledge. Therefore, we have posed this conceptual question which deals with the characterization of knowledge with respect to its contents and representation.

It is postulated in the literature on tool-use in animals [1] that "*a non-invasive tool selection in humans or animals alike is facilitated by conceptual knowledge about objects, especially, knowledge about their physical and functional properties and relationship between them*". It is not going to be any different for a robot in substitute selection situations where conceptual knowledge about objects will allow it to make such selections in a non-invasive manner. Therefore, if the conceptual knowledge about objects for a robot is to be consisted of physical and functional properties, the questions that need to be answered are: what are the desired contents of such conceptual knowledge about objects and how are they to be characterized such that the knowledge can be used for a substitute selection purpose? With respect to the physical and functional properties of objects, it is essential to specify what constitute physical and functional properties in this work. After all, we humans are capable of describing an object in term of its, for instance geometrical properties such as shape, size; mechanical properties such as rigidity, weight; thermal properties such as boiling point, melting point; chemical properties such as reactivity, corrosion resistance.

Moreover, while designing the contents, another issue that must be dealt with is the desired granularity of the knowledge to be acquired. In other words, how much detailed knowledge about objects is necessary and sufficient for selecting a substitute. It is worth noting that humans can describe an object at a macroscopic level as well as a microscopic level. The question is, what is the desired granularity of knowledge in substitute selection.

2) How to acquire such conceptual knowledge?

This research question is concerning the Objective 1 and builds upon the conceptual question about the nature of the conceptual knowledge. The acquisition of knowledge, in our view, can not be addressed without addressing the nature of the desired knowledge as the acquisition approach will be influenced by the contents and the granularity of the knowledge.

A straight forwards answer to this question would be to hand-code the desired knowledge, but note that it is not possible to hand-code knowledge about every possible object in the world. A second option would be to use the existing knowledge bases such as WordNet, ConceptNet etc. This could be a viable option, despite of them being hand-coded by humans. Another possible way is let a robot acquire the desired knowledge by interacting with the objects. The underlying question, in that case, is what kind of interactions can be designed to acquire such knowledge. Given that a robot has limited perception capabilities, it is also possible to fill the missing knowledge gap by combining the hand-coded knowledge or existing knowledge bases with a robot-acquired knowledge. Another question which needs to be addressed is, whether the knowledge should be acquired in an online manner or offline or a combination of both. Note that, in case of an online knowledge acquisition, the knowledge is usually acquired in an incremental manner. Some of the common online knowledge acquisition methods are: through interaction with the environment, human-robot interaction etc. In contrast, an offline knowledge acquisition allows the knowledge acquisition in bulk, for instance, acquiring knowledge from the existing knowledge bases like WordNet, ConceptNet, DBpedia or reasoning about the already acquired knowledge.

Acquisition of knowledge leads to two additional aspects about knowledge: knowledge representation and symbol grounding. As the acquired knowledge is to be used for substitute selection purposes, the question is how to represent knowledge that is suitable to reason about a substitute? Additionally, while considering the formalism, the desired characterization of the knowledge needs to be taken into account as well. Moreover, as we gain more experiences, learn about properties, exposed to new objects, new instances of the known objects, our knowledge about objects is reviewed and updated. As a consequence, our knowledge about objects is constantly evolving. It also holds true for a robot and therefore, when devising an acquisition process and representation formalism for the knowledge, we have to take this into account.

In artificial intelligence, in general, the knowledge is represented in symbolic or sub-symbolic formalisms. Symbolic knowledge usually represents mental representation of the outside world and in such case, such knowledge should have a correspondence to the respective aspects of the real world. For instance, a symbol *cup* should correspond to a physical cup in the real world. This correspondence is commonly referred as symbol grounding. For a robot, when representing knowledge about objects, it is essential that the knowledge is grounded in robot's perception or sensory data to be specific. It is not enough for a robot to simply have an access to a symbol *cup* without knowing what it means. In other words, for a robot, a symbol grounding can be seen as providing a meaning to the symbols by means of the sensory perception of the world. In case of this doctoral research, the central question is, what purpose does a symbol grounding serve in substitute selection? Especially, as knowledge about objects consists of physical and functional properties, the question is why and how do we ground physical and functional properties of objects into robot's sensory data? While it is a straight forward process to ground symbols representing objects such as *cup* into sensory data of a real-world cup, the real challenge is, how can we ground an abstract physical property, which can not be seen, into robot's sensory data. The more intriguing question, however, is what kind of sensory data do we need to acquire to represent a physical property of an object and how to acquire such data?

3) What is a substitute?

In order to attain the Objectives 2 and 3, we first need to conceptualize what a substitute is. Note that a substitute is a part of a tool substitution and therefore, before we conceptualize a substitute, it is essential to address the questions: what is a tool substitution and where does a substitute fit in the tool substitution? Moreover, we have noted in the literature on tool substitution, discussed in Chapter 4, Sec. 4.4 that the terms tool substitution and substitute selection are used interchangeably, however the question is, is tool substitution same as substitute selection? If not, then the question is where does a substitute selection fit in the tool substitution?

In regard to conceptualizing a substitute, it is essential that we conceptualize a tool first. There are many definitions of a *tool* in the literature. At the core, a tool is a physical object in the environment as suggested in the literature on tool-use [3; 4], on the other hand, Butler in [5] suggested that *"Nothing is tool unless during actual use"*, thus providing a much broader perspective to what can be constituted as a tool. In our case, it is not only a tool that requires defining but also a substitute. It can be agreed that at the core, a tool and a substitute are both objects. The question is when does an object become a tool and when does an object become a substitute? Note that, it is likely that a choice of a substitute for a missing tool may not be universal and will be influenced by personal preferences. Such subjective selection of a substitute means there is no such thing as an accurate substitute for a missing tool. For instance, between a book and a tablet as substitute options for a tray, some users may select a book while some users may select a tray. Secondly, a substitute does not necessarily has to be a look-alike of a missing tool. For instance, a stone and a hammer do not look alike, however, a stone can be used as a hammer. In other terms, when we (humans) are looking for a substitute, we do not necessarily look for most accurate or maximum similar substitute but sufficiently similar object that can be used as a substitute. This begs the question: how is a tool differentiated from a substitute or what are the characteristics of a tool and of a substitute.

4) How to determine a substitute for a missing tool?

This is a technical question which is built upon the conceptual question about a substitute and it is concerning the Objectives 2 and 3. As noted, the definition of a substitute will form the basis for a computational model of a substitute selection. As our substitute selection approach is inspired by the way humans select a substitute, the research question can be divided into two parts. Firstly, as humans select a substitute on the basis of relevant properties of a missing tool, the question that needs to be addressed is how to identify the relevant properties of the missing tool. We should bear in mind that, the relevant properties of a tool are embedded in the conceptual knowledge about the tool, however, they are not marked as such during the knowledge acquisition process. After we identify the relevant properties, the subsequent question can be posed as how to determine the suitability of a substitute on the basis of the relevant properties of the missing tool. Note that, a tool substitution usually takes place during an ongoing task. It should also be noted that for any user or a robot operating in the environment, there are countless number of missing tool situations with countless number of tools which can occur in countless number of places at any given time. This means, it is nearly impossible to know beforehand what objects will be available for every missing tool. As a result, a robot has to select a substitute from the available objects and determine how to use it as a missing tool during run-time in order to finish the task in a timely manner. Therefore, when devising an approach to substitute selection, a run-time response needs to be taken into account. Moreover, while humans select a substitute on the basis of its approximate similarity with a missing tool, we establish its substitutability only after it is used. In other terms, the validity of a substitute's accuracy for a missing tool can only be determined after using it. If the task was finished successfully using a substitute, we ascertain that it is an accurate substitute for a missing tool. However, if the task could not be finished using the substitute, then we proceed with failure analysis. The failure analysis may lead to various conclusions such as: 1) a substitute was used incorrectly; 2) a substitute can not be used due to the user's limited capabilities; 3) unforeseen environmental changes which may have affected the task etc. The challenge is can a substitute be validated by a robot in a similar manner and if not, what are the other manners in which a substitute can be validated?

1.4 Contribution

1.4.1 Overview of the system

Before we lay down our contribution we have made in this research work, we would like provide a context which will allow us to place our contribution. In the following, we provide a workflow of a tool substitution system which consists of typical steps that are to be followed in order to perform tool substitution. We have also provided a brief description of each step.

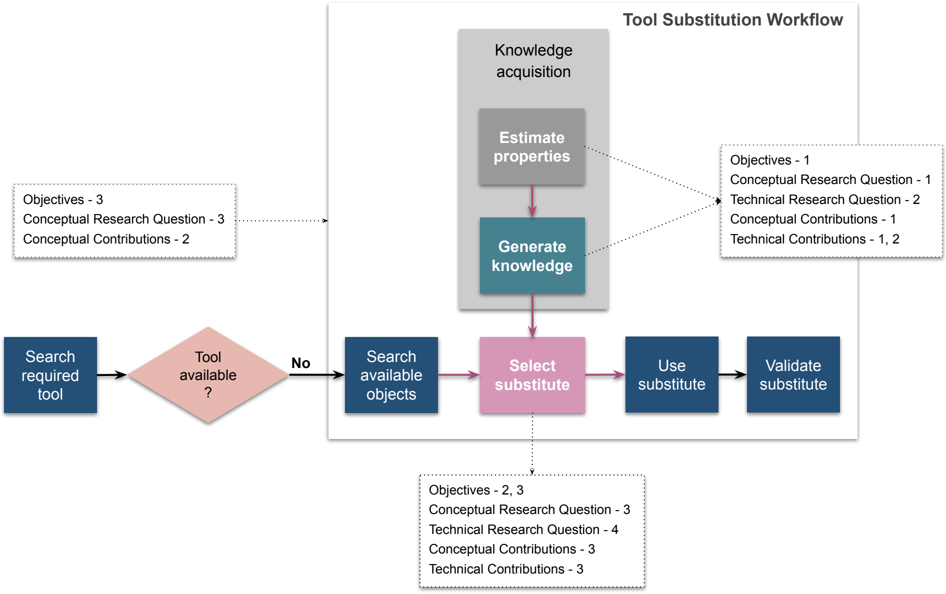


Figure 1.4: An illustration of a tool substitution workflow we have proposed in this research work. In the figure we have also included the modules that were the result of our research work. Additionally we have indicated in the workflow the respective objectives, conceptual and technical research questions, conceptual and technical contributions.

Fig. 1.4 illustrates graphically an aerial view of a workflow we have proposed in this work which is based on the literature on tool use in humans and animals as reported in [3; 4]. It highlights the processes involved in a typical tool substitution task. As the objective of the figure is to offer an overview of the workflow, the processes in the figure are stated without specifying the respective inputs and outputs for each process. We have elaborated the workflow further in the subsequent chapters.

The tool substitution is typically caused by the unavailability of a required tool in an ongoing task. A typical workflow of the tool substitution for a robot would be executed as follows: When a required tool is unavailable, *Search available objects* searches for available objects in the environment and sends the list of the available objects to *Select substitute* process. The *Knowledge acquisition* process sends the knowledge about the available objects to the *Select substitute* process. From the available objects, the *Select substitutes* determines a substitute on the basis of conceptual knowledge about objects and forwards the selected substitute to the *Use substitute* process. A robot uses the substitute as a missing tool in the ongoing task. The *Use substitute* sends the substitute performance feedback to the *Validate substitute* process. The use of the substitute is validated by the *Validate substitute* after evaluating the task performance using the substitute.

In this doctoral research work, we focus on the substitute selection process in the tool chain. Besides the substitute selection process, the research work also focuses on the acquisition of knowledge required for the selection process. The knowledge acquisition process consists of two modules: property estimation and knowledge generation. In property estimation, the physical and functional properties are estimated from household objects. The estimated properties are then used to generate the desired knowledge about objects. These three processes are our primary approaches we have proposed in our work. In the following, we lay down the contributions of this doctoral research.

1.4.2 Contributions

Our contributions lie on two different levels: some of our work focuses on the theoretical aspects wherein we propose a conceptual framework for tool substitution and for substitute selection. The conceptual framework forms the foundation for contribution at the technical level. In this thesis, we have implemented a proof-of-concept based on the proposed approaches and the conceptual frameworks. Each approach and its proof-of-concept are evaluated by performing various experiments which are discussed in the corresponding chapters. In the following, we briefly discuss our contributions which address the research questions posed in the previ-

Conceptual Contributions	Objectives	Conceptual Research Questions	Chapter
1. Conceptual Knowledge	**1.** Acquire conceptual knowledge about objects which contains knowledge about properties of objects	**1.** What is the nature of the conceptual knowledge about objects desired in substitute selection?	**2.** Property Estimation **3.** Knowledge Generation
2. Tool Substitution	**3.** Determine a substitute on the basis of relevant properties of a missing tool	**3.** What is a substitute?	**4.** Substitute Selection **5.** Discussion
3. Tool vs Substitute	**2.** Identify relevant properties with respect to the primary purpose of a missing tool **3.** Determine a substitute on the basis of relevant properties of a missing tool	**3.** What is a substitute?	**4.** Substitute Selection
Technical Contributions		**Technical Research Questions**	
1. Generation of Conceptual Knowledge	**1.** Acquire conceptual knowledge about objects which contains knowledge about properties of objects	**2.** How to acquire such conceptual knowledge	**3.** Knowledge Generation
2. Property Estimation	**1.** Acquire conceptual knowledge about objects which contains knowledge about properties of objects	**2.** How to acquire such conceptual knowledge	**2.** Property Estimation
3. Substitute Selection	**2.** Identify relevant properties with respect to the primary purpose of a missing tool **3.** Determine a substitute on the basis of relevant properties of a missing tool	**4.** How to determine a substitute for a missing tool?	**4.** Substitute Selection **5.** Discussion

Figure 1.5: A guide to how each contribution is related to the objectives and research questions. We have also noted the chapter/s that discuss/es each contribution.

ous section and which are elaborated in detail in the subsequent chapters (see fig. 1.5).

(I) Conceptual Contributions

1. **Conceptual Knowledge:** We have proposed the desired nature of the conceptual knowledge about objects that is suitable for substitute selection wherein we suggest that the conceptual knowledge should consist of the physical and functional properties of the objects. The nature of such knowledge primarily focuses on the contents of the knowledge, the characterization of the contents and its representation. Our proposal is inspired from the insights we gained from the literature on tool use in animals and humans.

2. **Tool Substitution Workflow:** We have proposed a tool substitution workflow which consist of typical processes that are required for tool substitution. In the workflow, we have proposed what inputs are required to each process wherein we have also proposed the characterization of each input. The primary objective of the tool substitution

workflow is to illustrate the overall complexity of the system which include how many different functionalities of the robots are required and how the integrated system would look like. Additionally, it also places the proposed substitute selection system within the workflow.

3. **Tool vs Substitute:** We have suggested a definition for a substitute on the basis of the definition of a tool suggested in the literature on tool use in animals and humans. The definition of a substitute also specifies the scope of a substitute as we have noted that a substitute has many forms. In addition to the definition, we have also proposed the characterization of a substitute which differentiates it from a tool.

(II) Technical Contributions

1. **Generation of Conceptual Knowledge:** Our first contribution focuses on the generation of conceptual knowledge about objects. We propose that the knowledge should be generated from quantitative measurements of object properties in order to capture the proposed characterization of the conceptual knowledge stated in the conceptual contribution. The knowledge generation module in conjunction with property estimation make up the knowledge acquisition process which is decoupled from the substitute selection process in order to manage evolving knowledge. Since the knowledge is generated from quantitative measurements of the properties of objects, the knowledge about objects gets grounded into the property measurements estimated using robot's sensory capabilities. In this manner, we bypass a separate symbol grounding process altogether.

2. **Property Estimation:** In this work, we propose light-weight estimation methods for rigidity, hollowness, size, flatness and roughness, requiring a minimal experimental set-up to generate quantitative measurements of a respective property of an object. Our proposed methods estimate the properties from a single instance at a time and do not require any prior training data for estimation. Additionally, we have proposed an *extensible* property estimation framework called **Ro**bot-**C**entric Data**s**et Framework (RoCS) wherein multiple property estimation methods reside. Moreover, given that multiple property estimation methods can be developed for a same property,

the framework is designed such that new estimation methods can be plugged-in. Our proposed framework is flexible in that it separates the sensory data acquisition from the actual property estimation methods. Such separation allows for redefining the estimation methods with a different set of sensory data than the existing one.

3. **Substitute Selection:** We present an approach to identify relevant properties of a missing tool where we have exploited the relationship between the functional properties and physical properties. In the next step, we have proposed an approach to select a substitute on the basis of relevant properties. Since it is a light-weight approach and do not require any prior training, it is suitable for a run-time response required in tool substitution.

1.5 Publication List

1. **Madhura Thosar**, Christian A. Mueller, Georg Jaeger, Johannes Schleiss, Narender Pulugu,Ravi Mallikarjun Chennaboina, Sai Vivek Jeevangekar, Andreas Birk, Max Pfingsthorn, Sebastian Zug. From Multi-modal Property Dataset to Robot-Centric Conceptual Knowledge About Household Objects.Frontiers in Robotics and AI, 8:87, 2021.

2. **Madhura Thosar**, Christian A. Mueller, Georg Jäger, Max Pfingsthorn, Michael Beetz, Sebastian Zug, Till Mossakowski. Substitute Selection for a Missing Tool Using Robot-Centric Conceptual Knowledge Of Objects. In 35th Annual ACM Symposium On Applied Computing, Brno, Czech Republic, 2020.

3. **Madhura Thosar**, Christian Mueller, Sebastian Zug, Max Pfingsthorn. Towards a Prototypical Approach to Tool-Use Improvisation - Extended Abstract. In International Conference on Autonomous Agents and Multiagent Systems, Montreal, Canada, 2019.

4. **Madhura Thosar**, Christian Mueller, Sebastian Zug. What Stands-in for a Missing Tool?: A Prototypical Grounded Knowledge-based Approach to Tool Substitution. In 11th International Workshop on Cognitive Robotics in 16th International Conference on Principles of Knowledge Representation and Reasoning, Tempe, Arizona, 2018.

5. Georg Jäger, Christian A. Mueller, **Madhura Thosar**, Sebastian Zug, Andreas Birk. Towards Robot-Centric Conceptual Knowledge Acquisition. In Robots that Learn and Reason Workshop in IEEE/RSJ International Conference on Intelligent Robots and Systems, Madrid, 2018.

6. **Madhura Thosar**, Sebastian Zug. From Senses to Knowledge: A Multi-layered Dataset ForGrounded Knowledge About Household Objects - Poster. In International Workshop on Concepts in Action: Representation, Learning, and Application, Osnabrueck, 2018.

7. **Madhura Thosar**, Sebastian Zug, Alpha Mary Skaria, Akshay Jain. A Review of Knowledge Bases for Service Robots in Household Environments. In 6th International Workshop on Artificial Intelligence and Cognition, Palermo, Italy, 2018.

8. **Madhura Thosar**. Rock, Paper, Scissors: What Can I Use In Place of a Hammer (Extended Abstract).In Robotics Fellowship Talk at The Thirtieth AAAI Conference on Artificial Intelligence (AAAI-16),Phoenix, Arizona, USA, 2016.

9. **Madhura Thosar**. Rock, Paper, Sissors: What Can I Use In Place of a Hammer. In Doctoral Consortium at the 38th German Conference on Artificial Intelligence, Dresden, Germany, 2015.

10. **Madhura Thosar**. Can I Use a Sandal Instead of a Hammer?: A Cognitive Approach to a Tool Substitution (Extended Abstract). In Learning Object Affordances: a Fundamental Step to Allow Prediction, Planning and Tool Use? Workshop at The 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems, Hamburg, Germany, September 2015.

11. **Madhura Thosar**. I Don't Find a Hammer, Can I Use a Rock? (Extended Abstract). In Women in Robotics Workshop at The 2015 Robotics: Science and Systems Conference, Rome, Italy, 2015.

1.6 Reader's Digest

The remainder of this thesis is organized as follows:

Chapter 2 begins with the motivation behind proposing the property estimation. In the methodology, we discuss the intent behind the property estimation framework, its components and how do they function together. In the next section, we describe our property estimation methods for six physical and four functional properties along with the grounds for selecting these properties. The chapter proceeds with the description of the dataset generated using the proposed property estimation methods which is used to evaluate different aspects of the estimation methods. The publications related to this chapter are: [6], [7]

Chapter 3 contains the detailed discussion of our conceptual contribution concerning the nature of conceptual knowledge. We will describe the knowledge generation approach from property measurements in the next section. We will also provide the literature review on nine existing knowledge bases to investigate whether they contain the desired nature of conceptual knowledge about objects. The publications related to this chapter are: [8], [6], [9], [10]

Chapter 4 details our approach to identify relevant properties of a missing tool and to select a substitute on the basis of the relevant properties. We will also discuss our proposed definition of a substitute and describe how it is differentiated from a tool. We will perform various experiments to demonstrate the applicability of our proposed substitute selection approach. We will validate our approach by comparing its substitute selection with the experts' selection of substitutes. The main publications related to this chapter are: [11], [9], [10],

Chapter 5 recaps our proposals and approaches followed by a discussion on the open questions related to substitute selection. Additionally, we will also discuss an experiment where we will integrate our substitute selection system with an object perception system.

2
Property Estimation

2.1 Motivation

In case of a tool use, conceptual knowledge about objects is essential for humans as well as for animals. Such conceptual knowledge about objects is also desired in robotic systems (from household to industrial robots) in order to efficiently perform tool-use related tasks such as tool selection or substitute selection [12; 13] with latter task being our primary focus. The conceptual knowledge about an object can take many forms such as temporal relationship between the object and the environment it resides it, spatial relationship with other objects, part-based relationship with its parts, the structure of an object, various functionalities or affordances of objects or various properties of objects such as mechanical properties, chemical properties, geometrical properties etc. We are primarily interested in the conceptual knowledge that considers the properties observed in the objects.

For our research work which focuses on the problem of substitute selection, we took inspiration from the way humans select a substitute from the existing objects for a missing tool wherein humans take into consideration object's physical and functional properties. For instance, consider a scenario in which one has to choose between a plate and a mouse pad as an alternative for a tray. In that regard, a tray will be considered as a rigid, rectangular, flat, wooden, brown colored object while a plate as a rigid, circular, semi-flat, white colored object and a mouse pad as soft, rectangular, flat, leather-based object. This knowledge will be used to determine the similarity between a tray and the other two objects in order to determine a possible substitute. The question is how to acquire such knowledge about the properties. There are three possible ways to go about it: hand-code the desired knowledge OR use existing knowledge bases such as WordNet, ConceptNet OR generate knowledge. Hand-coded knowledge has some well-known limitations, for instance, they are cumbersome to create, problem specific and biased. The existing knowledge bases such as WordNet, ConceptNet do not contain knowledge about the properties of objects (see Chapter 3, Sec. 3.2.2 for the discussion on the contents of WordNet and ConceptNet). The last option is to generate knowledge which raises a question: how should such knowledge be generated. An argument has been put forth in cognitive science for bottom-up generation of knowledge in which humans and animals alike develop conceptual understanding of objects based on their own perceptual experiences with objects [3]. We have followed suit and propose that knowledge about properties should be generated from the sensory measurements of the properties. We have termed such bottom-up generated knowledge from the property measurements as *robot-centric* which is elaborated in detail in Chapter 3, Sec. 3.2.2. The primary application of such robot-centric knowledge, in this work, is to select a substitute for a missing tool. Fig. 2.1 outlines how the substitute selection system is connected to the knowledge acquisition system. The modules *knowledge generation* and *property estimation* are part of the

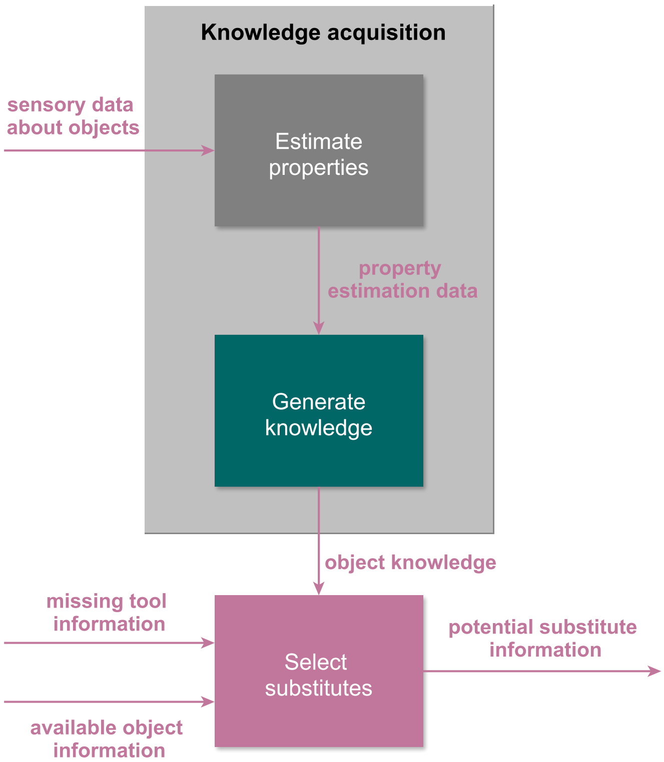


Figure 2.1: The figure shows a typical black box based architecture for substitute selection and the desired knowledge acquisition process wherein knowledge is generated from the sensory measurements of the various properties. Consequently, knowledge acquisition consists of two modules: property estimation and knowledge generation.

knowledge acquisition process which provides the desired knowledge to the *substitute selection* module.

The primary motivation for pursuing a robot-centric aspect stems from the research on cognitive aspects of tool use in humans and animals. Especially the theory that tool selection is a *first-person-perspective* activity which is driven by a relationship between the *user's own conceptual knowledge* about a tool and their ability to use that tool [3]. We noted earlier that one of the aspects of conceptual knowledge that needs to be expressed is subjective knowledge or as we call it *robot-centric knowledge* and in order to capture the subjectivity, the knowledge should be grounded in robot's own sensory perception of objects' properties. As it has been argued in cognitive science studies on concept formation, conceptual knowledge of an object is *grounded* in an individual's multi-modal perceptual and interactive experiences with various objects [14; 15; 16; 17]. This suggests that a conceptual understanding of any object differ from person to person [18]. This also holds true for robots as, in general, robots come in a multitude of perception and manipulation configurations. As a consequence, the individual perception and manipulation of the world similarly varies from robot to robot. Therefore knowledge generated about an object by a KUKA KR1000 Titan (maximum payload of 1300kg, 3.6m reach), for example, will not be the same as knowledge acquired by a Universal Robot UR3 (maximum payload of 3kg, 0.5m reach). Essentially we are suggesting that each robot should gather their own conceptual understanding about objects that they encounter, as the transfer between robots of such subjective understanding may not be desired, especially in the cases of substitute selection. We have elaborated this point further in the discussion on robot-centric aspect of the conceptual knowledge in the Chapter 3, Sec. 3.2.2.

2.2 Methodology

As our primary objective is to generate robot-centric conceptual knowledge about objects from measurements of properties of objects estimated from the sensory data, the question is what properties of objects should be measured. It is postulated in the literature on tool-use in animals [1] that "*a non-invasive tool selection in humans or animals alike is facilitated by conceptual knowledge about objects, especially, knowledge about their physical and functional properties and relationship between them.*" It is not going to be any different for a robot in substitute selection scenarios as conceptual knowledge about objects will allow it to make substitute selection in a non-invasive manner.

As noted earlier, conceptual knowledge about objects, in this case, is considered as a representation of objects in terms of their physical and functional properties generalized over the observations and daily interactions with them. In order to achieve this goal, we propose a multi-layered knowledge acquisition system (see Fig. 2.2) that can be used to generate robot-centric conceptual knowledge about household objects, where each layer is built upon a layer below by abstracting over the lower layer, consequently denoting the different levels of abstraction at each layer.

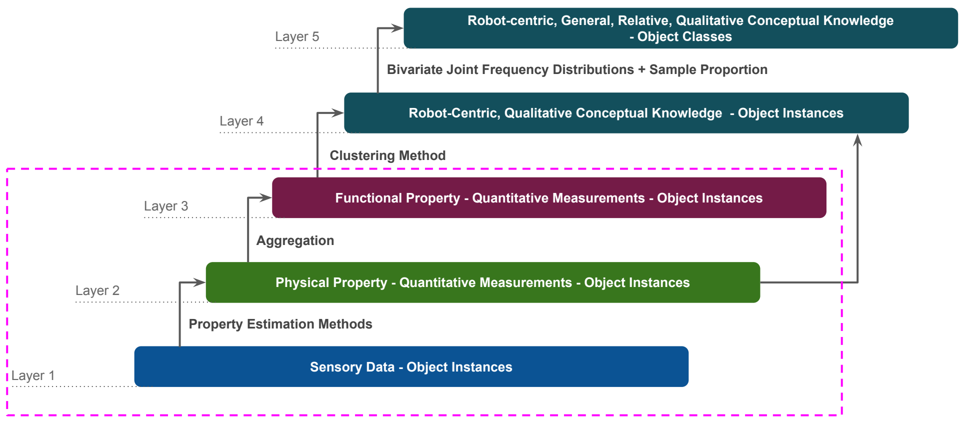


Figure 2.2: The figure illustrates the process layers for our bottom-up robot-centric knowledge generation where each layer is abstracted over the layer below. In this chapter, we focus on the bottom three layers, enclosed within a pink colored boundary, which focus on sensory data extraction followed by the physical and functional property estimation.

According to our proposal, acquisition of the robot-centric conceptual knowledge in a bottom-up fashion should take place in two steps: First, we capture the sensory data about various physical properties of objects. The sensory data is then processed to estimate *quantitative* measurements of physical and functional properties observed in the objects which is then used to generate the desired knowledge about objects. In this chapter, we discuss our approach to extract the sensory data and estimate quantitative measurements of properties of objects. Figure 2.2 illustrates such a multi-layered system wherein the bottom three layers form the proposed sensor data extraction of the properties of objects followed by property estimations. The top two layers form the proposed generation of robot-centric conceptual knowledge about objects which is discussed in the next chapter. In the following, we will discuss how the property estimation is approached and implemented in this work.

2.2.1 Property Estimation Framework

We propose a property estimation framework called **Ro**bot-**C**entric Data**s**et (RoCS) framework that contains multiple property estimation methods which can be used *to estimate the measurements* of various physical and functional properties of objects. The primary objective behind the pro-

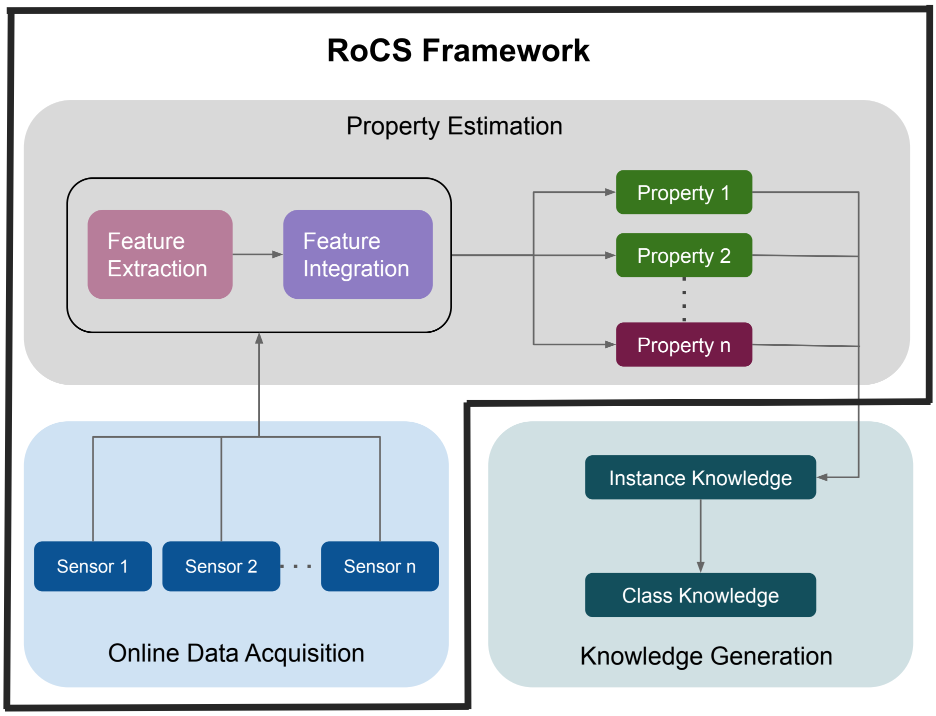


Figure 2.3: The figure depicts the RoCS property estimation framework for extracting sensory data related to various properties which is then used to generate robot-centric conceptual knowledge about objects

posed framework is to build an *extensible* system such that it *encases* various property estimation methods that can estimate the measurement of a property *from a single object instance* in real-time. In order to attain that we have proposed expert-defined estimation methods which estimate a property measurement in a single object instance. In contrast, data-driven models typically need many training examples for each property in order to estimate its measurement in a single instance which may not be a feasible solution.

Fig. 2.3 illustrates the modular structure of the RoCS framework and the resulting measurement data to be supplied to the *Knowledge Generation* system. It primarily consists of two modules: *Online Data Acquisition* and *Property Estimation*. The *Online Data Acquisition* module is responsible for capturing the raw sensory data using different sensors from a single object instance in real-time. Fig. 2.4 illustrates the implementation of both the modules. For instance, in the figure, we have captured the sensory data from an RGB-D sensor, a robotic arm and a household scale. The sensory data is supplied to the *Property Estimation* module which consists of three phases as depicted in Fig. 2.3. In the *Feature Extraction* phase, the desired features are extracted from the sensory data. These features form the basis for estimating a property measurement in an object instance. For instance, in Fig. 2.4, the *Feature Extraction* phase contains six features extracted from the RGB-D sensor data, three features from a robotic arm and a single feature from a scale. Note that the selection of sensors and the subsequent features to be extracted from the sensory data depend on how a property is computationally defined for estimating the measurements. We have elaborated this aspect in the next Sec. 2.2.2 where we discuss estimation method for each property. The extracted features are integrated in the *Feature Integration* phase to form the primary parameters required for estimating the measurements of the properties. For instance, in Fig. 2.4, the *Feature Integration* phase illustrates which features are integrated for each property estimation. The last phase consists of computing the quantitative measurements using the proposed expert-based estimation methods discussed in the next section (see Fig. 2.4 - *Physical and Functional Property Estimation*). The quantitative measurements of the properties are then forwarded to the *Knowledge Generation* system for generating conceptual knowledge about objects.

It is worth noting that a property can be computationally defined in myriad number of ways as it depends on what sensors are used, what kind of sensory data is available, how is it processed, what kind of features are extracted and how are they integrated. One can design an estimation method to calculate absolute measurements of a property in an object instance using physics based methods, on the other hand, one can calculate approximate measurements of the property using available sensor driven expert-defined estimation methods. We have opted for a latter approach due to the fact that each robot model is equipped with a different set of sensory capabilities. As a result, note that, it is possible that a property estimation method designed for an iCub robot [19] may not be applicable for a PR2 robot [20]. Our ultimate vision, therefore, is to develop an online system where developers can plug-in their estimation methods (simple or more complex) for the same property or a new property to the framework requiring minimal or more sophisticated experimental set-ups. The idea

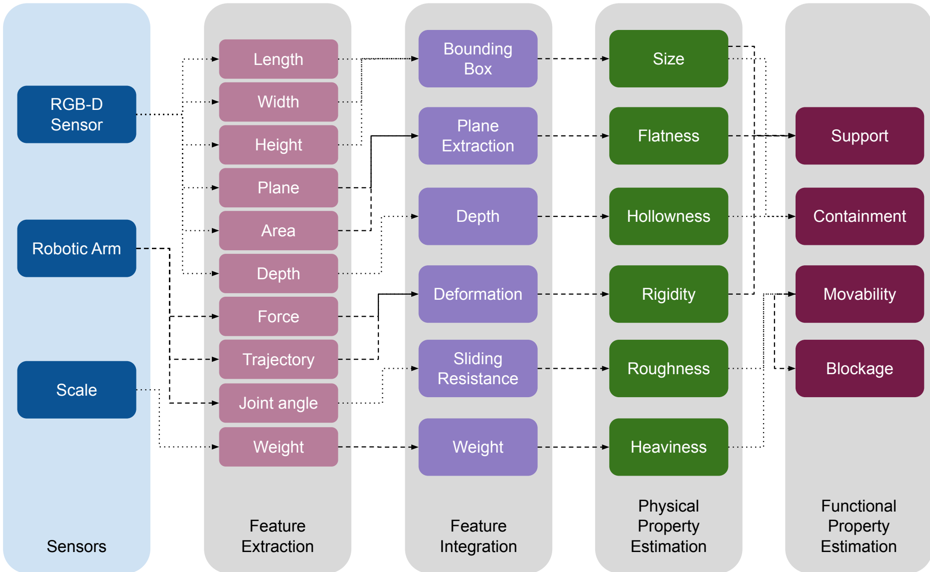


Figure 2.4: The figure depicts the progress of data at each step in order to estimate a property. The steps include sensory data acquisition, feature extraction, feature integration and physical property estimation followed by functional property estimation.

is to allow users to select the estimation methods or to design new estimation methods based on the available hardware at their end in order to estimate the measurements of the object properties. In order to achieve this goal, our proposed framework is designed such that it separates the sensory data acquisition from the actual property estimation methods in order to offer the flexibility. The decoupling of data acquisition, feature extraction and feature integration allows the flexibility for redefining the existing property estimation methods or proposing estimation methods for new properties with a different set of sensory data than the existing one. Additionally, such flexibility also allows developers to extract different set of features from the sensory data as opposed to the existing features OR re-purpose the features for redefining the properties or design estimation methods for new properties. For instance, in the current system, hollowness is defined on the basis of depth. However, it can be redefined on the basis of size and depth as well. Such flexibility, in our opinion, is necessary for robot-centric measurement acquisition since, as noted earlier, sensory and manipulation capabilities vary from robot to robot. The desire for flexibility is also driven by one of the pressing issues which is interpreting the meaning of the properties. The meaning can be complex where various facets of a property and their relationship to the various parts of an object are perceived and interpreted accordingly, or it can be primitive or simplistic. In either case, *the interpretation of a property forms a basis for computational definition of the property and the resulting designing of a hardware set-up and a subsequent estimation method.* Additionally, the proposed framework is also used to create a multi-layered dataset about household objects where the layers denote the different levels of abstraction (See Fig. 2.2).

2.2.2 Property Estimation Methods

How is a property measured in an object instance? The question seems trivial as, in many cases, we do have access to mathematical formula to calculate the measurements of a property in the object instance. Let us take *size*, for instance, of the object. How do we measure *size* of the object? The answer depends on how do we specify an object in a space which leads to the question how many dimensions are needed to specify the space. If we consider one-dimensional space, then the object could be viewed as a line and as a result, the *size* of an object would be a *length* of the line. In case of a two-dimensional space, the object could be viewed as a two dimensional shape and the *size* would be the area of the two dimensional (2D) shape. The formula, however, to calculate the area of a 2D shape depends on what kind of shape it is, such as, circle, rectangle, triangle. The above discussion demonstrates how seemingly a simple property like size depends on so many parameters. The task gets even more complicated when we want to automate the estimation of the size of the object with the help of the sensors deployed on a robot. There are various factors that need to be considered: how do we define size? what parameters do we have to consider to estimate the size? what sensors are available? how can they be used to obtain the sensory values of the parameters? etc. In this section, we will focus on the definition of the properties, the parameters or features we have considered, and the proposed estimation methods for obtaining their quantitative measurements in the object instances. Currently the framework supports the estimation of six physical properties namely rigidity, roughness, flatness, size, hollowness and weight, and four functional properties namely containment, blockage, movability and support.

In this work, when interpreting the meaning of the properties, simplistic interpretations were formed which allowed for a minimal set-up and lightweight estimation methods. The primary inspiration for simplistic interpretation of the properties - which forms the basis for estimation methods - is a level of understanding of properties demonstrated by animals as reported in various literature on tool use in animals [4; 21; 22; 23; 24; 25]. The intent behind a simplistic approach is that, for instance, it allows the use of a simple mobile manipulator whose limited capabilities can be exploited. Additionally, such a minimal experimental set-up can easily be reproduced as they do not require high-end robotic platforms. Our proposed methods estimate the properties from a single instance at a time and do not require any prior training data for estimation, in contrast to the methods proposed in [26; 27; 28; 29]. Moreover, the proposed methods do not require any complex manipulation or grasping capabilities as opposed to some approaches [30; 31]. The notion of the physical properties is based on the physical properties in solid-state physics, where they are considered as properties which can be observed, measured and quantified. We have extended the notion of functional properties in the similar fashion where they are measured and quantified. Depending on how a property is computationally defined, a measurement is either a scalar value or a vector. The selection of these properties are inspired by literature on the tool use in humans and animals [3; 4; 21; 22; 23; 24; 25; 32; 33; 34; 35; 36].

In the following, each property is described in a two-fold manner. First, for each property a general *definition* is provided where we aim for a *simplistic* and *intuitive* characterization for each property. The property definitions considered in this work are not unique. The proposed framework can be extended by plugging in separate estimation methods for the same property based on more complex and/or different characterizations. Second, for each property an *estimation method* is proposed. Note that, as a result of the rudimentary nature of the property estimation methods, the assigned labels are chosen such that they relate to the property definitions as closely as possible. This, in some cases, results in abuse of terminology, for instance, the estimation of size can also be renamed to shape or the estimation for flatness can be renamed to surface area. Note that, although the property definitions are formulated from a human perspective, our ultimate aim is towards enabling a robot to assemble its own understanding about objects, given its own perceptual capabilities in form of vision and manipulation feedback. Hence, we have derived estimation methods allowing a robot to interpret its *sensory data* about objects for generating numeric representations of *physical* and *functional* properties (see Fig. 2.4 for reference). While the presented methods consider features acquired from our robotic platform (Kuka youBot [37], see Fig. 2.6) and an RGB-D sensor (Asus Xtion Pro [38]), we aim to propose a light-weight set-up (Fig. 2.5) and methods that are transferable and adoptable to other robotic platforms by considering common hardware interfaces and data representations such as images, point clouds or joint states of robotic manipulators. We may summarize, that the following proposed estimation methods represent a possible mechanism to express these properties to achieve a continuous-valued property feedback. Depending on the robot capabilities, various estimation methods can be introduced based on different modalities such as vision, tactile or auditory feedback. Therefore, first and foremost, the following methods serve as a possible basis to receive feedback of the targeted properties from a robotic perspective (robot-centric).

Physical Properties

Humans tend to conceptualize tools in terms of their function, i.e., the outcome that a given kind of artifact, due to its designed physical structure, helps to bring about when used in goal-directed actions. [36]. In other words, in order to enable any functionality in an object, a certain assemblage of physical properties are essential prerequisites [3]. For humans, the first step towards understanding this causal relationship is by assessing various physical properties of an object and examining the functionalities enabled by them [36].

In this work, we have selected *flatness, hollowness, size, roughness, rigidity* and *heaviness* as physical properties given their significance reported in the literature on tool use in humans and animals[3; 33; 34; 39]. The main inspiration behind selecting these properties was the prominent roles these properties played in various tool use scenarios in humans and animals alike, as widely reported in the literature. For instance, human infants begin exploring their abilities to use any object by studying and interacting with it to understand its *weight*, *texture*, and *shape* [34]. While designing and manufacturing a tool, humans and animals alike pay closer attention to the properties such as *shape*, *size*, *rigidity*, *roughness*, and *heaviness* [3]. It has been observed that wild animals select the tools based on the *size*, *shape* or mechanical properties such as *strength*, *hardness* [33]. For example, otters have been observed carrying *flat* rock on their chest which they use to break the shellfish [39]. On the other hand, researchers found that the monkeys are able to select the hardness of the stone with respect to the hardness of the nut they want to cut open [40].

In the following, we provide a *definition* for each physical property and subsequently an *estimation method* is proposed for each property. Note that, across all estimation methods, we assume that an object is placed in its most natural position, for instance, a cup is most commonly placed in such a way that its opening points upwards. Additionally, the estimation methods are designed such that the resulting measurements can be used to generate knowledge and subsequently to select a substitute. Therefore, we aim at a bounded property value, i.e. an estimated property value that is mapped into a $[0,1]$ interval, by means of a normalization process, in order to enable a subsequent unbiased property analysis which is not affected by object-specific characteristics or scales. Such bounded values provide the abstractions over the feature values desired to generate the conceptual knowledge about objects. Note that, as a prerequisite, each object is segmented a priori through a table-top object segmentation procedure, particularly for the *size*, *flatness* and *hollowness* property. Moreover, estimated property values of each object are captured through the given capabilities of the robot in form of vision-based (e.g. featuring particular image, point cloud resolution or viewpoint) as well as manipulation-based (e.g., featuring particular joint-states, limits or force-feedback) input. As a result, these property values are originated from a robot-centric perspective on the perceived objects. Note that, as the property measurements of objects are to be obtained from the sensory data, the estimation method and the resulting measurements of an object instance are relative to: the sensors being used; viewpoint of the sensors such as camera from which an object instance is being observed, and the position of an object instance. Alteration in any of the above factors will affect the subsequent property measurements. From an object instance perspective, it means

Figure 2.5: Light-weight experimental setup consisting of two cameras and fiducial markers [41], for acquiring physical properties.

that the measurements of any property represent only a partial perspective of the instance. This is especially true in the cases of *size*, *flatness*, *hollowness* and *roughness*.

Size

Definition: *Size* of an object is defined intuitively by the object's spatial dimensionality in form of *length*, *width* and *height*.

Estimation Method: The size of an object is defined by the length, width and height. As it can be estimated by determining an object's bounding box, we use an RGB-D sensor to obtain point clouds of the object from a lateral perspective. Using marker detection to define a region of interest (ROI), we segment the object and transform its point cloud to an axis-normal representation, i.e. the z-axis is aligned with the object's height. Subsequently, an axis-aligned bounding box is approximated given the extracted object point cloud. The *size*=[*length*, *width*, *height*] of an object is directly derived from the object point cloud as distances between the minimal and maximal value in each spatial dimension of the bounding box. In order to retrieve a bounded property value range $[0,1]$ for the property *size* (si), each spatial dimension of *size* [*length*, *width*, *height*] is normalized by the largest dimension of the object (*max(size)*) (see Eq. 2.1). As a result, si is defined as a three dimensional property.

$$si = \left[l = \frac{\text{length}}{\max(\text{size})}, w = \frac{\text{width}}{\max(\text{size})}, h = \frac{\text{height}}{\max(\text{size})} \right] \tag{2.1}$$

Note that, *max(size)* merely abbreviates *max(length, width, height)*. In this implementation, the size is represented by a vector that can be interpreted as an aspect ratio of bounding box of a 3D object.

Flatness

Definition: As *flatness* describes a particular aspect of an object's shape, we define it as the ratio between the area of an object's greatest horizontal plane and its overall surface area. For instance, a sheet of paper features an upper bound of *flatness* whereas a ball features a lower bound of *flatness*.

Estimation Method: The *flatness* value of an object is estimated similarly to its *size*: We firstly observe the object from above (Fig. 2.5) and extract its greatest plane using RANSAC (RAndom SAmple Consensus [42]). In order to increase the confidence, a candidate plane is only selected if at least 95% of the surface normal vectors of the plane points are directed in the same direction, up to a threshold. In this manner, round surfaces (as they may be observed in *balls*) are rejected and subsequently a *flatness* value of zero is assigned to the considered object. Furthermore, if the candidate plane p is accepted, the plane size $|p|$, i.e the number of object points corresponding to p, is divided by the total number of points $|o|$ representing the observed object o in order to obtain a bounded numeric measure of its *flatness* fl (Eq. 2.2). Consequently, the retrieved *flatness* property is bounded within a value range of $[0,1]$.

$$fl = \frac{|p|}{|o|} \tag{2.2}$$

Hollowness

Definition: *Hollowness* is the amount of visible cavity or empty space within an object's enclosed volume. It contrasts *flatness* as it focuses on a another particular aspect of an object's shape.

Estimation Method: *Hollowness* contributes to the characterization of object shape. According to its definition, an object may enclose a volume which is not filled. For the sake of simplicity, we measure the internal depth d, which resembles the enclosed volume, and height h of an object o: the ratio defines the *hollowness* value. In order to retrieve a reasonable measure of object's depth and height, a two camera and fiducial marker [41] setup is introduced as illustrated in Fig. 2.5. Given the side camera view, the height h of an object can be obtained by estimating the respective bounding box (see Section 2.2.2). In order to retrieve depth, two fiducial markers $\{m_r, m_h\}$ are introduced (see samples in Fig. 2.5): m_r serves as global reference and is placed next to the object; m_h is placed inside the hollow volume of the object. Exploiting the top camera height c_t perpendicularly pointed to the object, the distances $d_r = \|m_r - c_t\|$ and $d_h = \|m_h - c_t\|$ can be obtained. Given object height h and the distances d_r and d_h, hollowness ho can be approximated as shown in Eq. 2.3b, where b (Eq. 2.3a) is introduced to consider the base height of the object, i.e. distance between the table (global reference plane) and the bottom inside the object's hollow volume.

$$b = d_r - d_h \tag{2.3a}$$

$$ho = \frac{h - b}{h} \tag{2.3b}$$

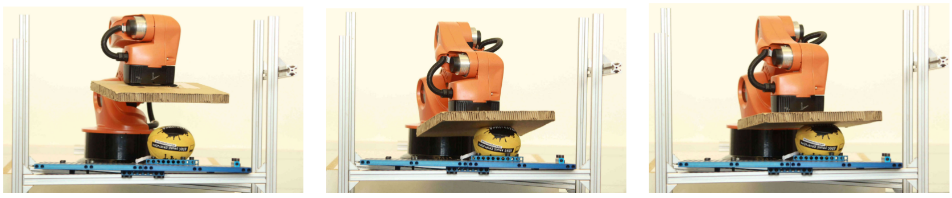
Note that, ho is inherently bounded within the interval $[0, 1]$. Furthermore the proposed method may be susceptible to noise originated in the point clouds from which the bounding box was approximated to infer the object's height h. Hence, if the difference between an object's height h and distance d_h (fiducial marker inside the object) is smaller than 1cm it is cumbersome to differentiate between sensor noise and the actual *hollowness* due to the low signal-to-noise ratio. To sanitize the property in such situations (particularly in case of flat objects), default value of zero is assigned.

Heaviness

Definition: Following our basic premise of using straight forward definitions, we borrow the definition of *heaviness* from physics: the object's *heaviness* is the force acting on its mass within a gravitational field.

Estimation Method: *Heaviness* he of an object o can be directly derived by weighing an object with a *scale* (Eq. 2.4); a scale with a resolution of 1g provides an adequate precision for our scenario. Note that, he is normalized by the carrying capabilities of the robotic arm.

$$he = scale(o) \tag{2.4}$$



Rigidity extraction at time t=1 Rigidity extraction at time t=2 Rigidity extraction at time t=3

Roughness extraction at time t=1 Roughness extraction at time t=2 Roughness extraction at time t=3

Figure 2.6: Light-weight experimental setup consisting of a camera-manipulator combination, for acquiring physical property *rigidity* (top row) and *roughness* (bottom row).

While it may require additional hardware, a robot may lift an object and calculate the *heaviness* by converting the efforts observed during the process in each of its joints.

Rigidity

Definition: *Rigidity* of an object is defined as the degree of deformation caused by an external force vertically operating on it.

Estimation Method: *Rigidity* of an object is estimated using a robotic arm. The arm is equipped with a planar end-effector that is used to vertically exert a force onto an object until predefined efforts in the arm's joints are exceeded, see Fig. 2.6; by setting the predefined efforts to the limits of the robotic arm, the final rigidity value is specific to the robot executing the estimation method. During this process we record the trajectory $tr(t)$ of the arm as well as the efforts in all of its joints. By analyzing them using an adaptive threshold-checking, we detect the first contact of the end-effector with the object o at time t_0. Using the final position of the arm when the efforts are exceeded at t_1, we can calculate the deformation *def* of an object as the vertical movement of the end-effector, that is, its movement along the z-axis between t_0 and t_1:

$$def(o) = tr_z(t_0) - tr_z(t_1) \tag{2.5a}$$

$$ri = \frac{def(o)}{h} \tag{2.5b}$$

In that way, the deformation *def*(o) is nothing but the distance the arm pushed into the considered object. For rigid objects, this deformation is zero while it is increased continuously for non-rigid objects. Finally, we normalize the deformation by the height h of the object to obtain its *rigidity* value *ri*. As we use a distance as a measure of an object's deformation, *def*(o) will always be positive. Furthermore, as an object may not be deformed more than its own height, the value of *ri* is naturally bound to the interval of $[0,1]$.

Roughness

Definition: *Roughness* provides information about an object's surface. Therefore, we simplify the physical idea of friction and define *roughness* as an object's resistance to sliding.

Estimation Method: *Roughness* ro requires interaction as well to measure an object's resistance to sliding. The robotic arm is exploited to act as a ramp on which the considered object is placed, see Fig. 2.6. Starting horizontally, with an initial angle of $a_i = 0°$, the ramp's angle is increased and thereby causes an increasing gravitational force pulling the object down the ramp. When the object begins sliding, a fiducial marker that is a priori placed underneath the object, is unveiled and subsequently detected. As this means that the object's sliding resistance is exceeded, the ramps' angle a_r is observed and exploited as a measure of *roughness* as shown in Eq. 2.6. In this setup, a 90° ($\frac{\pi}{2}$) ramp angle represents the upper bound that induces an object to slide. Hence, it is used to normalize *roughness* value ro within $[0,1]$.

$$ro = \frac{|a_i - a_r|}{\frac{\pi}{2}} \tag{2.6}$$

Functional Properties

In contrast to physical properties, functional properties describe the functional capabilities or affordances [43] of objects. It is proposed that functional properties do not exist in isolation, rather certain physical properties are required to enable them [44]. In tool use, functional properties play an important role especially when perceiving an object as a possible tool, since humans in general characterize an object in terms of its functional properties rather than its physicality [43; 45]. The question is how does a functional property emerge? In other terms, what are the required qualifications for an ability to be recognized as a functional property? Various theories have been proposed to address this question [43; 45; 46] and among them is a theory proposed by Kuhn in [47]. According to Kuhn in [47], *image schema* capture the necessary abstractions to model functional properties. *Image schema* is a theory proposed in psychology and cognitive linguistics and it concerns with a recurring pattern abstracted from the perceptual and motor processes [47]. Some of the examples of image schema are *containment, support, path,* and *blockage* which form the basis for functional abilities to *contain, support, move,* and *block* respectively.

As we noted earlier, the physical properties which are selected in this work are fundamental in nature. For functional properties, we wanted to follow suit. However, the question is, how to identify fundamental functional properties of objects. This is where the theories of image schema come in the picture. Image schemas are expressed as generic conceptual building blocks for concepts [48; 49]. For example, in [50], it is demonstrated how abstract concepts in mathematics could be broken down into the bodily experiences and image schemas. Commonly mentioned image schemas are: *containment,* the notion that objects can be within other objects; *support,* the notion that objects can rest on top of other objects; and *source-path-goal,* the notion of object movement along a trajectory between two different points. Kuhn in [47] suggested that image schemas in (some) cases can model the essential properties of objects. Based on our research, we have selected the following functional properties as fundamental properties. To *contain* is the ability of objects to hold within themselves other objects which is based on the image schema *containment.* It is one of the most investigated image schema and it appears in different levels of specification [51]. The system presented in this thesis takes a straight forward interpretation of containment as either full or partly enclosed. *Support* is another essential object relation for many objects. Similar to containment, *support* has different levels of specifications [47], however, we focus on the surface based support. For example, *support* appears as a necessary functional property for objects such as tables and trays that have the main function of carrying/supporting other objects on their surfaces. *Move* is one of the most fundamental [52] functional properties of any object derived from the image schema *path*. The last functional property is *block*, which captures the notion of hindered movement of one object. While *block* is derived from the image schema *blockage*, the schema itself is a type of an abstract image schema called *force*. Like *path*, it is also considered as one of the most fundamental schema [53].

Note that functional properties of objects are also known as affordances of objects, which form a widely researched area in robotics. Affordances in robotics is primarily focused on learning a computational model of an affordance in terms of actions a robot can perform on an object and its effects observed on the object using the available sensors [54; 55]. Our proposed approach bypasses learning affordances of objects, as we are primarily interested in the quantitative measurements of affordances.

It is suggested in [3] that a certain assemblage of physical properties is essential prerequisite to enable a functional property and such knowledge is used by humans and animals alike in tool selection. We have exploited this notion and have designed our substitute selection approach around it (See Chapter 4 for detailed discussion). It is stated in [2] that designing a tool is not an arbitrary act, but rather requires thoughtful consideration of many factors such as the purpose of the tool, the intended user of the tool, the assumption about the physical capabilities of the use. Therefore, a tool is designed or rather should be designed in such a manner that it is comfortable to handle, manipulate and use in order to achieve the intended outcome [2]. It is no surprise that the discussion on the relationship between functional properties and its enabling physical properties are notable in the discussion on tool design. Especially how the presence of physical properties in a tool can affect the functionality of the tool. Some of the notable examples noted in [2] are: 1) the weight of a tool will affect the lifting of the tool, for example, power drills or hammer; the author even suggested that "for most users, a weight of around 4.5kg represents the maximum load for manipulation and handling."; 2) the roughness of a surface of a tool will affect the movability or the sliding ability of the tool; 3) length of a tool matters when the tool requires swinging action such as in a hammering action; 4) shape allows different types of gripping; 5) size, texture and the weight of a tool can block the movement of another object such as a stone blocking the movement of a door; One can also find similar discussion in [39] where the following observations have been made: 1) size of a hollow object can affect what it can or can not contain; 2) flat surface is needed in a tool to allow other objects rest on it.

As a result, while the measurements of physical properties are computed using mathematical formulations (except for heaviness), for functional properties the same treatment is not used. In this work, a functional property is measured in terms of the measurements of the physical properties that enables it. For example, for *containment*, *size* is relevant as in concrete situations it is not possible for an object to contain a larger object than itself. Likewise, *rigidity* and *weight* are essential properties for the *support*, as the rigidity of a supporting object needs to (on a physical level) correspond to the weight of the object being supported. The mapping between a functional property and its enabling physical properties is derived from the findings reported in the literature on tool use in animals and humans [3; 4; 32; 33; 34; 35; 36]. This approach is therefore based on prevalent theories in cognitive science rather than being data driven as in, for instance, machine learning based methods.

Support

Definition: *Support* describes an object's capability to support, i.e. to carry another object. Therefore, an object is attributed with *support*, if other objects can be stably placed on top of the supporting object.

Estimation Method: *Support* requires to consider three aspects of an object. Firstly, the considered object needs to be rigid. Secondly, for carrying another object, the sizes of both may feature similar spatial proportions. Thirdly, the object's shape needs to be sufficiently flat in order to enable the placing of another object on top of it. Consequently, *size*, *flatness* and *rigidity* are considered as core elements of the *support* property, Eq. 2.7.

$$su = [si, fl, ri] \tag{2.7}$$

Containment

Definition: An object is attributed with *containment* if it is capable to enclose another object to a certain degree.

Estimation Method: *Containment* property requires to consider two aspects. In order to contain something, an object needs to be *hollow*. On the other hand, it's *size* itself needs to be respected when considering whether it can contain another object. Thus, the value of the object's *containment* co property is formed by combining its *size* and *hollowness* property values, Eq. 2.8.

$$co = [si, ho] \tag{2.8}$$

Movability

Definition: *Movability* describes the required effort to move an object.

Estimation Method: *Movability* is based on a robot's primary ways of moving objects: either by lifting or pushing. In both cases, *heaviness* of an object affects the *movability* of an object. Additionally, when pushing an object, its sliding resistance expressed in form of *roughness* (see Fig. 2.6), needs to be considered as well. Therefore *movability* property mo constitutes of *heaviness* and *roughness*, Eq. 2.9.

$$mo = [he, ro] \tag{2.9}$$

Blockage

Definition: *Blockage* describes the capability of an object of being impenetrable, i.e. the object cannot be moved by other objects, therefore it stops the movement of other encountered objects.

Estimation Method: *Blockage* of an object can be derived from its *movability*. Note that, given the set of physical properties, we can interpret that the *blockage* property is related to *roughness* and *heaviness* of an object as these properties affect the intensity of being capable to block another object. Accordingly, *blockage* property bl states to which degree an object is able to stop another object's movement. Thus, the object itself needs to be not movable by the other object, which is the inverse of its *movability*, Eq. 2.10.

$$bl = -mo = [-he, -ro] \tag{2.10}$$

2.3 ROCS Dataset

For the sake of a thorough evaluation of our conceptual framework, the **Ro**bot-**C**entric data**S**et (RoCS) is introduced. Note that we propose a Robot Operating System (ROS) [56] based implementation to acquire object data used in the following evaluation. It consists of 11 different object classes where each class consists of 10 unique object instances that leads to a total number of 110 object instances. In the following, we briefly introduce the hardware setup and procedures for acquiring raw object data, describe its parameters (e.g. thresholds) and the contents of the final dataset.

2.3.1 Hardware Setup

Figure 2.4 illustrates the sensors we used as data sources. For visual and non-invasive estimation methods, RGB-D sensors are required. More specifically, the *size* property requires a lateral view on objects while the *hollowness* property relies on a birds-eye view. Hence, we employ two Asus Xtion Pro depth sensors [38] (see Fig. 2.5). To estimate the physical properties *rigidity* and *roughness*, a robotic arm is required to interact with objects. In this interaction the proposed property estimation methods require arm joint state values which are generally provided by manipulators, such as the one we have used: a Kuka youBot [37] manipulator. Finally, a common kitchen scale with a resolution of 1g is used to estimate the weight for *heaviness* of objects.

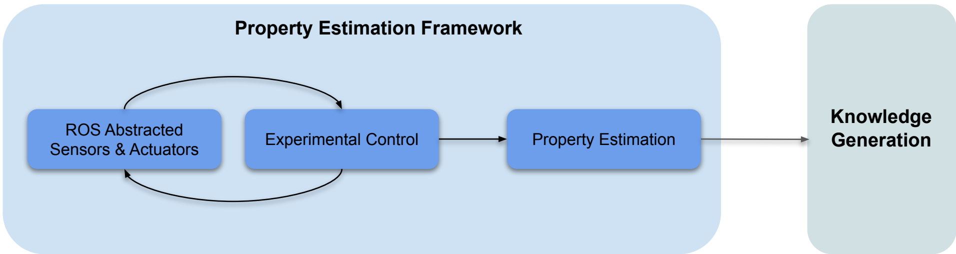


Figure 2.7: Data flow within the dataset creation framework.

2.3.2 Object Property Acquisition Procedure

Using the described hardware, we implemented a ROS-based framework to estimate the physical and functional properties of objects. A schematic overview of the framework is given by Fig. 2.7.

The interface for operating sensors and actuators is provided to our framework by ROS. This interface is used by different experiments for observing and interacting with objects to acquire the necessary sensory data. Together, both blocks (*ROS Abstracted Sensors & Actuators* and *Experiment Control*) form a control loop enabling to generate feature data (see Fig. 2.4). According to the selected properties, four control loops are implemented as separate experiments. The first experiment is non-invasive and gathers the visual feature data required for *hollowness*, *flatness* and *size*; Fig. 2.5 illustrates the camera setup. Initially a table-top object detection is introduced that uses a RAndom SAmple Consensus (RANSAC) based plane fitting approach in order to detect object candidates on the table. The RANSAC algorithm is parameterized with a leaf size of 0.0025m, a maximum of 10^4 iterations and a $0.02m$ distance threshold between points and the estimate plane model. Note that, RANSAC is also used in this experiment for segmenting planes for the property *flatness*. Furthermore, fiducial markers (ArUco Library [41]) with sizes of 14 cm and 3 cm are used for the *hollowness* property. The second experiment uses the robotic arm to deform objects to facilitate the estimation of *rigidity* (see Section 2.2.2). We set the efforts to exceed in each joint to ±8 Nm. Within the third experiment, the robotic arm is used as a ramp to estimate an object's *roughness* (see Section 2.2.2). To achieve an appropriate resolution, the angular speed of the joint lifting the ramp is set to 0.05 rad/s. Finally, the last experiment employs a kitchen scale with a resolution of 1g to estimate the ob-

Figure 2.8: RoCS dataset samples: Point cloud and RGB images of a *ball*, *bowl*, *paper box*, and *cup* (for visualization purposes, images are scaled and 3D points are magnified).

jects' weight. Following the *Experiment Control*, the individual estimation methods process the generated feature data as described in Section 2.2.2 to estimate physical and functional property values of the considered object. Finally, this data can be accumulated for a set of objects and further processed to generate conceptual knowledge.

2.3.3 Dataset Structure

For the RoCS dataset we consider 11 different object classes (*ball*, *book*, *bowl*, *cup*, *metal_box*, *paper_box*, *plastic_box*, *plate*, *sponge*, *to_go_cup* and *tray*) featuring various object characteristics – from appearance to functional purpose. Each class consists of 10 unique object instances that leads to a total number of 110 object instances; Fig. 2.8 illustrates sample object instance of RoCS dataset.

In order to evaluate the performance of the proposed property estimation methods, such as stability, for each object instance we capture 10 repetitions without modifying the setup. As a result we captured 1100 object observations for which physical and functional property values are generated. The dataset is publicly available as a git repository. Please check the appendix D for the git links.

2.4 Evaluation

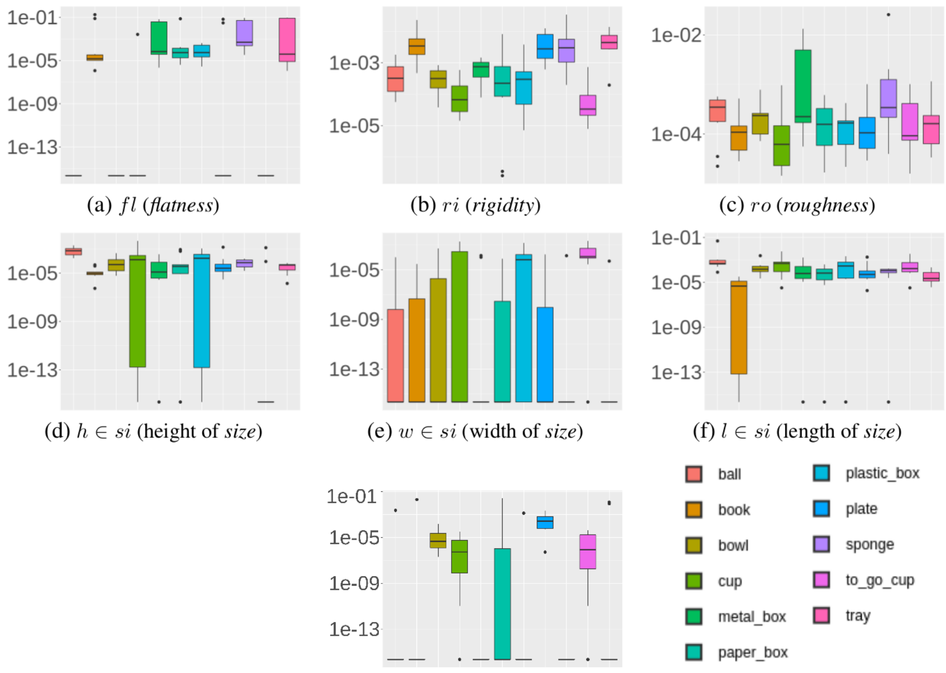
The ultimate application of the estimated measurements of the properties of objects is to generate conceptual knowledge about objects which, in the end, will be used for substitute selection purposes. It is, therefore, vital that the estimation methods are performing efficiently and are producing the quality results. The lack of either will affect the quality of the conceptual knowledge and the subsequent substitute selection for a missing tool. In the following, we have discussed various experiments which are performed to evaluate the proposed estimation methods and their estimated measurements. We investigate various aspects of our proposed approach such as the stability of the estimation methods, the quality of the estimated measurements and the semantics of the estimated measurements.

2.4.1 Property Estimation

The objective of the evaluation is to analyse the estimations computed by the property estimation methods as described in the Section 2.2.2. At this level, we only focus on physical properties as functional properties are built on the basis of an object's physical properties. First, we analyze the stability of the estimation methods to determine how *deterministic* and *reproducible* the estimation is for each property and object. Furthermore, we explore the coverage of our data set to determine the *variance* and range of objects reflected in the different classes and properties. Lastly, we inspect the *correlation* among different properties in our data.

Estimation Stability

The abstraction process from raw sensor data to symbolic object property knowledge requires a stable processing. However, sensors are influenced by external and internal factors which can affect the quality of the sensory data [57]. The resulting variations in the quality of data is often called as noise which can affect ultimately the overall quality of the measurement data. To compensate for such uncertainty caused by the noisy data, the property estimation of each object instance consists of 10 repetitions. We use these repetitions in the following to analyze the stability of the proposed property estimation methods. For that, the variance of each physical property of each object instance is analyzed. More specifically, given the 10 repetitions of a particular object instance for each of its physical properties, we calculate the *variance of the property values of its 10 repetitions.* As the measurements of 6 physical properties are based on 8 features, we obtain 8 values per object instance and therefore 880 values in



(a) fl (*flatness*) (b) ri (*rigidity*) (c) ro (*roughness*)

(d) $h \in si$ (height of *size*) (e) $w \in si$ (width of *size*) (f) $l \in si$ (length of *size*)

(h) ho (*hollowness*) (i) legend

Figure 2.9: Variances for physical properties [*fl*, *ri*, *ro*, *si*, *he*, *ho*] illustrated in form of a box plot (in log-scale to provide insights of respective intra property variances compared to linear-scale shown in Table 2.1). Note that, in order to be able to display all variances (including zero) in log-scale, we add an epsilon on each value before computing log. Heaviness is excluded as all variance values are zero for this property due to the resolution of the scale.

total. We further reduce the data, by calculating the *mean of the object variances* for a particular object class and property as shown in Table 2.1, whereas Fig. 2.9 illustrates the variances of all object instances within one object class as box plots; the colored middle box represents 50% of the data points and the median of the class is indicated by the line that divides the box. It is worth noting that the box plots illustrates the effect of a noise in the sensory data during the estimation of a property measurement for each instance. In an ideal case, if the noise is absent, the variance of measurements would be zero for each object instance as evident by the measurements of heaviness. If, however, the noise is present, the variance will be greater than zero. As a result, higher the variance is, more is the noise in the sensory data which in turn will affect the quality of resulting measurements.

Table 2.1: Mean variance for each physical property.

class	flatness	rigidity	roughness	size_length	size_width	size_height	heaviness	hollowness	class_mean
ball	0	0.00053	0.00032	0.00538	0.00001	0.00083	0	0.00023	0.00091
book	0.02554	0.00583	0.00015	0.00001	0.00001	0.00002	0	0.002	0.00419
bowl	0	0.00037	0.00025	0.00038	0.00006	0.00012	0	0.00003	0.00015
cup	0.00026	0.00015	0.00017	0.00098	0.0003	0.00079	0	0.00001	0.00033
metal_box	0.01939	0.00074	0.0039	0.00028	0.00002	0.00007	0	0	0.00305
paper_box	0.00747	0.00115	0.00021	0.00011	0.00002	0.00017	0	0.0035	0.00158
plastic_box	0.00015	0.00071	0.00016	0.00056	0.00021	0.0003	0	0.00013	0.00028
plate	0.00971	0.00481	0.00022	0.0003	0.00003	0.00017	0	0.0005	0.00197
sponge	0.02503	0.00705	0.00313	0.0001	0.00001	0.00008	0	0	0.00443
to_go_cup	0	0.00016	0.00031	0.00061	0.00044	0.00013	0	0.00001	0.00021
tray	0.03486	0.00569	0.00024	0.00005	0.00001	0.00004	0	0.00206	0.00537
prop_mean	0.01113	0.00247	0.00082	0.0008	0.0001	0.00025	0	0.00077	0.00204

Each value represents the mean variance of estimated property values of an particular object class consisting of 10 instances and their respective repetitions. Variances are scaled by color in ascending order from transparent (0) to red (highest variance).

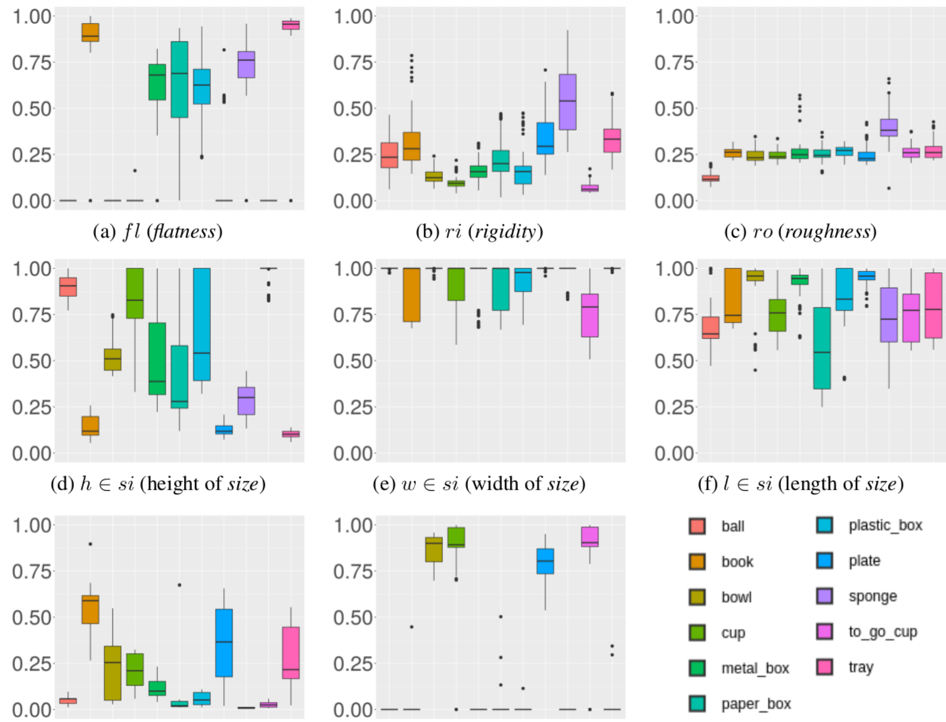
The results of the Table 2.1, elaborated in the Fig. 2.9 reveal that the class variances are overall low, which implies stable property estimation methods in general. The highest variances can be found for the *flatness* property. The estimation of the *flatness* property for small and flat object instances such as sponge, plastic box, is particularly affected by noise due to the low signal-to-noise ratio. Furthermore, it can be observed that for *ball*, *bowl* and *to_go_cup* the variance of the *flatness* property is zero due to the fact that no top-level plane can be extracted for instances of these classes as they feature either round or negligible small top-level surfaces (see Section 2.2.2). Similarly, a higher variance can be observed for the *rigidity* property which is caused by the thinner object instances (object instances with shorter height), such as *book*, *plate*, *sponge* and *tray*. Here the detection of the first contact with the object causes false positives and therefore introduces varying deformation values.

In contrast, for the *hollowness* property the variance for *metal_box* and *sponge* are zero. Such object instances predominantly feature flat surfaces and negligible degree of hollowness. Considering sensor quantization effects, such negligible degree for hollowness cannot be confidently distinguished from sensor noise under such conditions (see Section 2.2.2). As a consequence a default hollowness value of zero is set for instances that fall in a negligible range of hollowness, i.e. below 1cm distance between marker. Concerning the *heaviness* property, a zero variance is observed due to the accurate measurement by a scale – considering a resolution of 1g which is a sufficient resolution for our scenario.

Property Coverage of RoCS

The objective of this experiment is to evaluate the intra-class variance for each property in order to determine the range of data covered in each object class for one particular property. For this experiment, the mean estimated property value over the 10 repetitions is used. The result for each of the physical properties is shown in Fig. 2.10 in form of a box plot in which all object instances of a particular class are considered.

Several observations can be made. For instance, *hollowness* and *flatness* are complementary in our dataset. Objects with *flatness* values close to zero are commonly exhibiting increased *hollowness* values (above 0.5) and



(a) fl (*flatness*) (b) ri (*rigidity*) (c) ro (*roughness*)

(d) $h \in si$ (height of *size*) (e) $w \in si$ (width of *size*) (f) $l \in si$ (length of *size*)

(g) he (*heaviness*) (h) ho (*hollowness*) (i) legend

Figure 2.10: Category-wise coverage for each physical property.

vice versa. Only balls form an exception as they are neither flat nor hollow. While this means that we cover a wide range of values for the *flatness* property, we miss such coverage for *hollowness* values in the interval $[0, 0.5]$. Moreover, for *roughness* most object classes are in a similar range – except sponge and ball instances. As we place the objects in their most natural position we can conclude that the sponges' ground surfaces have a higher *roughness* due to their open-pored surfaces. Due to their roundish surfaces, ball instances feature obviously a low *roughness* value. Furthermore, it is unlikely to observe objects featuring *roughness* values close to one as none of the considered object classes has the ability to *stick* to the ramp.

For the *rigidity* values an interval of $[0, 0.9]$ is covered, ranging from rigid objects such as *metal_box* to non-rigid objects such as *sponge*. Suspiciously, only a limited number of objects has a value of zero which indicates that sensor noise has its greatest effect on these objects. Analyzing the *size* values, it becomes apparent that *width* commonly is the greatest

Table 2.2: Pearson Correlation on the mean values of physical properties

	flatness	**rigidity**	**roughness**	**s_length**	**s_width**	**s_height**	**heaviness**
flatness	-						
rigidity	0.45	-					
roughness	0.45	0.35	-				
size_length	0.03	0.12	0.15	-			
size_width	0.16	0.34	0.02	0.21	-		
size_height	-0.65	-0.59	-0.38	-0.26	-0.45	-	
heaviness	0.09	-0.04	-0.13	0.19	0.02	-0.37	-
hollowness	-0.71	-0.36	-0.08	0.24	-0.1	0.24	0.13

dimension among the considered objects while the objects' height varies along the range of possible values.

Property Correlation

In this experiment, we investigate the linear correlation in the physical properties of our data. In the natural world some of the properties are not correlated such as size and roughness or rigidity and heaviness or roughness and rigidity. On the other hand, properties such as hollowness and flatness may show negative correlation due to the lack of flat solid surface in case an object is hollow, for instance, a cup or a bowl. The objective of this experiment to investigate whether two properties are correlated with each other in the dataset as a result of our proposed estimation methods and due to the estimated measurements of the properties.

In order to investigate the linear correlation, we performed Pearson correlation [58] on the data. Given estimated values of a particular property, we compute the mean property value $\overline{o_x}$ (Eq. 2.11a) over the 10 repetitions for each object instance o. Based on these mean variances, the pearson correlation ρ_{XY} is obtained between two sets of mean variances X and Y corresponding to respective properties, see Eq. 2.11b, where cov is the covariance and σ_x the standard deviation of X, respectively.

$$X = \{\overline{o_{x_1}}, \overline{o_{x_2}}, \overline{o_{x_3}}, ...\} \tag{2.11a}$$

$$\rho_{XY} = \frac{\text{cov}(X, Y)}{\sigma_x \sigma_y} \tag{2.11b}$$

Table 2.2 shows the pearson correlation among all physical properties with a color scale.

It can be observed that the correlation of our data is low in general. However, a strong negative correlation between *flatness* and *hollowness* is found which may indicate that in our data objects with high flatness are likely to have low hollowness. This matches our observation in Section 2.4.1, where we noted the complementary nature of these properties in our dataset. The object instances of our dataset may also show some negative correlation between *size-height* and *flatness* as well as *size-height* and *rigidity* which, as we have noted earlier, is the result of the false positives where shorter object instances are regarded as being rigid.

2.4.2 Property Semantics

Given a stable property estimation (Section 2.4.1) from noisy real world data, the following experiment focuses on the semantic interpretation of the estimated measurements of the properties of objects. More specifically, the question is, when categorized on the basis a property with varying number of categories, how robustly object instances from various object classes share similarity. In other words, we are interested to examine object instances of which object classes would be contained in each category when categorized on the basis of various properties. In that sense, we can view these categories as *artificial object categories (AOCs)* which would be created from a robot's perspective. Additionally, we also want to examine, with different properties, how the categorization of the instances vary. We propose an experiment that categorizes object instances of our RoCS dataset in an unsupervised manner by considering a particular set of properties. Given the small dataset to be categorized, it is essential that every single data is categorized and not left out. For this reason and to conduct a preferably unbiased (machine-driven) categorization, we used a clustering technique K-means as a baseline technique where it is applied by gradually increasing the value of k=$\{2,\ldots,11\}$. Here, 11 is selected as an upper bound as it represents the number of object classes considered in the RoCS dataset.

Fig. 2.11 focuses on the categorization of the object instances on the basis of the functional properties. On the other hand, the Fig. 2.12 focuses on the categorization of the object instances on the basis of non-visual properties which include roughness, rigidity, heaviness; visual properties containing flatness, hollowness, size; and finally on the basis of all the physical properties. Both figures consist of pyramid charts that shows the gradual categorization process for the respective property. A category or a cluster is depicted as a pie-chart illustrating the distribution of assigned object instances with their labeled class. Therefore, each row of the pyramid-like structure shows the results of one application of the k-means clustering. The number of pie-charts in each row equals to number of clusters (k value). As stated earlier, each categorization on the basis of the properties can be viewed as an *artificial object category (AOC)* who share certain the respective property-based similarity. In other terms, since each category partitions the property space, assigned instances within the category share similar degree of the property. For instance, in case of a partitioning on the basis of *containment* in the Fig. 2.11, the object instances grouped together share a similar degree of *containment*, similarly the object instances grouped together in the Fig. 2.12 on the basis of visual properties, share a similar degree of visual properties.

Generally on higher levels in the pyramids (lower k), the distribution of the instances and the classes in each *AOC* is higher compared to lower levels (higher k). It can be interpreted as, *AOCs* on higher levels appear to feature more generic attributes since the distribution of object classes is higher compared to lower levels (higher k) while lower levels encompass *AOCs* featuring more specific attributes. As a consequence, *AOCs* in the higher levels appear more generic as opposed to the lower levels where *AOCs* appear more specific. Moreover, as the levels in a pyramid progresses in both figures, patterns in the distribution of classes can be observed which are carried forward in the subsequent levels. Such patterns in each pyramid hints towards semantic relations between class labels and *AOCs*. For example, instances of *bowl*, *cup*, *to_go_cup* share similar *AOCs* regarding the *containment* property (see *AOC* annotated with ○ in Fig. 2.11) which is also reflected over multiple levels. Such emerging patterns can also be observed and tracked over multiple levels in other pyramids in Fig. 2.12, for instance, for all the physical properties, pattern containing the instances of *plastic_box*, *metal_box*, *paper_box*, and *sponge* (see *AOC* annotated with ○). Notably, a pattern observed in one pyramid will not be necessarily observed in every other pyramid which indicate that with properties, the categorization of object instances vary accordingly. For instance, the aforementioned pattern containing the instances of *bowl*, *cup*, *to_go_cup* is not

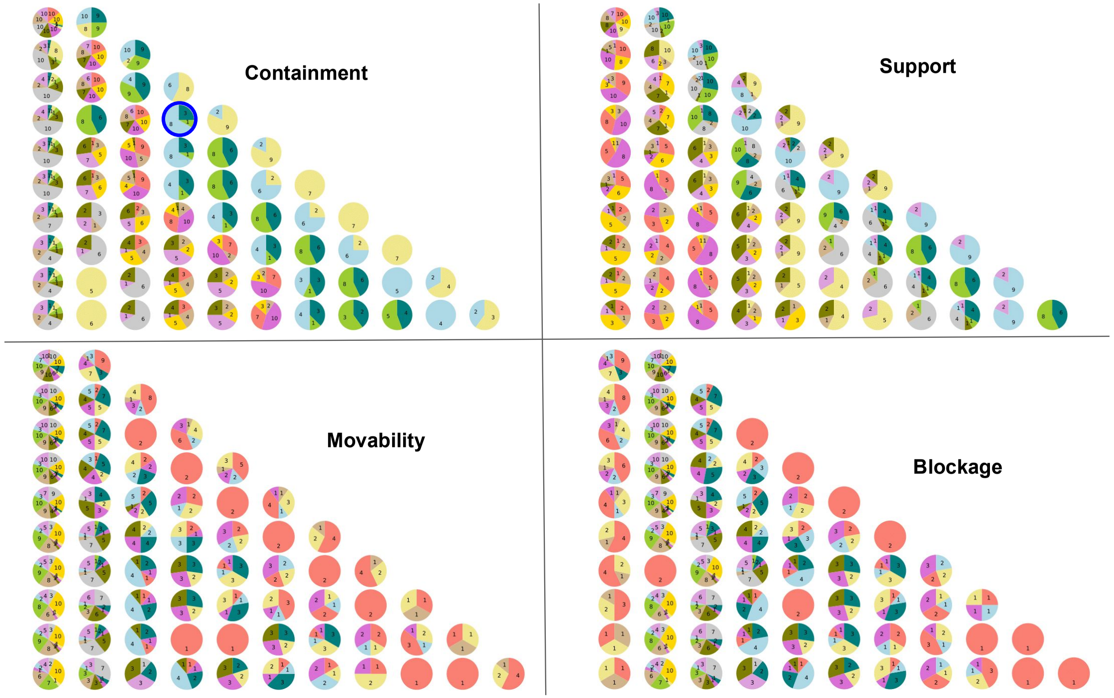


Figure 2.11: Gradual categorization of instances to particular artificial object category (*AOC*) given a particular functional property describing each instance. Each *AOC* is illustrated as a pie chart showing the object class label distribution of instances assigned to the respective *AOC*. A sample *AOC* is annotated as circle (○) which illustrates a grouping of instances of object classes *bowl*, *cup*, and *to_go_cup* featuring similar attributes regarding the property *Containment*. The legend is provided in the figure 2.12.

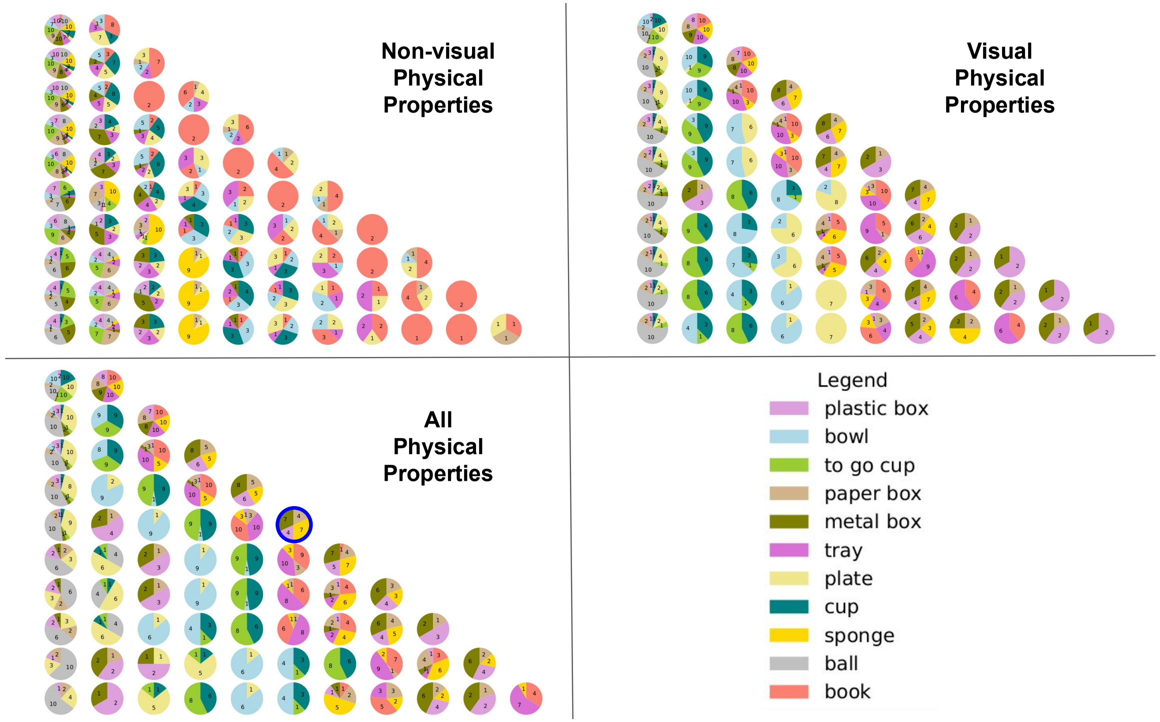


Figure 2.12: Gradual categorization of instances to particular *AOCs* given a set of visual properties (flatness, hollowness, size), or a set of non-visual properties (roughness, rigidity, heaviness) or set of all physical properties describing each instance.

reflected in the *movability* or *blockage* pyramid. This suggest that the patterns are closely associated with a property along which categorization of object instances is performed. It is also worth noting that the number of object classes in AOCs in each level in all the pyramid structures varies. In some cases, an AOC contains only one object class, for instance, in containment, movability, blockage, visual and non-visual properties. In some cases, an AOC contains more than six object classes, for instance, in movability, blockage and non-visual properties. We also noticed a tendency in all the pyramid structures where all the instances of an object class are grouped together across all the levels, for instance, *tray* in containment, *bowl* in support (almost all the instances), *sponge* in movability, blockage and non-visual properties, *ball* in visual and all the physical properties.

As a result, the proposed property estimations may allow to *describe* object instances encompassing a variety of characteristics – from appearance to functional purposes – and also allows to *discriminate* the object instances on the basis of the properties. In the figure, property generality can be observed across certain object classes, i.e. *AOCs* on different granularity levels may feature dedications to instances of different object classes as they feature similar characteristics or trends regarding the property. This interrelation of object classes is reflected by the heterogeneity of the distribution of instances within an *AOC* – even in case of k=11 when considering 11 object classes. On a critical note, some *AOCs* may seem questionable, for example, in movability and blockage, two *AOCs* across all levels contain eight object classes. On the other hand, in movability, blockage and non-visual properties, there is an *AOC* that contains only a one or two instances of the class book. We can also note that, in the upper three levels, the categorization of object instances is more discriminating than the lower most levels. As we increase the number of clusters for categorization, it does get difficult to group together the instances of the same class. However, we can also observe that in almost all the pyramid-structures, across all the levels, there is at least one *AOC* where at least one object class dominates in terms of the number of its instances assigned to that *AOC*. Moreover, it can also be observed that the property measurements has allowed the meaningful groupings of certain object classes from a human perspective. For instance, in containment and support, across all the levels, the instances of *cup* and *to go cup* are grouped together, or *tray* and *book* are grouped together; the similar observations can be made in visual and all physical properties; in movability and blockage, across all the levels, *sponge*, *to go cup* and *paper box* are grouped together.

Note that, this experiment is not aimed at finding the optimal number of clusters for each property based categorization or perform an accurate object categorization on the basis of each property. Instead, our focus is to investigate the semantics of the property measurements or how meaningful the property-based categorization of object instances is from a human's perspective. Consequently, these insights gained from this experiment provided us with a basis for the generation of conceptual knowledge about objects and the subsequent substitute selection which are discussed in the next chapters.

2.5 Conclusion

Property Estimation Framework

In this chapter, we propose an *extensible* property estimation framework called **Ro**bot-**C**entric Data**s**et Framework (RoCS) wherein multiple property estimation methods can be used to measure various physical properties and functional properties. Currently, the framework consists of six physical properties and four functional properties. Our proposed framework is flexible in that it separates the sensory data acquisition from the actual property estimation methods. Note that our framework is mere a skeleton that deploys a decoupling approach to process the sensory data and estimate the measurements of the properties observed in the objects. Such separation allows for redefining the estimation methods with a different set of sensory data than the existing one. Our ultimate vision is to develop an online system where developers can plug-in their estimation methods (simple or more complex) for the same property or new property to the framework requiring minimal or more sophisticated experimental set-ups. This way, we wish to create a community of users who can select the estimation methods based on the sensor and robot availability at their end. Additionally, the proposed framework is also used to create a multi-layered dataset about household objects where the layers denote the different levels of abstraction (Fig. 2.2 for reference). The property measurement data generated by the proposed framework can then be used to generate robot-centric symbolic knowledge.

Properties Estimation Methods for Physical Properties

In this work, we propose light-weight estimation methods for physical properties rigidity, hollowness, size, flatness and roughness, whereas functional properties are estimated in terms of the measurements of the physical properties that enable them. Our proposed methods estimate the properties from a single instance at a time and do not require any prior training data for estimation, in contrast to the methods proposed in [26; 27; 28; 29]. Moreover, the proposed methods do not require any complex manipulation or grasping capabilities as opposed to some approaches [30; 31]. Our proposed methods are light-weight, requiring minimal experimental set-up where even simplistic robotic platforms with just an arm and minimum sensors, such as YouBot, can be used for estimation. However, it should be noted here, we do not claim that our proposed methods are the only way or a better way to estimate the properties of objects, which is why, a proposed framework also allows for additional estimation methods for the same properties besides the one that already exist. The primary inspiration for interpretation of the meaning of properties - which forms the basis for estimation methods - is a level of understanding of properties demonstrated by animals as reported in various literature on tool use in animals.

The estimation of physical properties for substitute selection is not unique to our work. For instance, in the research work on tool selection in robotics, various properties are used for selecting a tool such as length [59; 60; 61], width [60], shape [60], a function label [59; 60], or hand-coded symbolic knowledge about geometric properties of tools [13]. Similarly, in substitute selection approaches, various properties are estimated and used such as metric data about position, orientation, size, and symbolic knowledge about hand-picked relations such as *similar-to* and *capable-of* extracted from ConceptNet [62]; visual and physical understanding of multi-object interactions demonstrated by humans [63]; metric data about size, shape and grasp, as well as a human estimate of an affordance score for task + mass [64]; attributes and affordances of objects are hand-coded using a logic-based notation, and a multidimensional conceptual space of features such as shape and color intensity [65]. Moreover, the application of physical property estimation is not limited to generate conceptual knowledge, substitute selection or tool selection. Various approaches for estimating physical properties of objects such as rigidity, shape, texture, size, etc. have been proposed for applications such as object recognition/categorization, grasping and manipulation [26; 27; 28; 29; 30; 31; 66].

In summary

In this chapter, we discussed our approach to estimate the physical and functional properties observed in the objects. The approach consists of two constituents: 1) property estimation framework; 2) property estimation methods. The estimation methods and the resulting measurements were investigated by performing various experiments to analyse various aspects such as variance of the property measurements, intra-class variance, correlation between the properties and the semantic relations between object instances. The primary application target of our proposed approach is to generate robot-centric conceptual knowledge about objects (Chapter 3) which will be used in substitute selection for a missing tool (Chapter 4).

3
Knowledge Generation

3.1 Introduction

Given how vital the tool-use ability is, robotics researchers have been developing approaches to enable a robot to use tools in various tasks [12; 61; 67; 68; 69; 70; 71; 72]. While these approaches focus on learning tool-use behavior, one of the underlying issues in tool-use is *how can an appropriate tool be selected that is required in a task?* In the literature on tool selection in robotics, the proposed selection mechanism is often integrated with a proposed tool-use system in which a robot selects an appropriate tool and uses it in a pre-designed task [13; 59; 60; 61]. All the ap-

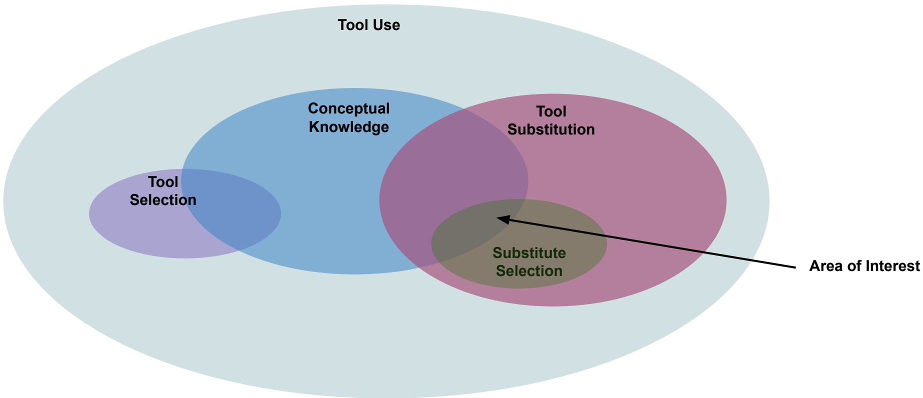


Figure 3.1: The figure shows our primary area of interest within the domain of tool-use. Conceptual knowledge is desirable in tool use, however our focus is on generating conceptual knowledge required for substitute selection. The figure also illustrates the positioning of substitute selection within tool-use. While tool-use also consists of areas such as grasping, planning, manipulation, validation etc. we have left them out for the sake of clarity. Besides our primary area of interest, our intent is to distinguish it from tool-use or tool selection. Note that in tool selection a robot does not have any prior knowledge of what tool is appropriate in a given task whereas in substitute selection, the robot does have such prior knowledge.

proaches use machine learning techniques where training examples typically include parameters used for selecting a tool such as length [59; 60; 61], width [60], shape [60], a function (affordance) label [59; 60], or hand-coded symbolic knowledge about geometric properties of tools [13]. A typical problem scenario in tool selection in the literature is - from the given tool options, what is the most appropriate tool for a task at hand. But what if the required tool is missing? This is the question we are primarily interested in (see Fig. 3.1).

In order to be recognized as a plausible substitute for a missing tool, the substitute needs to be similar to the missing tool in some way without having to interact with it. The question is what is needed to determine the similarity. In the literature on substitute selection, typically a substitute for a missing tool is determined by means of knowledge about objects, and the knowledge-driven similarity between a missing tool prototype and a potential substitute. Such knowledge about objects varies in its contents and form across the literature: metric data about position, orientation, size, and symbolic knowledge about hand-picked relations such as *similar-to* and *capable-of* extracted from ConceptNet [62]; visual and physical understanding of multi-object interactions demonstrated by humans [63]; metric data about size, shape and grasp, as well as a human estimate of an affordance score for task + mass [64]; attributes and affordances of objects being hand-coded using a logic-based notation, and a multidimensional conceptual space of features such as shape and color intensity [65]; hand-coded models of known tools in terms of superquadrics and relationships among them [73]; potential candidates extracted from WordNet and ConceptNet if they share the same parent with a missing tool for predetermined relations: *has-property*, *capable-of* and *used-for* [74]; hand-coded object-action relations [75]; as well as hand-coded knowledge about inheritance and equivalence relations among objects and affordances [76].

While for tool selection, metric data of certain properties are primarily considered, where such data is extracted in real-time from an instance of an object, for substitute selection, symbolic knowledge about the object category or class is considered. In such cases, either the proposed approaches use existing common sense knowledge bases such as WordNet, ConceptNet or knowledge is hand-coded. Regardless of the use of existing knowledge bases or hand-coded relational knowledge, it is apparent that knowledge about objects in some form is sought by the aforementioned approaches. What also seems to be a common theme in the literature is that some form (metric or symbolic) of physical and functional understanding of objects is used as essential drivers for substitute selection as well as for tool selection. This is in line with the findings noted in the literature on tool use in animals and humans [3; 21; 22; 34; 77; 78] which state that conceptual or semantic knowledge about object forms the foundation for not only tool use but also for tool selection and substitute selection.

3.2 Building Blocks

As we noted, conceptual knowledge about objects is desired in robots (from household to industrial robotics) in order for substitute selection, where selection is facilitated by the knowledge about various (physical and functional) properties observed in the objects [12; 13]. We stated in the Sec. 1.2 and will elaborate further in Sec. 4.1.2, that our primary inspiration behind our proposed approach to substitute selection is based on the notion of relevant properties of a missing tool. In our approach, the determination of the relevant properties and the subsequent selection of a substitute on the basis of the relevant properties is aided by conceptual knowledge about objects. Therefore, in order to utilize such conceptual knowledge about objects (e.g. in a household environment) for substitute selection, the following questions need to be answered:

- What should conceptual knowledge about objects be constituted of for substitute selection?
- How can such conceptual knowledge about objects be acquired?
- How should the acquired knowledge be represented?

These questions form the primary building blocks of our work, namely: *Conceptual Knowledge*, *Robot-centric Knowledge*, and *Knowledge Representation*. In this section, we address how the building blocks are realized in this work.

3.2.1 Conceptual Knowledge

It is postulated in the literature on tool-use in animals [1] that non-invasive tool selection in humans or animals alike is facilitated by conceptual knowledge about objects, especially, knowledge about their physical and functional properties and relationship between them. For instance, knowledge about what physical properties of a hammer enable the hammering action can facilitate the decision between a stone and a plastic bottle as a substitute. It is not just humans, but animals such as crows, chimps have demonstrated that in order to perform tool use, a tool is conceptualized in terms of its physical as well as functional abilities [3; 4; 21; 22; 23; 24; 25; 34].

One of the key components in substitute selection by human experts is the knowledge about physical properties of an object [79]. It is postulated that conceptual knowledge about objects is generalized over our observations and daily interactions with them [3]. As a consequence, while on one hand humans tend to express an object in linguistic form by giving it a label such as a mug [80], on the other hand an object label is not merely a reference to the corresponding physical entity in an environment but also incorporates knowledge about its physical and functional abilities [3; 43; 81]. Therefore, based on our observations and interactions with various instances of a cup, conceptual knowledge of the *cup* may, for example, consist of an object that is rigid, hollow, cylindrical, made up of ceramic material and also has a primary function, for instance, hold liquid [82]. For a robot to select a substitute in a non-invasive manner, we propose that conceptual knowledge about objects should consists knowledge about physical and functional properties observed in objects.

But is knowing merely "whether a cup is rigid or not" enough? Consider, for instance, a choice between a cup and a stone as a substitute for a hammer. While both the objects are rigid, we have *general knowledge* that a stone is *usually more rigid* than a cup and quite possibly as rigid as a hammer. As a result, we will choose the stone over the cup for hammering. Take another instance where we have to choose between two stones of different sizes and weight for hammering. It is possible that some people may *prefer* one stone over the other for various reasons. For some people, one stone has a better shape which allows them to grip better for hammering while some people *prefer* a stone that is less heavy than the other and for some, they may *prefer* the smaller sized stone for better grasping. Another example is the choice between a mobile phone and a plate as a substitute for a tray to carry a drink. Since both the objects are flat, they should be viable substitutes. However, since we *know* that a plate is *usually larger in size than* a mobile phone, and a plate is *closer to* a tray *in size* than a mobile phone is, we will vote for the plate. There are two pieces of information worth noticing: firstly, our knowledge about properties of objects is *generalized, relative, subjective* and *qualitative*, and secondly, the selected substitutes are not necessarily visually similar to the missing tools but are rather qualitatively similar. We are proposing a similar approach to conceptual knowledge about objects for substitute selection.

The conceptual knowledge about objects, in this work is characterized as *qualitative*, *generalized*, *relative*, and *subjective*. In the following, we provide the proposed interpretation of the aforementioned characterizations of the conceptual knowledge.

Generalized: The conceptual knowledge about objects primarily consists of knowledge about an object class as opposed to a specific instance of the object class. However, it is derived from the instances of the object class.

Relative: The general knowledge about any object class is based on its instances that have been encountered. It means, the knowledge about any object class is not absolute and is subject to change as more instances of the object class are encountered.

Subjective: The knowledge about an object class is acquired from the sensory experiences and interaction with the object's instances as opposed to extracted from other sources. In our work, the subjective knowledge is knowledge acquired from a first-person perspective.

Qualitative: The knowledge about object classes is expressed in terms of the properties where the properties observed in the objects are interpreted qualitatively as opposed to quantitatively. However, the qualitative knowledge about properties of objects is obtained from the quantitative data about properties of objects.

In summary, the conceptual knowledge about objects, in this work is characterized as *generalized* over a robot's observation and interaction with the objects; *relative* to the encountered objects by the robot; *subjective* with respect to the robot's sensory experiences and interaction with objects; and *qualitative* in the representation of knowledge. The primary contents of the proposed conceptual knowledge about objects consist of the properties observed in the objects. The properties are divided into *physical* and *functional* properties where the physical properties describe the physicality of objects and the functional properties ascribe the (functional) abilities or affordances to the objects. In this work, we have focused on the six physical properties: *rigidity, weight, hollowness, roughness, flatness, size* and four functional properties: *containment, blockage, support, movability*.

3.2.2 Robot-centric Knowledge

Given that we need conceptual knowledge about objects that is *general, relative, subjective,* and *qualitative,* the fundamental question is how should a robot acquire such knowledge. One way to go about it, is by using existing knowledge bases such as WordNet [83] and ConceptNet [84] which consist of commonsense knowledge. The commonsense knowledge is considered as knowledge that is commonly shared by *most people* and it usually is considered as implicit in nature, meaning that not every known fact about an object will be included in such knowledge bases [85]. As a result, there is no clear consensus about what falls precisely under commonsense knowledge. Figures 3.2, 3.3, 3.4 and 3.5 show the knowledge about a cup and a plate as described in WordNet and ConceptNet. It can be noted that the pieces of knowledge about both the objects in WordNet and ConceptNet, while sharing some similarity, differ in the contents. Moreover, not everything we know about a cup and a plate is included. ConceptNet provides much broader and categorized overview about a cup and a plate, while WordNet contains more condensed knowledge. It is also worth noting that the knowledge about the two objects does not contain *explicit* information about the physical and functional properties of objects. In order to verify whether such commonsense knowledge base is enough to select a substitute, we performed an experiment discussed in Sec.4.3.3 where we have demonstrated the use of WordNet in substitute selection and have argued that why the WordNet *alone* is not an adequate source of knowledge about objects for substitute selection.

While commonsense knowledge is concerned with commonly known knowledge by most people, subjective knowledge is concerned with knowledge held by an individual. The robot-centric notion proposed in this work is concerned with the subjectiveness of the knowledge. The primary motivation for pursuing a robot-centric notion stems from the research on cognitive aspects of tool use in humans and animals. We are especially interested in the theory that tool selection is a *first-person-perspective* activity which is driven by a relationship between the *user's own conceptual knowledge* about a tool and their ability to use that tool [3]. Here the term *user's own conceptual knowledge* is vital since it deals with the knowledge acquired by individual using their own senses and personal interactive experiences with the objects. As suggested before, such personalized or subjective knowledge plays a crucial role during tool or substitute selection process. It has been argued in the cognitive science studies on concept formation that conceptual knowledge of an object is *grounded* in

Noun

- S: (n) **cup** (a small open container usually used for drinking; usually has a handle) *"he put the cup back in the saucer"; "the handle of the cup was missing"*
- S: (n) **cup**, cupful (the quantity a cup will hold) *"he drank a cup of coffee"; "he borrowed a cup of sugar"*
- S: (n) **cup** (any cup-shaped concavity) *"bees filled the waxen cups with honey"; "he wore a jock strap with a metal cup"; "the cup of her bra"*
- S: (n) **cup** (a United States liquid unit equal to 8 fluid ounces)
- S: (n) **cup** (cup-shaped plant organ)
- S: (n) **cup** (a punch served in a pitcher instead of a punch bowl)
- S: (n) **cup** (the hole (or metal container in the hole) on a golf green) *"he swore as the ball rimmed the cup and rolled away"; "put the flag back in the cup"*
- S: (n) **cup**, loving cup (a large metal vessel with two handles that is awarded as a trophy to the winner of a competition) *"the school kept the cups is a special glass case"*

Verb

- S: (v) **cup** (form into the shape of a cup) *"She cupped her hands"*
- S: (v) **cup** (put into a cup) *"cup the milk"*
- S: (v) **cup**, transfuse (treat by applying evacuated cups to the patient's skin)

Figure 3.2: The knowledge about a cup as described in WordNet.

an individual's multi-modal perceptual experiences with various objects [14; 15; 16]. This suggests that conceptual understanding of any object may differ from person to person, thus making conceptual knowledge of an object a subjective understanding of an object [18; 86]. In other words, the understanding of a property observed in an object depends on one's perception and subsequent interpretation of that property [86]. For instance, a heavy object for one person may be heavier for the other or a big size object for one person may be medium size object for another. As a result, such distinct understanding of heaviness or size of an object may influence one's selection of a substitute where heaviness or certain size is desired. It means the selected substitutes may differ from person to person (see Sec. 4.3.2 for a more detailed discussion). Consequently, such subjective understanding of an object may not be transferable between individuals as they differ in their interpretation of perceived properties in objects. This also holds true for robots in general, as robots come in a multitude of perception and manipulation configurations. As a consequence, the individual perception and manipulation of the world similarly varies from robot to robot. Therefore conceptual knowledge acquired about an

Noun

- S: (n) home plate, home base, home, **plate** ((baseball) base consisting of a rubber slab where the batter stands; it must be touched by a base runner in order to score) *"he ruled that the runner failed to touch home"*
- S: (n) **plate** (a sheet of metal or wood or glass or plastic)
- S: (n) **plate** (a full-page illustration (usually on slick paper))
- S: (n) **plate** (dish on which food is served or from which food is eaten)
- S: (n) **plate**, plateful (the quantity contained in a plate)
- S: (n) **plate**, crustal plate (a rigid layer of the Earth's crust that is believed to drift slowly)
- S: (n) **plate** (the thin under portion of the forequarter)
- S: (n) **plate** (a main course served on a plate) *"a vegetable plate"; "the blue plate special"*
- S: (n) **plate** (any flat platelike body structure or part)
- S: (n) **plate** (the positively charged electrode in a vacuum tube)
- S: (n) **plate**, photographic plate (a flat sheet of metal or glass on which a photographic image can be recorded)
- S: (n) **plate** (structural member consisting of a horizontal beam that provides bearing and anchorage)
- S: (n) **plate**, collection plate (a shallow receptacle for collection in church)
- S: (n) **plate**, scale, shell (a metal sheathing of uniform thickness (such as the shield attached to an artillery piece to protect the gunners))
- S: (n) denture, dental plate, **plate** (a dental appliance that artificially replaces missing teeth)

Verb

- S: (v) **plate** (coat with a layer of metal) *"plate spoons with silver"*

Figure 3.3: The knowledge about a plate as described in WordNet.

object by a KUKA KR1000 Titan (maximum payload of 1300kg, 3.6m reach), for example, will not be the same as conceptual knowledge acquired by a Universal Robot UR3 (maximum payload of 3kg, 0.5m reach). It is worth noting that, despite of the fact that all humans are equipped with the same set of sensory organs, the humans perceive properties of objects differently. In case of the robots, even the vision sensors differ from robot to robot. This is where acquisition of robot-centric knowledge about objects is desired where the underlying idea is, in order to capture the subjectivity, the knowledge should be grounded in robot's *own* sensory perception of objects' properties [87]. Given that substitute selection is influenced by one's understanding of an object, we propose that every robot should be equipped with its own robot-centric knowledge acquisition system which takes into account the sensory system fitted into its system.

Figure 3.4: The knowledge about a cup as described in ConceptNet.

Figure 3.5: The knowledge about a plate as described in ConceptNet.

3.2.3 Knowledge Representation

According to [88], a representation formalism is a medium where knowledge can be organized such that it allows for efficient reasoning. The choice of a specific formalism is driven by a desired outcome and a world in which a robot is operating [89]. As we stated earlier, the conceptual knowledge about objects should consists of knowledge about physical and functional properties observed in the objects. In other words, an object should be *represented* in terms of the physical and functional properties observed in it. For representing conceptual knowledge about objects for substitute selection, we need a formalism that represents the four aspects of such knowledge: *generalized, relative, subjective,* and *qualitative*. In the following, we lay out the foundation for representation formalism we have sought in this work that allows us to represent all the aforementioned aspects.

For *generalized* knowledge, a representation formalism should express knowledge about an object category in terms its properties which is generalized over the properties observed in individual instances. As we noted in 3.2.1 that when representing any object as a concept, humans usually omit quantitative measurements of the properties observed in the object, but assign what we have termed as *qualitative measurements* to the properties of an object to reflect to what degree that property is reflected [80]. For instance, a cup is *generally light* weight, *medium* rigid, and can *fully* contain solid or liquid. This observation is generalized over the observations made in individual instances of the cup. Therefore we need a formalism that takes into account instance-specific knowledge to generate class-specific knowledge about any object. Moreover, as knowledge is to be represented as a set of generalized observations regarding the properties reflected in the instances of an object, a formalism should be able to incorporate the degree with which the property is observed in the object.

The generalized knowledge about any object category should be *relative* to a robot's experience with different instances of the respective category and other categories too. Consider, for instance, *size* of an object. Our understanding of size and variations in size is relative to the different sizes we observe in different objects we encounter over period. As a result, our understanding of any property is subject to change as we encounter more objects and their size. This also entails that our understanding of size *generally* observed in an object (category) may change as we encounter more instances of the object category. Accordingly, we need a formalism that can incorporate such relativeness of understanding of size globally (regardless of an object category) and locally (object category specific). Moreover, the formalism should be adaptable and tractable, since the conceptual knowledge about objects needs to be updated as the robot acquires experiences with new instances of the known object category or a new object category.

In order to acquire subjective or robot-centric knowledge, it is necessary that the knowledge is grounded in the robot's own sensory perception of the properties of objects. This grounding process is typically called a symbol grounding process which bridges the gap between symbolic knowledge and sensory perception by creating a correspondence between them. This correspondence either refers to a physical entity in the real-world a.k.a. perceptual anchoring [90] or assigns a meaning to a symbol by means of a respective sensory-motor process [91] (*what I sense is what I know*). We propose that in order to capture robot-centricness in conceptual knowledge, it should be generated in a bottom-up fashion, that is, knowledge is generated from the sensory data about objects. In order to achieve that, first, we capture the sensory data about each property from various objects using robot's sensory systems. The sensory data is then processed to estimate *quantitative* measurements of properties observed in objects using the property estimation method discussed in chapter 2. The quantitative measurements are instrumental in bringing about the desired qualitativeness in the knowledge wherein they are used to generate property specific *qualitative* measurements. Conceptual knowledge about objects is then generated for given objects on the basis of the qualitative measurements of various properties. This very bottom-up generation of qualitative knowledge from the quantitative measurements of properties of objects is behind the notion of robot-centric knowledge. Such bottom-up knowledge generation not only grounds the knowledge into robot's sensory perception but it is generated from the data acquired using the robot's sensory system.

For our proposed knowledge base, we deemed the symbolic formalism based on the notion of fuzzy sets [92] and attribute-value pairs as suitable formalism. The knowledge base is provided with a set of symbols which form the vocabulary using which the proposed knowledge base is built. The question we are trying to address here is: How can we generate robot-centric knowledge on the basis of robots' sensory perception and a given vocabulary or a set of symbols for properties and object categories?

3.3 Methodology

As suggested in Sec. 3.2.2, the robot-centric aspect of the knowledge in this work lies on the notion that for symbolic knowledge to be grounded in robot's perception of the world, it is to be generated by abstracting over the robot's perception data. In order to attain that, we have proposed a bottom-up method which consists of five levels of abstractions as illustrated in Fig. 3.6. The bottom three levels, covered in the chapter 2, focus on gathering the sensory data and estimating the properties measurements on the basis of the sensory data. For each property, we have uploaded the estimated measurements in the git repository. Please refer appendix D for the git links. The top two levels, on the other hand, concentrate on generating the conceptual knowledge about objects from the property measurements. For generating robot-centric conceptual knowledge, the data about the objects' physical and functional properties is processed in two stages: *sub-categorization* (Layer 4) and *conceptualization* (Layer 4 and 5) which was implemented in Python programming language. In the following, we discuss each stage in detail.

Consider **O** as a *given* set of object classes where (by abuse of notation) each object class is identified with its label. For example,

$$\mathbf{O} = \{\ \textit{cup, bowl, book, plate, ball}\ \}$$

where the labels *cup, bowl, book, plate, ball* represent the said object classes.

Let each object class $O \in \mathbf{O}$ be a *given* set of its instances where each instance is identified with a label. For instance, an object class *Cup* is a set of the labels of its instances given as,

$$\text{cup} = \{\ \textit{cup\_1, cup\_2, cup\_3,...}\ \}$$

Let $\bigcup\mathbf{O}$ be the union of the sets of all object classes such that $|\bigcup\mathbf{O}| = n$.

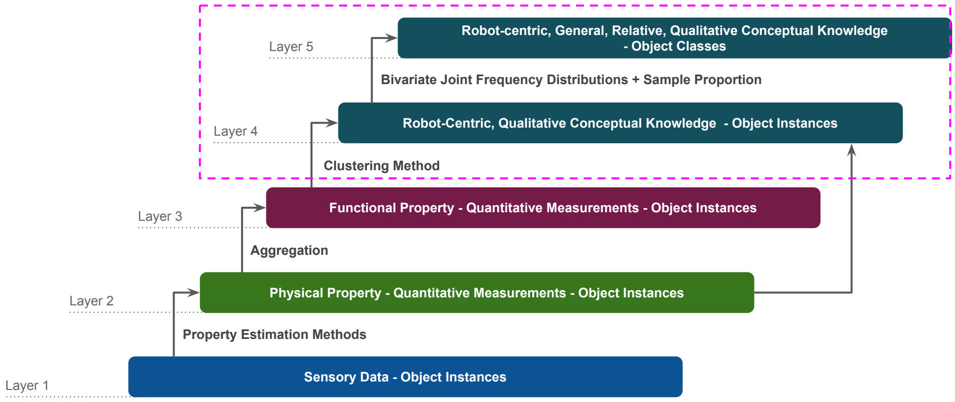


Figure 3.6: The figure illustrates the process layers, enclosed within a pink colored boundary, for our bottom-up robot-centric knowledge generation where each layer is abstracted over the layer below.

Let **P** and **F** be the given sets of physical properties' labels and a set of functional properties' labels respectively. By abuse of notation, each physical and functional property is identified with its label. For example,

$$\mathbf{P} = \{\, \mathit{flatness, hollowness, size, rigidity, roughness, weight} \,\}$$

where *flatness, hollowness,* etc. are the labels representing the respective physical properties. Similarly,

$$\mathbf{F} = \{\, \mathit{containment, support, movability, blockage} \,\}$$

where *containment, support,* etc. are the labels representing the respective functional properties.

For each physical property $P \in \mathbf{P}$ as well as for a functional property $F \in \mathbf{F}$, property measurement is estimated from each object instance $o \in \bigcup \mathbf{O}$. The measurements are estimated using the property estimation method discussed in chapter 2.

Let P_n and F_n represent sets of property measurements of a physical property P and a functional property F respectively, estimated from all the instances given in $\bigcup \mathbf{O}$. Note that n represents the total number of instances in $\bigcup \mathbf{O}$.

3.3.1 Sub-categorization – from Continuous to Discrete

As we have stated in Sec.3.2.1, the desired conceptual knowledge about objects is required to be qualitative. In other words, an object (instance or class) should be represented in terms of the *qualitative* measurements of the properties observed in the object as opposed to *quantitative* measurements of the properties (see the top two layers of conceptual knowledge generation in the Fig. 3.6). The *sub-categorization* process is the first step in creating conceptual knowledge about object classes. The process generates (more intuitive) qualitative measures, to represent the degree with which a property (physical or functional) is reflected in an object instance, unsupervisedly using a clustering mechanism (see Fig. 3.8) on the quantitative measurements of a property estimated in the object instances. In that, a cluster of the property measurements can be seen as a qualitative measure of the corresponding property.

In this process, P_n and F_n representing measurements of a physical property $P \in \mathbf{P}$ and a functional property $F \in \mathbf{F}$ respectively estimated from n number of object instances are categorized into a *given* number of discrete clusters η using a clustering algorithm. For the sake of clarity, henceforth a qualitative measure of a physical property is referred to as a physical quality and that of a functional property as a functional quality.

Let P_η and F_η be the sets of labels, expressing physical qualities and functional qualities, generated for a physical property $P \in \mathbf{P}$ and a functional property $F \in \mathbf{F}$ respectively. For example, in

$$size_\eta = \{\, small,\ medium,\ big,\ huge \,\}$$

size is a physical property and *small, medium, big, bigger* are its physical qualities. Similarly, in

$$support_\eta = \{\, no\_support,\ weak\_support,\ good\_support,\ strong\_support \,\}$$

support is a functional property and its functional qualities are *no_support, weak_support, good_support, strong_support.*

As illustrated in figure 3.7, in case we decide on three qualitative measures, the property measurements of a property, for instance, *rigidity*, estimated from all the objects instances will be clustered into three clusters and they can be represented as,

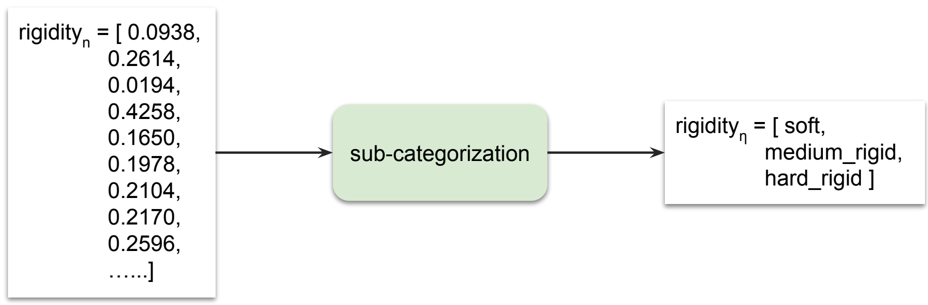


Figure 3.7: The figure illustrates the sub-categorization process wherein quantitative measurements of rigidity property are transformed into qualitative measurements using a clustering algorithm.

$$rigidity_{\eta}=\{\ soft,\ medium\_rigid,\ hard\_rigid\ \}$$

Note that, in this work, the number of qualitative measures remain the same for all the properties as opposed to different number of qualitative measures for different properties. The optimum number of qualitative measures for substitute selection are decided empirically as discussed in the experiment 4.3.1. Additionally, it is to be noted that the number of clusters essentially describes the granularity with which each property can qualitatively be represented. A higher number of clusters suggest that objects can be qualitatively described in a finer detail, which may obstruct the selection of a substitute since it may not be possible to find a substitute which is similar to a missing tool down to the finer details (see experiment 4.3.1 in the next chapter on substitute selection).

Note that, the aforementioned physical quality labels are only provided for illustration purpose as they are commonly used qualitative labels for the stated property; however, the quality labels for any property are internally represented by combining a property label P and a cluster label (created by the clustering algorithm). For example, the physical quality labels for *size* are expressed as,

$$\{\ size\_1,\ size\_2,\ size\_3,\ size\_4,\ ...\}$$

At the conclusion of the sub-categorization process, the clusters are mapped to the generated symbolic labels for qualitative measures. It is worth noting here that the qualitative measures generated by the process are *relative* to the given number of instances and their corresponding quantitative data. As the number of instances increases, with the addition of quantitative data, the sub-categorization process will have to be repeated. As a result, the quality labels assigned to a property measurement of an object instance will not remain permanent and will shift with respect to the repeated sub-categorization process.

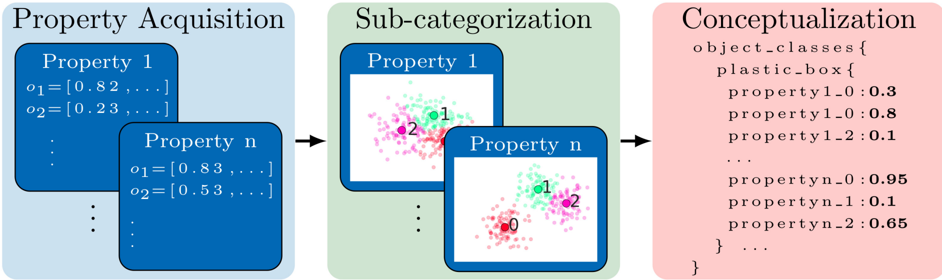


Figure 3.8: The robot-centric conceptual knowledge generation process is illustrated where acquired continuous property data of objects $\{o_1, o_2 ...\}$ is sub-categorized into multiple clusters. Using Bi-variate joint frequency distribution and sample proportions, conceptual knowledge about object classes (e.g. *plastic_box*) is generated.

3.3.2 Conceptualization – Knowledge about Objects

The conceptualization process is twofold: in the first step, it generates the knowledge about all the instances and in the second step, it generates knowledge about object classes on the basis of the knowledge about the instances. As illustrated in the top two levels of Fig. 3.6, the resulting conceptual knowledge about objects consist of two layers: knowledge about object instances and knowledge about object classes.

Knowledge about object instances:

Let $\mathbf{P}_\eta$ and $\mathbf{F}_\eta$ be the families (sets) of sets containing the physical quality labels P_η and the functional quality labels F_η for each physical property $P \in \mathbf{P}$ and functional property $F \in \mathbf{F}$ respectively. Note that, η is the number of clusters a.k.a. qualitative measures.

In order to generate knowledge about each object instance, we aggregate all the physical and functional quality labels assigned to physical and functional property estimation values of each object instance in the *sub-categorization* step. The knowledge about object instances is a compilation of knowledge of each object instance.

Thus, each object instance $o \in \bigcup \mathbf{O}$ is represented as a set of all the physical as well as functional qualities attributed to it, which are expressed by a symbol **":"** in our work. For the sake of clarity, we will substitute the symbol **":"** with *holds* and it is represented as:

$$holds \subset \bigcup \mathbf{O} \times \bigcup (\mathbf{P}_\eta \cup \mathbf{F}_\eta).$$

For example, knowledge about the instance *plate₁* of a *plate* class can be given as,

$$holds(plate_1, medium),$$
$$holds(plate_1, harder),$$
$$holds(plate_1, good\_support)$$

where *medium* is a physical quality of *size* property, *harder* is a physical quality of *rigidity* property and *good_support* is a functional quality of *support* property.

Figure 3.9 illustrates how the knowledge about object instances is expressed after aggregating the physical and functional quality labels assigned to it. The figure depicts the knowledge about instances of *ball, cup, metal box* and *book*. It consists of qualitative measures of the following physical properties: *weight, size, roughness, hollowness, flatness, rigidity* and functional properties: *support, movability, containment, blockage*. As shown in the figure, we have used the Python data type Dictionary to express the knowledge about the instances. At this stage, as illustrated in the figure 3.6, we have captured *robot-centric, relative* and *qualitative* aspects of our proposed conceptual knowledge. It should be noted that the relativeness of knowledge at this stage is constrained to the total number of instances.

Knowledge about object classes:

The second stage of the conceptualization process generates the knowledge about an object class by aggregating knowledge about its instances.

"ball_9": ["support_1", "movability_0", "weight_0", "size_1", "roughness_3", "hollowness_0", "containment_2", "flatness_1", "rigidity_1", "blockage_1"]	***"cup_1":*** ["support_1", "movability_2", "weight_2", "size_3", "roughness_2", "hollowness_1", "containment_0", "flatness_1", "rigidity_2", "blockage_3"]
"book_8": ["support_3", "movability_2", "weight_2", "size_2", "roughness_2", "hollowness_0", "containment_1", "flatness_2", "rigidity_0", "blockage_3"]	***"metal_box_10":*** ["support_0", "movability_0", "weight_0", "size_0", "roughness_2", "hollowness_0", "containment_1", "flatness_0", "rigidity_1", "blockage_1"]

Figure 3.9: An illustration of instance knowledge generated by the Conceptualization process. The illustration depicts the knowledge about four object instances: *ball_9, cup_1, metal_box_10, book_8* which contains physical and functional qualities observed in the instances. In the illustration the physical and functional property measurements are clustered into four qualitative measures.

The underlying principle is to represent an object class in terms of the overall observations made about its instances which encompasses the *frequently* observed properties as well as *less frequently* observed properties associated with it. In order to represent the frequency of qualitative measures of the physical or functional properties observed in an object class, we employed a statistical technique called *bivariate joint frequency distri-*

bution. A *frequency distribution* is typically concerned with how often different values of a single variable appear within a sample. The *bivariate joint frequency distribution* focuses on two variables and it describes how often a pair of values of two variables appear within a sample.

In order to apply *bivariate joint frequency distribution*, we first obtain knowledge about instances of an object class. Given that the knowledge about object instances consist of physical and functional qualities, for each physical and functional quality, we count how often it is assigned to the instances of the object class. Figure 3.10 illustrates the application of bivariate joint frequency distribution to the instances of an object class *plate*. We considered three physical qualities of the properties *roughness* and , and three functional qualities of the properties *support* and *movability*. Based on the distribution, we can see that the functional quality *support_1* and the physical qualities *_1* and *_2* have not been observed in any of the instances of *plate*. Note that, this circumstance may change in case of an instance assigned any or all of the aforementioned qualities. This demonstrates that the knowledge generated about any object class is relative to the observed instances and is subject to change when the status quo is confronted with new observations.

While the frequency offers the number of times a certain quality is observed in the encountered instances of an object class, it does not state the share of the quality within the encountered instances of the object class. In order to determine the share of the quality, *sample proportion* is calculated by using the following formula:

$$\textit{sample proportion of a physical or functional quality} = \frac{\textit{number of instances containing the physical or functional quality}}{\textit{total number of instances}}$$

In figure 3.11, on the left, we have knowledge about an object class *plate* which contains physical and functional qualities along with the corresponding frequency. Note that, the sum of the frequencies of physical or functional qualities of a corresponding property is same as the total number of instances. On the right side of the figure, the knowledge about the *plate* class is given in terms of the sample proportion of each physical or functional quality observed in the instances of the *plate* class. A row in the sample proportion table can be read as, a physical quality *roughness_0* is observed in the 50% of the instances of the object class *plate*. Note that,

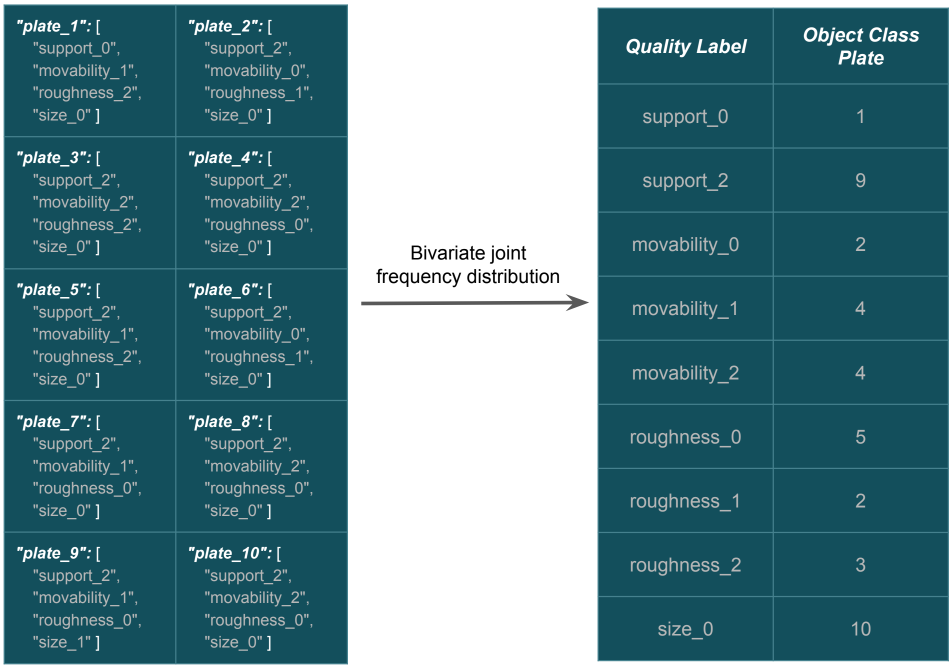


Figure 3.10: An illustration of class knowledge generated by the Conceptualization process by applying *bivariate joint frequency distribution*. The illustration depicts the knowledge about 10 instances of an object class *plate*. On the left is knowledge about individual instances of plate containing physical and functional qualities observed in each instance. On the right is knowledge about class *plate* which contain physical and functional qualities and their respective frequency. In the illustration the physical and functional property measurements are clustered into three qualitative measures.

similar to the frequency distribution, the knowledge represented in terms of the sample proportion is *relative* to the encountered instances and thus, it is subject to change. As a consequence, the sample proportion of any physical or functional quality will have to be updated as new instances of an object class are encountered. As stated in Fig. 3.6, the resulting knowledge about object class represents the *generalised* aspect of the proposed conceptual knowledge about objects. Fig. 3.8 illustrates graphically the resulting conceptual knowledge about an object class at the conclusion of the *Conceptualization* process where we have used the Python data type Dictionary to express the desired fuzzy set and attribute-value based formalism for our proposed knowledge base. The sample proportion is later exploited in the next chapter to identify a substitute for a missing tool from the available objects.

Quality Label	Object Class Plate
support_0	1
support_2	9
movability_0	2
movability_1	4
movability_2	4
roughness_0	5
roughness_1	2
roughness_2	3
size_0	10

Sample proportion →

Quality Label	Object Class Plate
support_0	1/10 = 0.1
support_2	0.9
movability_0	0.2
movability_1	0.4
movability_2	0.4
roughness_0	0.5
roughness_1	0.2
roughness_2	0.3
size_0	1.0

Figure 3.11: An illustration of class knowledge generated by the Conceptualization process by applying *sample proportion* to the *bivariate joint frequency distribution*. The illustration depicts, on the left, the knowledge about object class *plate* represented in terms of the *frequency distribution* of physical and functional qualities observed in 10 instances of *plate*. On the right is knowledge about the object class *plate* represented in terms of the sample proportion of each physical and functional qualities observed in the instances.

Function models:

In addition to conceptual knowledge about objects, the *conceptualization* process also creates knowledge about functional qualities, termed as *function models*, by associating the occurrence of physical qualities in an object instance given the occurrence of a functional quality in the instance and aggregating the result of such conditional occurrences. The role of a functional model is discussed later in the chapter 4, section 4.2. The idea behind the function model is to identify the physical qualities which are correlated with functional qualities. The functional model generation follows the similar two-step process as the conceptual knowledge generation: first, the *bivariate joint frequency distribution* is applied which is then followed by *sample proportion* calculation. Note that, in order to generate a function model for each functional quality, the knowledge about all instances are considered.

Figure 3.12 demonstrates how a preliminary functional model for a functional quality is generated using *bivariate joint frequency distribution* wherein the frequency of a physical quality observed in the object instances given that a functional quality is observed in those instances is counted. In the illustration, we concentrate on a functional quality *support_1* of a functional property *support*. On the left side of the figure, we have knowledge about ten object instances, each of which contain *support_1* and the physical qualities of three physical properties: *weight*, and *roughness*. The bottom two rows contain instances with the physical qualities of the aforementioned physical properties , however, different functional qualities than *support_1* of a functional property *support*. The right side of the figure indicates the preliminary functional model of *support_1* after applying *bivariate joint frequency distribution* to the knowledge about twelve object instances. The preliminary function model indicates the number of times each physical quality was observed in an instance given that *support_1* was also observed in the instance. Therefore, although *weight_0* is observed in ten instances, in only eight instances, it is observed along side *support_1*. A similar observation can be made for the physical qualities *_1* and *roughness_2*.

In the second step, similar to knowledge about an object class, a function model is generated by applying *sample proportion*. The *sample proportion* offers the proportion of each physical quality in the instances in which a functional quality is observed. As a result, it can be determined which physical quality was observed often in the instances where the functional quality is observed. Figure 3.13 illustrates a function model generated for a functional quality *support_1*. On the left side of the figure, we have a preliminary function model of *support_1* (see Fig. 3.12) which contains the frequency of the physical qualities of three physical properties *rigidity*, *weight* and *roughness* in the instances where *support_1* is also observed. On the right side of the figure, *sample proportion* is employed which calculates the proportion of each physical quality by dividing the frequency

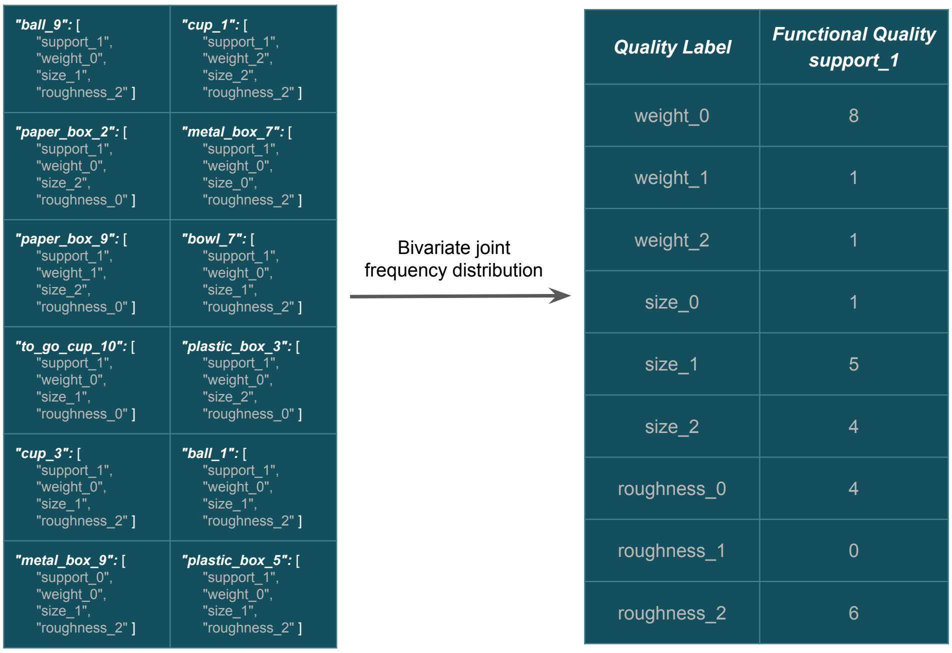


Figure 3.12: An illustration of a preliminary function model of a functional quality *support_1* generated by the Conceptualization process by applying *bivariate joint frequency distribution*. The illustration depicts the knowledge about 10 instances of different object classes. On the left is knowledge about individual instances containing physical qualities and the functional quality *support_1* observed in each instance. On the right is the function model of *support_1* which contain physical qualities observed when *support_1* is also observed and their respective frequency. In the illustration the physical property measurements are clustered into three qualitative measures.

of the physical quality by the total number of occurrences of *support_1* in the knowledge about instances. For instance, in the example, *support_1* is observed in total ten times out of which eight times it was observed along side the physical quality *weight_0*. One can infer that the presence of *support_1* in the instances is highly correlated with the presence of *weight_0* in the knowledge about object instances. A contrast observation can be made about *weight_1* and *weight_2* as their presence is less correlated with the presence of *support_1*. Such correlation among the physical and functional qualities is later exploited in the Chapter 4, Sec. 4.2.3 to identify relevant properties (see Sec. 4.1.2) of a missing tool. Note that, similar to conceptual knowledge about any object class, the correlation between physical and functional qualities is *relative* to the number of instances observed and is subject to change as the robot encounters more instances. Consequently, a function model of any functional quality will also be subjected to change as it is relative to the number of instances observed so far.

Quality Label	Functional Quality support_1
weight_0	8
weight_1	1
weight_2	1
size_0	1
size_1	5
size_2	4
roughness_0	4
roughness_1	0
roughness_2	6

Sample proportion →

Quality Label	Functional Quality support_1
weight_0	8/10 = 0.8
weight_1	0.1
weight_2	0.1
size_0	0.1
size_1	0.5
size_2	0.4
roughness_0	0.4
roughness_1	0
roughness_2	0.6

Figure 3.13: An illustration of a function model of a functional quality *support_1* generated by the Conceptualization process after applying *sample proportion* to the *bivariate joint frequency distribution*. The illustration depicts, on the left, the preliminary functional model of *support_1* represented in terms of the *frequency distribution* of physical qualities observed in 12 instances given that *support_1* is also observed in those instances. On the right is the final functional model of *support_1* represented in terms of the sample proportion of each physical observed in the instances where *support_1* is observed.

3.4 Literature Review

As stated in chapter 1, our proposed approach for substitute selection performs knowledge-driven computation to identify the relevant properties of a missing tool and determines the most suitable substitute on the basis of those properties. Since the computation requires an access to the conceptual knowledge about properties of a missing tool and of existing objects in the environment, we set out to explore the existing knowledge bases. The primary objective of this exploration was aimed at determining whether the knowledge about objects from the existing knowledge bases can be exploited in our substitute selection approach.

The demand for such conceptual knowledge about objects has been increasing over the years (see Fig. 3.14). Especially, for the developers of reasoning systems such as tool selection, task planning or an action selection aimed at service robots where they are expected to perform household tasks, an unhindered access to a stack of knowledge about objects or the environment is a primary concern. Since there are myriad knowledge bases developed for service robots, it can be cumbersome to scrutinize each one of them to examine its usefulness to the intended system. The objective of this literature review is to provide an overview of the existing knowledge bases and examine their characterization with respect to the proposed *robot-centric*, *general*, *relative* and *qualitative* conceptual knowledge.

3.4.1 Knowledge Base Selection

There has been an increasing interest in the knowledge-based systems aimed at various applications in robotics such as human-robot interaction [93], action recognition [94], task planning [95], robot navigation [96]. While there are myriad amount of knowledge bases designed for either specific application or for wider range of applications, it is a challenging task to identify the most suitable one for our specific demand [87]. After determining that there is no comparison of knowledge bases containing the relevant information for the robotic applications exist, we executed a systematic investigation of the state of the art into three phases to identify the relevant knowledge bases:

b) Literature Search: In order to find the relevant papers for this review article, we automatically aggregated publications from publication databases by referencing the following combinations of keywords : `knowledge engine robot`, `knowledge database robot`, `knowledge household objects`, `knowledge data household` and `knowledge base robot`. The crawler provided 313 papers after removing the duplicates.

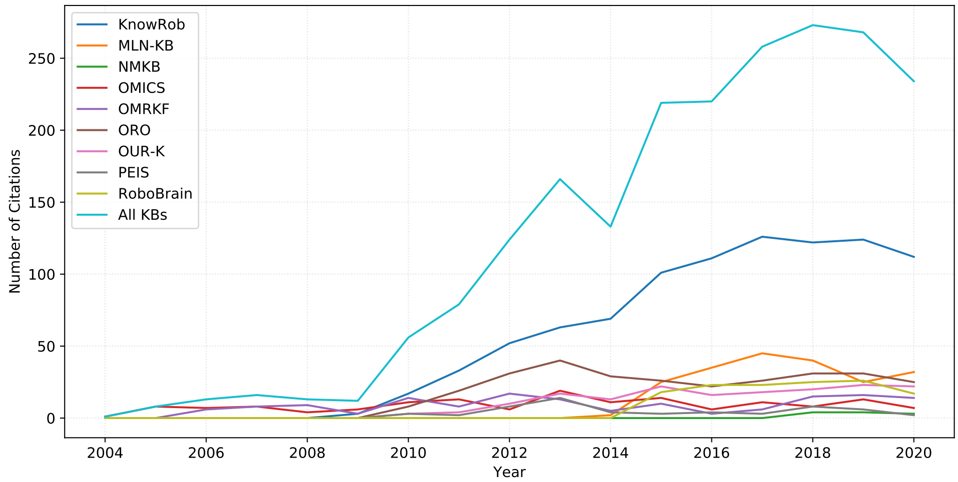


Figure 3.14: The plot illustrating the number of citations to knowledge base papers increasing

c) Literature Filtering: In this phase, the paper selection was manually evaluated and assessed. We removed the papers which:

- focused on the development of knowledge bases for non-robotic applications.
- were written from the application perspective, without a discussion of the underlying knowledge base.
- do not cover knowledge about household objects.
- focused primarily on knowledge acquisition without a framework in place to store the acquired knowledge or update the existing knowledge.

As a result, we selected 39 papers covering 9 knowledge bases for evaluation[1]. The involved knowledge bases are summarized in Table 3.1 along with their acronyms by which they are identified. The plot in Figure 3.15 illustrates the life span of each knowledge base where the knowledge bases that are still actively being researched, since their inception, are highlighted in green color. The knowledge bases which are not active anymore are highlighted by blue color. Their lifespan indicates the duration - from

[1] https://tinyurl.com/KBPaperList

Table 3.1: List of selected knowledge bases and their names

Knowledge Base	Name
Knowledge processing system for Robots	KnowRob [97]
Knowledge Base using Markov Logic Network	MLN-KB [98]
Non-Monotonic Knowledge-Base	NMKB [99]
Open Mind Indoor Common Sense	OMICS [100]
Ontology-based Multi-layered Robot Knowledge Framework	OMRKF [87]
OpenRobots Ontology	ORO [101]
Ontology-based Unified Robot Knowledge	OUR-K [102]
Physically Embedded Intelligent Systems	PEIS [103]
Knowledge Engine for Robots	RoboBrain [104]

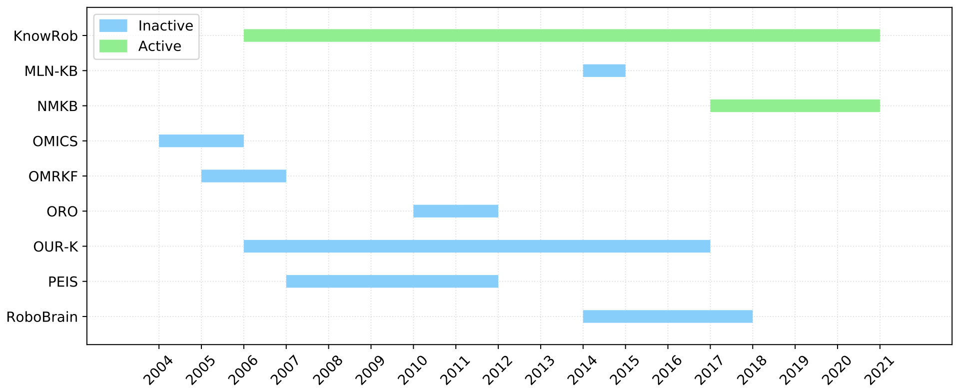


Figure 3.15: The plot illustrating the knowledge bases that are still actively researched according to the published work indicating the life span of each knowledge base.

the year of their inception until the year they were active. For instance, OMICS has a lifespan of two years, starting from year 2004 and was active until year 2006.

d) Final Literature Selection: In the last step we revised the final list and extracted the most important papers according to:

1. Content - we looked for papers providing detailed descriptions of configurations, content, performance, interfaces, etc. of the knowl-

Table 3.2: The selected knowledge bases, along with the pointers to the papers about and an overview about their impact (number of citations provided by google scholar as of 8th April 2021)

Knowledge Base	**Pointers**	**Citations**
KnowRob	[97], [105], [106], [107]	957
MLN-KB	[98]	216
NMKB	[99]	13
OMICS	[100], [108], [109]	160
OMRKF	[110], [111], [87]	149
ORO	[101], [112], [113]	296
OUR-K	[102]	172
PEIS	[103], [114]	59
RoboBrain	[104]	139

edge base. This information is necessary to assess the knowledge bases with respect to different criteria.

2. Impact - we examined the impact of each paper on the basis of the number of citations the selected papers have received and how those numbers have evolved over the years as illustrated in 'Impact of the paper' column of table 3.2. In terms of the number of citations, KnowRob is by far the most influential knowledge base since its inception while individual papers referencing OMICS and OMRKF are continuously cited. Figure 3.14 illustrates the citation history of selected papers of each knowledge base over the years. The citation history demonstrates how over the years the interest in knowledge bases has increased as its applicability in robots is increased.

For KnowRob, however, we have isolated 4 papers from over 40 papers. For the comprehensive list of the papers, please visit the web page of KnowRob[2]. As a result, the original list was filtered and eventually 20 research papers were selected covering the 9 knowledge bases (see Table 3.2).

[2] http://knowrob.org/publications

3.4.2 Knowledge Base Review

Typically for any knowledge base, three components play foundational roles: *knowledge acquisition*: how is the knowledge acquired, *knowledge representation*: how is the acquired knowledge represented and *knowledge processing*: how is the knowledge processed. Each component comes with their own set of challenges which need to be taken into account while designing and building a knowledge base. For knowledge base reviewing, we focus on similar components as they distinguish knowledge bases from each other. For each component, we have selected the criteria that are relevant to the objective of this doctoral work, given in the Table 3.3, along which the knowledge bases are examined. Note that, our review provides an aerial view of how each component is realized with respect to the criteria. The review is structured in a tabular form wherein the particulars of each criteria considered in each knowledge base are specified.

Table 3.3: The list of criteria corresponding to the components Knowledge Acquisition, Knowledge Representation and Knowledge Processing used to review the knowledge bases

Component	**Criteria**
Knowledge Acquisition	Knowledge Source Knowledge Content
Knowledge Representation	Representation Formalism Modeling of Uncertainty Symbol Grounding
Knowledge Processing	Inference or Query Mechanism

Knowledge Acquisition

Knowledge Source

Knowledge acquisition is primarily concerned with acquiring the desired knowledge that populates any knowledge base. Evidently, a single source or multiple sources are needed from which the knowledge can be acquired. As a result, one of the primary tasks in knowledge acquisition is to identify such sources. In case of knowledge bases to be used by robots, the sources could be, for instance, the sensors deployed on a robot. However, despite of the sensors being deployed, the real-world is perceived differently by a robot than its human counterpart due to the limited perception capabilities of the robot [87]. Therefore, it is not unusual to provide hand-coded knowledge or use some existing human-made common sense knowledge bases such as WordNet, ConceptNet etc. As the knowledge can be acquired from multiple sources, therefore, when reviewing the 'knowledge acquisition' of the knowledge bases, we specify what sources were used to acquire knowledge by each knowledge base. In the Table 3.4, we have listed down the sources used by each knowledge base to acquire the desired knowledge.

Knowledge Contents

The other aspect of knowledge acquisition is concerned with the contents of knowledge in the knowledge base. In order to select a knowledge base for a certain application, it is important to know what kind of knowledge is contained in it. In the case of the proposed substitute selection application, for instance, we are interested in the conceptual knowledge about objects which consist of generalized, robot-centric, relative and qualitative knowledge about physical and functional properties of objects. As a result, in reviewing the knowledge bases, as we are primarily interested in the knowledge about objects, we focused on what kind of knowledge about objects is acquired by each knowledge base. During our review, we noted that various aspects related to objects were targeted by the knowledge bases. We have listed them down in Table 3.4 under Contents. Additionally, we have also distinguished between what knowledge was acquired from the sensors or from the robotic perspective (robot-centric) and what knowledge was acquired from non-sensory sources such as web pages, manually encoded. Table 3.4 catalogues each knowledge base, the kind of knowledge about objects it contains and the source/s from which, what kind of knowledge is acquired.

Knowledge Representation

Representation formalisms:

Knowledge representation is concerned with encoding of the acquired knowledge using a certain representation formalism. It is another vital component that separates knowledge bases from each other as there are variety of representation formalisms available. The knowledge encoded

Table 3.4: Comparison of the selected knowledge bases with respect to knowledge acquisition: what is the source of knowledge and what kind of knowledge was acquired using the source.

Knowledge Base	Contents: Object	Appearance	Properties of Objects	Spatial Relations	Temporal Relations	Uses of Objects	Topological Relations	Map of the Environment	Actions	How to Perform Task	Other	Source of knowledge
KNOWROB	●	●		●	●		●	●	●			Multi-Modal Sensor Systems
	●	●	●	●		●	●				A	OpenCyc, WordNet, OMICS
	●	●										Online Shops
										●		Observation of Human Activities or Shared by Other Robots
										●		Web Instructions
MLN-KB	●	●										ImageNet
	●	●	●									Freebase, Amazon, Ebay
	●		●									WordNet
	●					●						Manually Encoded
											B	Standford 40 Action Dataset
NMKB	●	●	●	●		●	●		●	●		An Interaction-Oriented Cognitive Architecture [115]
OMICS	●	●	●	●	●	●	●	●	●	●	A	non-expert users, WordNet
OMRKF	●	●	●	●			●	●				Multi-modal Sensors
	●		●	●	●	●	●					Manually Hand-coded
ORO	●	●	●	●					●			OpenCyc
	●	●	●	●			●					Multi-modal sensor system
	●	●	●	●			●					Human Interaction
OUR-K	●	●	●	●			●	●				Multi-modal sensor system
	●		●	●	●	●	●		●			Manually Hand-coded
PEIS	●		●	●		●	●					Cyc
	●		●	●		●	●					Vision and Localization System
RoboBrain	●		●			●			●	●		Robot Interaction
	●		●								A	WordNet, OpenCyc, Freebase
	●		●									ImageNet

A = Common Sense Knowledge about the objects and the environment
B = Human-poses and human-object relative position during object manipulation

Table 3.5: Comparison of the selected knowledge bases with respect to Representation Formalism

Knowledge Base	**Formalism**
KNOWROB	OWL-RDF
MLN-KB	Markov logic network
NMKB	Prolog - Horn Clause
OMICS	Relational Database
OMRKF	OWL-RDF
ORO	OWL-RDF
OUR-K	OWL-RDF
PEIS	Second Order Predicate Logic
RoboBrain	Graph Database

in such formalisms allows for a meaningful internal representation of the external world [88]. Given the complexity of the world around us, it is a daunting task even for humans to select what aspects of the world should be represented and what aspects of the world should be ignored. The representation formalisms facilitate in such a selection as each formalism is specialized in formalizing a very specific aspects of the world while ignoring others [88; 89]. Additionally, representation formalisms offer an efficient system to extract and reason about the knowledge [88]. The Table 3.5 lists down the formalisms used by each knowledge base to represent the knowledge stated in the Table 3.4.

Symbol Grounding:
The knowledge representation formalisms given in Table 3.5 are primarily symbolic where the symbols are typically based on a natural language [116] whereas a robot typically perceives the environment with the help of its sensors which is usually represented in a non-symbolic or numerical form. The knowledge represented in a symbolic formalism is abstract in nature in a sense that knowledge provided to a robot in this form may not have any bearing on the robot's perception of the environment. For a robot to use the knowledge effectively, it needs to know the meaning behind the symbols used to represent the knowledge and how do they correspond with the robot's perception of the environment. For instance, when a robot is given a knowledge that *cup is in the kitchen*, it needs to know

what is a cup and *what is a kitchen.* This is where the notion of *symbol grounding* comes into play where the symbolic representation of a cup is grounded into the robot's sensory representation of the cup. In other words: a symbol *cup* is mapped to the sensory model of it generated by, for instance, an object recognition system. The symbol grounding, thus, is an important aspect of knowledge representation and it is also closely connected with a robot-centric aspect proposed in this work. It is, therefore, no surprise that the knowledge bases reviewed in this work also address the issue of symbol grounding and offer approaches for this. It is, however, worth noting that due to the robot's limited perception capabilities, it is not possible to ground all the symbolic knowledge. The Table 3.6 summarizes what knowledge is grounded in the robot's sensory representations in each knowledge base.

Table 3.6: Comparison of the selected knowledge bases with respect to Symbol Grounding.

	Grounded Knowledge								
Knowledge Base	Object	Properties of Objects	Spatial Relations	Location	Affordances	Topological Relations	Actions	Task	Other
KNOWROB	●	●	●	●	●	●	●	●	
MLN-KB	●	●	●	●	●		●		A
NMKB	●		●				●		
OMICS									B
OMRKF	●	●	●			●			B
ORO	●	●	●			●			
OUR-K	●	●	●			●			
PEIS	●	●	●		●	●			
RoboBrain	●	●							

A = Weights of the objects
B = Knowledge is not grounded

Uncertainty

Understanding the environment or objects in it from a robot perspective has its own share of difficulties: one such difficulty being uncertainty present in the robot's perception of the world [116]. Uncertainty can be due to the noisy data or partial observability, limited sensor capabilities causing incomplete understanding of the world, to list a few. For instance, a 2-D sensor model of a cup is a partial representation of a cup in the real world. The uncertainty can also be present in the symbolic knowledge, for instance, *a cup is usually on the shelf*. Here the term *usually* represents the uncertainty that refers to the likelihood. Whether the knowledge is grounded into uncertain sensory representation of the world or the knowledge itself contains the uncertainty, a representation formalism must be able to capture this uncertainty in the representation. Table 3.7 summarizes how each knowledge base has handled uncertainty. As we noted earlier that uncertainty can take various forms, in the tabular summary we also noted what kind of uncertainty is captured along with the mechanism used to represent the uncertainty. Note that, in our work, uncertainty is referred by qualitativeness and relativeness to represent the general, robot centric knowledge about object classes.

Knowledge Processing

Knowledge processing is another vital component which distinguishes knowledge bases from each other. The knowledge processing is concerned with the question of usability of a knowledge base. For any knowledge base to be useful, it should be equipped with a knowledge processing system that contains mechanisms to access the knowledge, retrieve the knowledge and reason about knowledge [106]. Similar to databases, a knowledge base should be equipped with a query mechanism that retrieves the desired knowledge from the knowledge base. On the other hand, in order to perform reasoning, the knowledge base should also be equipped with an inference system which combines different pieces of knowledge and infers new knowledge. In Table 3.8, we have provided a tabular summary of what kind of query mechanisms and inference mechanisms are used in each knowledge base.

Table 3.7: Comparison of the selected knowledge bases and the modeling of the uncertainty with a focus on what knowledge is modeled and what mechanism is used to represent the uncertainty

Knowledge Base	Mechanism	Knowledge Content
KNOWROB	Probabilistic Model Statistical Relational Models	Noisy sensor information Relations between objects, types of objects
MLN-KB	Median-based	Noise in the web data
NMKB	Principle of Specificity	Incomplete Knowledge
OMICS	-	Uncertainty not considered
OMRKF	-	Uncertainty not separately modeled
ORO	Validation by Users	Unknown objects and its properties
OUR-K	Bayesian Inference	Unknown objects, action selection, context recognition
PEIS	Validation by Users	Disambiguate multiple groundings of a symbol
RoboBrain	Validation by Users	Inconsistencies due to knowledge coming from different resources, Disambiguate due to the same word having different meaning

Knowledge Base Size

So far, we have discussed the characteristics of the knowledge bases with respect to knowledge acquisition, knowledge representation and knowledge processing. For a knowledge base to be useful, size of the knowledge base is a critical piece of information. The size of the knowledge base can be measured in terms of quantities in which different kind of knowledge is available, for instance, number of objects, properties, relations etc. In Table 3.9, we have provided the information on the size of each knowledge base as reported in the respective literature.

Table 3.8: Comparison of the selected knowledge bases with respect to Inference/-Query Mechanism: what kind of knowledge is inferred and what mechanism is used.

	Contents										
Knowledge Base	Object	Properties of Objects	Spatial Relations	Localization	Affordances	Topological Relations	Context	Actions	Task	Other	Mechanism
KNOWROB	●	●	●	●	●	●	●	●	●	A	Prolog Query
	●	●	●	●		●		●	●	A	Probabilistic Inference
MLN-KB	●				●						ImageNet
NMKB										B	Prolog Query and Logic Inference
OMICS	●	●			●	●	●	●	●	A	SQL query
OMRKF	●	●	●			●	●	●			Logical Inference
ORO	●	●	●	●							Pellet
OUR-K	●	●	●			●	●	●			Bayesian Inference
PEIS			●	●	●	●	●	●	●		OWL Query
RoboBrain										A	RoboBrain Query Library

A = Retrieve knowledge from the knowledge base
B = Conceptual Inferences

Table 3.9: This table comprises the information about the size of knowledge bases reviewed in this paper. The size of knowledge bases is mainly quantified based on the number of objects, number of classes, instances etc.

Knowledge Base	**Quantification of size of KB**
KnowRob	Around 8000 classes that describe events, actions, objects, mathematical concepts and so on
MLN-KB	40 objects comprise 100 images and on average 4.25 affordance for each objects
NMKB	Not available
OMICS	As of 2004, 400 users with 26,000 accepted submissions, 400 images of indoor objects (current number of images unknown) comprising a total of 100,000 entries in the form of objects, actions, senses.
OMRKF	Knowledge about approximately 300 objects as per 2005
ORO	56 object classes and 60 predicates that states relation with objects
OUR-K	Knowledge about approximately 300 objects as per 2005
PEIS	15 objects that comprise 2 to 5 images for each object
RoboBrain	44,347 concepts and 98,465 relations

Knowledge Base Accessibility

The accessibility is concerned with the manner in which a knowledge base can be accessed. The knowledge bases should be developed such that they can be used by the developers around the world in various applications.The knowledge base accessibility criteria examines the ways in which each knowledge base is made accessible to the developers.In the accessibility, we have examined, if the knowledge bases are available to download or install, if there are tutorials or any other documentation available to get the user started and if there is information on API available. Additionally, we also check what kind of licensing is made available. Table 3.10 summarizes the accessibility of each knowledge base. Since for OMICS, OMRKF, OUR-K and PEIS, we were not able to find the required information, we have indicated NA (Not Applicable) in the table. Additionally, we have provided the available web pages for the knowledge bases in appendix C.

3.5 Conclusion

Since the demand for conceptual knowledge has been increasing in robotic applications, the development of knowledge bases has been undertaken by many researchers around the world [8; 117]. While there exist a multitude of knowledge bases, the question is how many existing knowledge bases about objects conform to the requirements which can be used for a substitute selection purpose: conceptual knowledge base containing knowledge about the objects' properties that is *general, relative, subjective (robot-centric), and qualitative*. In this chapter, we reviewed existing knowledge bases primarily containing knowledge about household objects and their underlying acquisition system. For the literature review, we selected 20 papers covering 9 knowledge bases about household objects on the basis of the contents of the paper with respect to the above mentioned requirements and overall impact of the paper on the basis of the number of citations (refer Table 3.1). Our review resulted in the following conclusions with respect to each building block discussed in the Sec. 3.2:

Conceptual Knowledge:
As our desired conceptual knowledge about an object class consists of

Table 3.10: Compendium of Knowledge bases accessibility features.

Knowledge Base	**Download?**	**Install?**	**License**	**Documentation**	**API**
KnowRob	yes	yes	Apache License	yes	yes
MLN-KB	yes	no	Open source	no	yes
NMKB	yes	yes	Golem Group License	yes	yes
OMICS	NA	NA	NA	NA	NA
OMRKF	NA	NA	NA	NA	NA
ORO	yes	yes	GNU General Public License	yes	yes
OUR-K	NA	NA	NA	NA	NA
PEIS	NA	NA	NA	NA	NA
RoboBrain	yes	yes	Creative Commons license	Yes	yes

qualitative knowledge about its physical and functional properties, we reviewed the existing knowledge bases to examine whether such conceptual knowledge was considered. In the Sec. 3.4.2, we reviewed the contents of the knowledge in each knowledge base. With regard to the knowledge source, we noted that the majority of the knowledge bases relied on the external human-centric commonsense (universal) knowledge bases such as ConceptNet [118], WordNet [119] (KnowRob, MLN-KB, OMICS, RoboBrain), Cyc [120] (PEIS-KB), OpenCyc [120] (KnowRob, ORO, RoboBrain) and the rest either relied on the hand-coded knowledge (OMRKF, OUR-K) or on knowledge acquired by human-robot interaction (NMKB), for the symbolic conceptual knowledge about objects. Our review concluded that while the existing knowledge bases do contain *general* common sense knowledge about objects, they do not contain *qualitative* knowledge about their properties as discussed in the Sec. 3.2.1. For instance, a *cup* is described in WordNet as *a small open container usually used for drinking; usually has a handle*. The description does not contain qualitative knowledge about various properties such as size, shape, weight, roughness or rigidity observed in a cup. Moreover, it is worth noting that the knowledge about objects in the existing knowledge bases is universal in nature and thus lacks subjectivity (robot-centricness), though some portion of the knowledge is grounded, and relativity aspects of knowledge. While common sense knowledge has its merits, in Sec. 4.3.3, we will discuss an experiment which illustrates the inadequacy of using WordNet in substitute selection.

Knowledge Representation:
Logic based representation formalisms were overwhelmingly used by a majority of the knowledge bases to represent knowledge: OWL-RDF (KnowRob, OMRKF, ORO, OUR-K), Markov Logic Network (MLN-KB), Prolog - Horn Clause (NMKB), Second Order Predicate Logic (PEIS), while database inspired formalisms were used by RoboBrain (Graph Database) and OMICS (Relational Database). Besides representing knowledge about objects, the knowledge bases also focus on representing various uncertainty factors such as noisy sensor information, incomplete knowledge, unknown objects or environment, and inconsistent knowledge. While all the above uncertainty factors are significant, the desired factors such as

relativity, and qualitative measures were not formalized while representing knowledge about object properties. For instance, when we think of a cup, although at the abstract level, it is a type of container, the degree of containment is different in a cup for espresso coffee and a cup for tea. Such variation in the containment is not reflected in the representations in the knowledge bases.

Robot-Centric:
Almost all of the knowledge bases (except for OMICS) addressed the problem of symbol grounding. While the object labels, appearance related properties (shape, size, etc.), and functional properties (KnowRob, MLN-KB, NMKB, PEIS) were grounded in the robot's perception, the reliance on human-centric symbolic knowledge did pose a disadvantage. Since the commonsense knowledge bases such as WordNet, ConceptNet are fully human-made, the depth and breadth of the knowledge is not perceivable by a robot due to its limited perception and manipulation capabilities. While a small portion of human-centric knowledge is grounded into robot's limited perception, the majority of the knowledge base remains non-grounded.

It should be noted that the knowledge bases existed independent of the sensory perception. The symbol grounding processes were introduced in the knowledge bases to correspond the sensory perception with the relevant symbolic knowledge. In contrast, our proposed approach generates knowledge from the quantitative measurements computed from the sensory data. As a consequence, the knowledge generated from the sensory data for a robot A may differ from the knowledge generated from the sensory data for robot B as the sensors. This is due to the different sensory capabilities of both the robots, thus reflecting the notion of robot-centricity: object understanding from a first-person perspective.

4

Substitute Selection

4.1 Introduction

Humans are sophisticated in their use of tools compared to their animal counterparts [32]. The sophistication pertaining to tool-use in humans involves not just the cognitive capabilities and dexterity in manipulating a tool, but also the diversity in tool exploitation [32; 121]. This ability to exploit the tools has enabled humans to adapt and thus exert control over an uncertain environment, especially when they are faced with unfavorable situations.

Given its role in our daily life, scientists across various disciplines are making consorted efforts to develop theories on tool use by conjecturing about various aspects of tool use. Baber in [1], for instance, proposed six forms of engagements with a tool to describe tool use and developed a theory of tool use on the basis of those engagements. He identified the six engagements as: *environmental, morphological, motor, perceptual, cognitive and cultural.* On the other hand, Vaesen has proposed in [32], nine cognitive capabilities which are essential for tool use: *enhanced hand–eye coordination, body schema plasticity, causal reasoning, function representation, executive control, social learning, teaching, social intelligence, and language.* So far, there has not been a consensus on a theory of tool use, however there is a wider agreement on tool use being a multidisciplinary, complex endeavor requiring integration of multitude of cognitive faculties and behavioural capabilities.

Tool substitution is a form of tool use where a substitute is selected in place of an intended tool such that the substitute can be used as the tool [1]. Like tool use, tool substitution is an elaborate endeavor which involves behavioral and cognitive aspects of problem solving [1]. For example, the primary function of a heeled shoe is to cover a foot while extending the height of a person wearing it, yet the heel can also be used to hammer a nail into a wall. One can observe the cognitive reasoning involving analogous thinking between the shoe and the hammer while taking into account the biomechanics of manipulating a shoe like a hammer. Suffice it to say that tool substitution occurs primarily in two stages: substitute selection and manipulation of a substitute as a tool. Our focus in this research is on the former. More specifically, the primary question that we address is, how to select object/s from the available objects that are suitable as a substitute for a tool.

4.1.1 Tool vs. Substitute

Before we delve into tool substitution further, it is necessary to clarify the meaning of a tool considered in this work and how it is distinguished from a substitute.

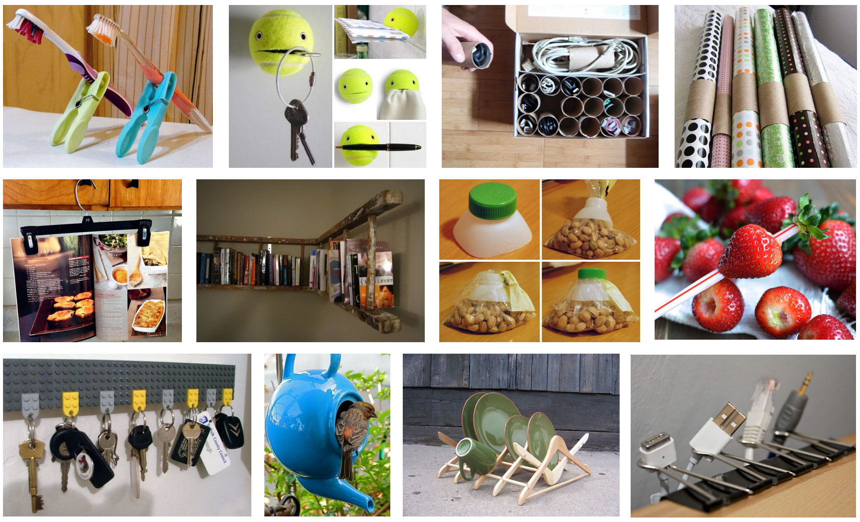
One may wonder what is perceived as a tool. So far the researchers have not agreed over a single definition of a tool and as a consequence one will find multiple definitions of a tool in the literature. For example, Butler in [5] suggested that "Nothing is tool unless during actual use". On the other hand, in [78] a tool is described as "any manipulable, physical implement that amplifies the user's sensorimotor capabilities." According to Baber in [1], a tool is "a physical object that is manipulated by users in such a manner as to both affect change in some aspect of the environment and also to represent an extension of the users themselves." Note that tools can occur naturally such as stones, tree branches etc. or they can be manufactured by altering a physical structure of an object or by combining multiple objects in a certain manner. It has been suggested that man-made tools are manufactured with a specific purpose in mind, and as a result such tools are primarily known for their conventional uses [1; 78]. For instance, a hammer is conventionally known for hammering or a tray is known for carrying food or drinks etc.

Given that there are multiple definitions of a tool reported in the literature on tool use in animals and humans, it is pertinent to define what is constituted as a tool as opposed to a substitute in the proposed approach. The definition of a tool considered in this work is inspired by aforementioned definitions: *a tool is foremost a physical object and is manufactured artificially for a designated purpose.* For instance, the designated purpose of a hammer is to hammer or of a tray is to carry food or drinks on it etc. *In contrast, a substitute is either a tool which is used for an unconventional purpose for which a conventional tool exists or it is a naturally occurring object.* The examples of substitutes, where a substitute is originally a tool, are: a shoe is used as a hammer, a plate is used for carrying drinks, a water bottle used as a vase for flowers etc. The examples of substitutes where a substitute is a naturally occurring object are: a stone is used as a paperweight or a banana leaf is used as an eating plate.

Note that a substitute for a tool can take many forms. Humans have proven time and again their creative ability to exploit a tool's form, shape or any other part of it to transform it into a desired substitute: sometimes using a tool as it is or in some cases by altering its physical structure or combining it with other tools. In this work, we do not consider substitutes that require any alterations in its physical form or combining it with other tools. Fig. 4.1 outlines the illustrations of substitutes that are considered in this work as opposed to the ones that are not considered in this work. Fig. 4.1(a) depicts substitutes-in-use scenario that we have envisioned for our work where the primary assumption is a conventional tool required in a task exists, however it is unavailable. In each scenario listed in the figure, a conventional tool for each task exists. In Fig. 4.1(a), starting from the left image in top row, the conventional tools are dip bowls, a match box stick, a hammer, hair sticks, chopsticks, a stool, a door stopper, a tray, a bottle opener, a vase, a camera pouch, and a plant pot. A substitute depicted in each task is not altered or combined in any way when being used in the task. On the other hand, figure 4.1(b) illustrates substitution scenarios which we have not considered in this work. The figure shows, in some cases, that a tool is being used creatively for which a conventional tool does not necessarily exist. For instance, an empty toilet roll is used to organize cables to hold gift wrapping papers to keep them from unraveling; a hanger is used to hold a recipe book, binder clips for holding cables and prevent them from falling off the table, a straw to remove stems from the strawberries, etc. In some examples, a tool is altered or combined with other tools before being used as a substitute: a tennis ball as key holder, a plastic bottle to pack nuts, hangers as dish dryer, ladder as book shelf etc.

The previous discussion and examples narrate the distinction between a tool and a substitute that we have made in our work. This distinction is essential when we determine suitability between them. Let us consider, for instance, two objects A and B. As we noted earlier, if an object A can be replaced by an object B then B is seen as a substitute for a tool A. However, the vice-versa need not be true, that is, a tool B can not necessarily be replaced by an object A. Within the context of a designated purpose, the substitutability relationship between a tool and a substitute is symmetric, for instance, *for hammering*, a hammer can be replaced by a heeled shoe and vice versa. However, it is not the case once you step outside the context, for instance, a hammer can not be used as a heeled shoe to cover foot. This indicates that the substitutability between two objects depends on which object is a tool and which is a substitute. As a result, the object B being a substitute of the object A does not make the object A a substitute of the object B by default. This aspect is examined further in the experiment discussed in the Sec. 4.3.3 to signify the distinction. In our work, this distinction is reflected in the notion of relevant properties of a tool and their relation to the properties of a substitute (see Sec. 4.1.2 and Sec. 4.2 for further details). We are proposing in this doctoral work, that relevant properties play vital role in determining an object's *suitability as a substitute* for a missing tool.

(a) Positive Examples of Substitutes



(b) Negative Examples of Substitutes

Figure 4.1: The examples of substitutes that are considered in this work are listed in the Fig. 4.1(a). The Fig. 4.1(b) illustrates the examples of substitutes that are not considered in this work [1]

4.1.2 Relevant Properties

Consider a scenario in which we have to select between a plate and a mouse pad as a substitute for a tray. Note that the tray is to be used for its designated purpose, that is, to carry drinks. In the given scenario, we will most likely select the plate as the substitute for the tray instead of the mouse pad. The question is what compelled us to reject the mouse pad and what made the plate the *suitable* substitute for the tray. We have stated earlier that deliberation for a tool selection in humans or animals alike is facilitated by conceptual knowledge about objects [1]. The conceptual knowledge, according to [3], about a tray can be defined as rigid, rectangular, flat top surface, wooden, brown colored, light weight, ability to support, movable while a plate can be defined as a rigid, circular, semi-flat top surface, white colored, light weight, ability to support, movable and a mouse pad as soft, rectangular, flat top surface, leather-based, light weight, ability to support and movable.

Baber has stated in [44] that a set of certain physical properties enable a specific purpose in a tool. A similar observation has been made in [36] which further states that the conceptual knowledge about objects is complemented by causal relationship that exist between physical properties of a tool and its functional properties. The question is how to determine which of the properties, physical and functional, of a tray enable its designated purpose.

The relevant properties of a tool in this work are considered as those properties which enable a tool's designated purpose. For instance, in order to enable a tray's designated purpose, the tray should be capable of supporting the predestined-for-the-tray objects that are placed on it. If the tray is not capable to support the objects that are placed on it, it can not be used for its designated purpose. As a result, *support* is considered as a relevant functional property of a tray. In order to enable support functional property in a tray, the physical properties rigidity and flat top surface are required [39; 78] which makes them the relevant physical properties of a tray. Note that if rigidity or flat top surface are absent in a tray, it will not be possible to support objects placed on the tray and consequently the

[1] credit: boredpanda.com/creative-life-hacks-diy, homesthetics.net, herzindagi.com, Pinterest - The Telegraph, cuinsight.com, flickr.com(Brett Patterson)

tray will not be able to carry objects on it. However, if a color of the tray is changed or material is changed, it will still be able to carry the objects on it as long as the relevant physical properties rigidity and flat top surface are present. The relevant properties of a tool, therefore, when absent in the tool, can not enable the tool's designated purpose. But then, is it enough for relevant properties of a tool to be present in a substitute?

Consider another functional property of the tray: movability. We know that the heavier any tool is, harder is its movability. Hence, if the tray is too heavy, a user may not be able to carry it. This makes movability and heaviness relevant functional and physical properties of the tray respectively. Consider instead of the plate and the mouse pad, we have a marble slab, a stool and the plate as possible substitute choices for a tray. Interestingly, all three choices are rigid and have flat top surface, thus allowing them to support objects that are placed on them and they are all movable too. It means, that all three are the substitutes for the tray, however, that may not be the case. From the given three options, the most likely substitute would be the plate. Baber in [2] said that, *A characteristic of a well-designed tool is that it feels comfortable and balanced when held.* As a result, when a tray is designed for its designated purpose, each relevant physical property needs to be present to a certain degree such that a user can use the tray comfortably. In other words, it is not arbitrary that heaviness is present in a tray to a certain degree. This notion also applies when deliberating on a substitute for the tray. In the above example, the plate is as heavy as the tray which makes it as movable as the tray and as a result, makes it to be used as comfortably as the tray. *Consequently, the notion of relevant properties of a tool consist of two aspects: a physical or functional property and a degree with which it is present in a tool.* This degree is manifested in our work as qualitative measures of physical properties as they are discussed in the Sec. 3.3.

4.1.3 Tool Substitution - Workflow

In this section, we discuss a workflow for a tool substitution system we have conceived in this work. The workflow is drawn from the insights we gained from the literature on tool use in animals and humans [3; 4; 22; 23; 24; 25; 32; 33; 35; 36; 77; 78; 121]. The objective of this section is twofold: 1) to portray the complexity of a computational model for the tool substitution system; 2) to place the focus of our research in the workflow. Fig. 4.2 outlines the primary processes, the required inputs and the desired outputs for each process. It is important to note that the presented workflow is one possible architecture of a tool substitution system.

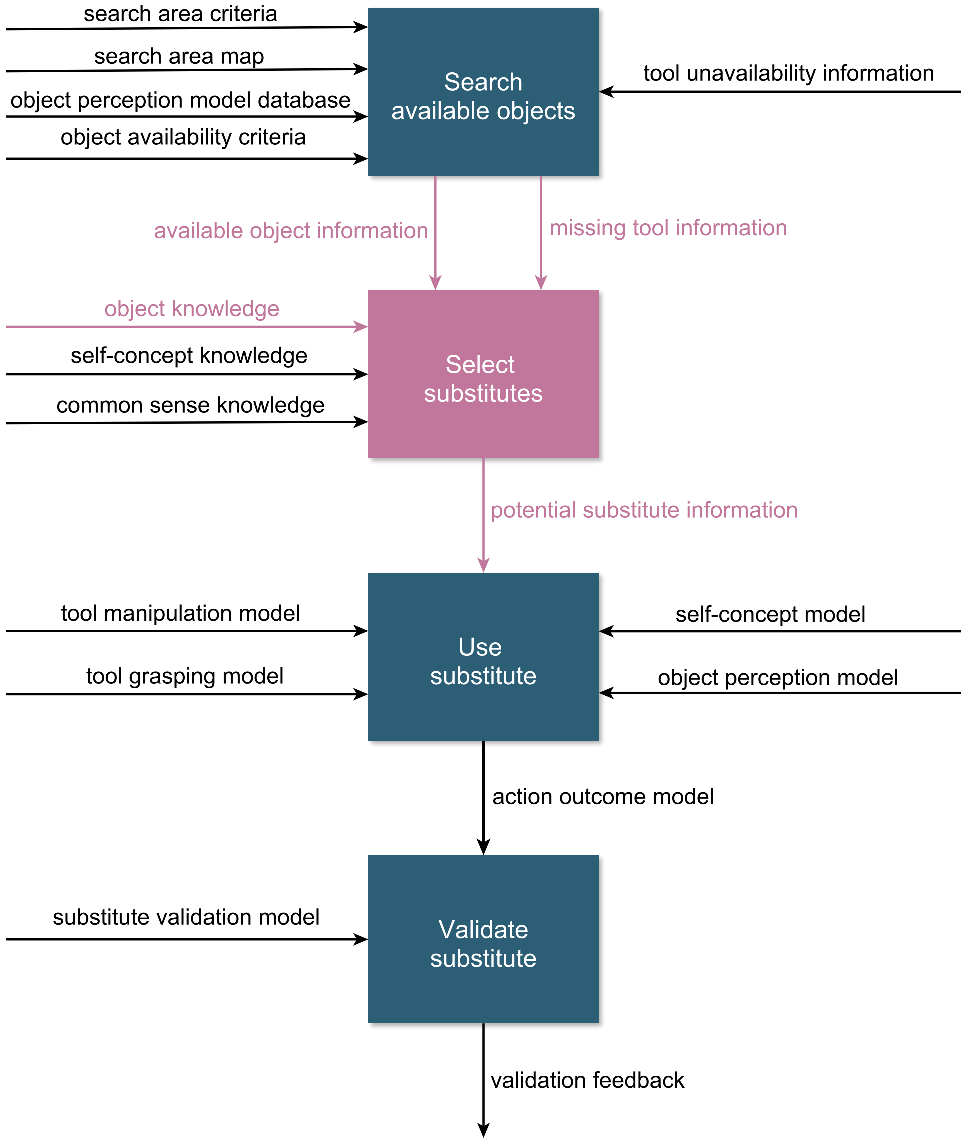


Figure 4.2: Tool substitution workflow envisioned in this work which consists of typical processes involved to perform tool substitution, typical primary inputs, primary outputs and supplementary inputs required for each process

Search available objects

Our proposed workflow is initiated when a robot requests for a substitute when it does not find a conventional tool required in an ongoing task. The workflow begins with a search process wherein available objects in the environment are discovered. The *tool unavailability information* can provide information about the missing tool such as tool label, perception model of the tool, the expected location of the tool etc. The *search area criteria* defines the search parameters for discovering the objects that are present in the environment. In any household environment, there are myriad number of objects which are spread across various locations. The *search area criteria* defines the region of interest for performing the search in the environment. The criteria can also include search duration, minimum and maximum number of objects to be discovered, their proximity to each other etc. The criteria in the end facilitate a more goal-directed and restricted search for the available objects. The *search area map* provides a map to reach a location of interest and to perform the desired search. The map can be refined further by adding information received from the *search area criteria* and *tool unavailability information*. The *object perception model database* consists of the learned perception models of household objects which can be used by the search process to detect and recognize objects. The *object availability criteria* contain the definition of availability of an object which facilitates the selection of available objects. An unavailability can occur as a result of an object being in use or being damaged, for example. In such cases, though the object is present in the environment, it can not be considered for the substitute selection purpose. Therefore it is necessary to define what it means for an object to be available. The search process takes all the above inputs into consideration and forms a search strategy to detect and select the available objects. Along with the *tool unavailability information*, the search process sends the *available object information* for determining possible substitutes among them. The *available object information* can contain information such as the labels of available objects, their location, their perception models etc.

Select substitutes

Based on the *available object information* and *tool unavailability information*, the *select substitutes* process identifies the substitutes among the available objects for a missing tool. For determining a suitable substitute, three kinds of knowledge are provided. *Object knowledge* consists of conceptual knowledge about objects. It can contain knowledge about physical properties of objects, functional properties of objects, different types of relations such as spatial relations between object and environment or between objects, temporal relations related to objects etc. *Self-concept knowledge* is based on the psychological term called self-concept proposed by a psychologist Carl Rogers [122]. According to him, self-concept encompasses one's knowledge about self, beliefs, dispositions, one's capabilities (mental as well as physical), preferences etc. The knowledge about self-concept consists of similar notions about a robot's self. It should be noted that besides *object knowledge*, *self-concept knowledge* such as one's physical capabilities or preferences can influence the selection of a substitute [2]. For instance, between a book and a tablet, it is possible that one may select the book over the tablet as a substitute for a tray and the other one may select the tablet. Such selection can be driven by one's *physical capabilities*: the book might be heavier for one person than a tablet, so the person selects the tablet. On the other hand, one may *prefer* to use the book in order not to use an electric device for carrying drinks. The *common sense knowledge* consists of general knowledge about the world a robot is operating which may contain naive physics knowledge, declarative (factual) knowledge about the world, causal relations between physical and functional properties etc. Such common sense knowledge can exert an influence on the selection of a substitute. The *select substitute* process takes into account these three kinds of knowledge when determining the suitability of a substitute for a missing tool. The selected candidates' information is then passed over in the form of *potential substitute information* to the next process: *use substitute*. The *potential substitute information* can consists of labels of the potential substitutes, their location, a map to reach the locations etc.

The primary focus of this doctoral research is on the substitute selection process for the given *available object information* and *missing tool information*. In our research work, we have considered *object knowledge* and have excluded *self-concept knowledge* and *common sense knowledge* for selecting substitutes.

Use substitute

After receiving the candidates for substitutes, the *use substitute* process takes a substitute provided in the *potential substitute information* and uses it in place of a missing tool. In order to use the substitute as the missing tool, multiple inputs are needed for the *use substitute* process to devise a plan to maneuver it so as to get the intended result had the missing tool been used. The inputs *tool manipulation model* and *tool grasping model* contain the manipulation and the grasping model of the tool. They essentially contain information about how a tool is to be grasped and manipulated which in turn can be used by the process to build a grasping and manipulation model for the substitute to be used as the tool. The input *object perception model* contains the perception model including vision and other modalities of the substitute which is essential to understand the physical structure of the substitute. Such understanding can be used by the process in building the required grasping as well as manipulation models for a substitute. Besides the perception model, the *self-concept model* provides vital information concerning the physical capabilities and preferences of the robot which forms the basis upon which the desired grasping and manipulation models can be built. Once the grasping and manipulation models are built, the *use substitute* process formulates a plan to use the substitute and executes the plan. After executing the plan, the *action outcome model* is sent to the *validate substitute* process to evaluate the action execution performance and whether the desired result was achieved.

Validate substitute

The *validate substitute* process determines the substitutability of a substitute by assessing whether the desired result was achieved after using the substitute. The selection of a substitute alone does not guarantee that the substitute is an accurate alternative for a missing tool. Humans determine the accuracy of a substitute for a missing tool or substitutability after using it and assessing the result. A similar approach is needed in case of a robot when assessing the substitutability. In order to determine the substitutability, two kinds of inputs are primarily needed. The *action outcome model* of the substitute-use provides information such as action execution plan, effects of the action, action parameters etc. The action parameters can contain information such as a substitute grasping model, a substitute manipulation model, substitute location, perception model of the perception model, the desired outcome of the action etc. The *substitute validation model* contains information about the desired effects of an action after using an actual tool. The validation process compares the validation model with the outcome model to determine whether the desired result was achieved. It is possible that a substitute may be accurate but the desired results was not achieved due to failure in action plan or execution or even in grasping or manipulation model [123]. The validation process can be equipped with a fault detection and diagnosis system where the system may offer a feedback on the cause of a failure and may suggest a recovery course.

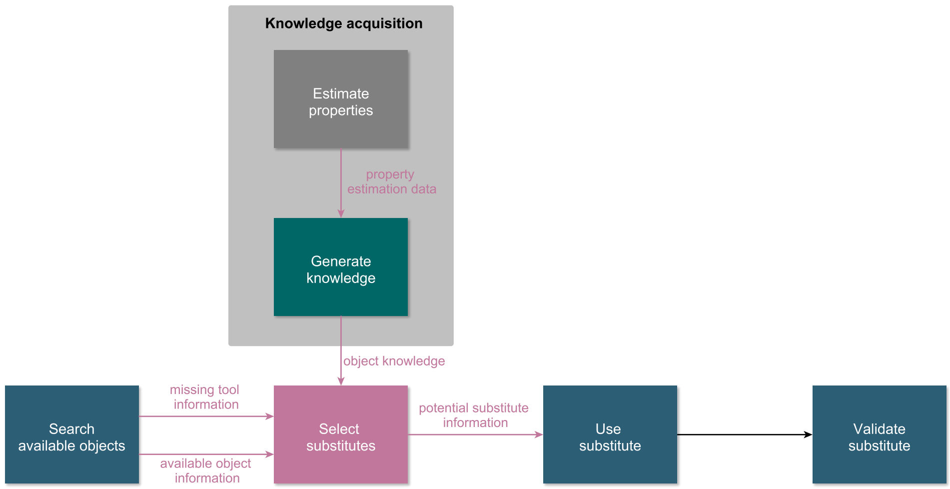


Figure 4.3: Outline of the tool substitution workflow and substitute selection workflow

Summary

It is evident from the workflow that building an autonomous tool substitution system for a robot requires integration of various capabilities such as object manipulation, object grasping, localization, navigation, fault diagnosis and recovery, object perception, knowledge acquisition, knowledge representation. It is a highly complex endeavour to accomplish and as a result, in this doctoral work we have focused on a single process: substitute selection (see Fig.4.3). The scope of the required inputs is as follows. The selection process receives a query which contains a label of a missing tool, and labels of available objects from which potential substitutes are to be determined. As for the required knowledge, we have taken into account object knowledge only. The self-concept knowledge or common sense knowledge is considered as out of scope in this work. It is also assumed that knowledge about objects received in a query exists in the knowledge base. In order to acquire the object knowledge, we have developed a knowledge acquisition system (see Fig. 4.3) which is independent of the substitute selection process. The knowledge acquisition system is composed of two modules: property estimation and knowledge generation. In the previous chapters, we have discussed both the modules in detail. The substitute selection process concludes when it outputs the labels of potential substitutes for a missing tool determined from the available objects.

4.2 Substitute Selection - Methodology

In this section, we discuss the methodology of our proposed approach to the substitute selection. We call our approach *ERSATZ* which is a German term for a *substitute*. Fig. 4.4 portrays a typical workflow of the proposed substitute selection approach and it was implemented in Python programming language. The workflow consists of four primary processes: *extract knowledge*, *generate representative models*, *determine relevant properties*, and *determine suitability*. In the following, we discuss each process in detail.

4.2.1 Extract knowledge

The ERSATZ workflow begins with the *extract knowledge* process when it receives a query which constitutes of the labels of the available objects and of a missing tool as input. The extraction process extracts the knowledge corresponding to the labels and it consists of qualitative symbolic knowledge about an object class instead of a specific instance of the ob-

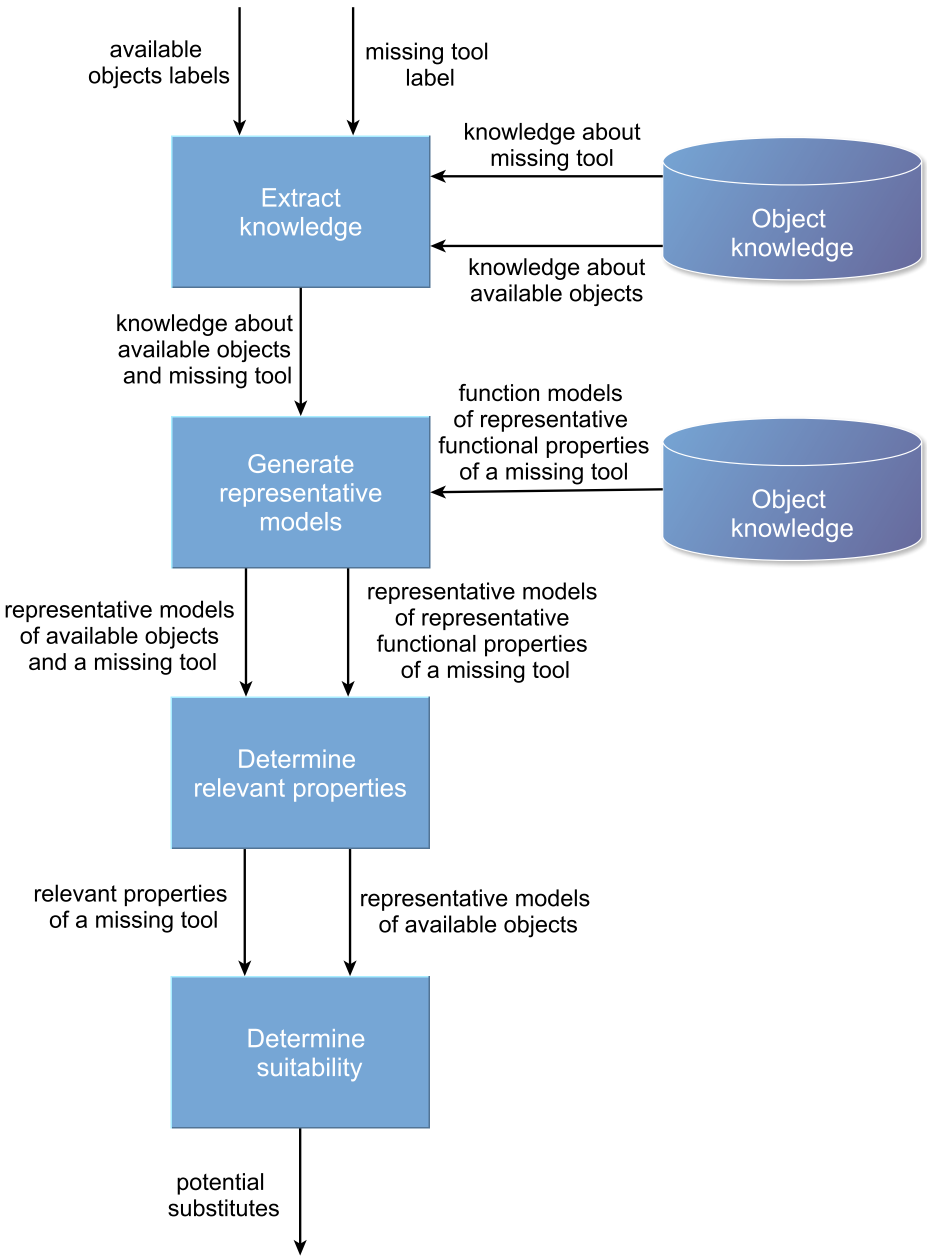


Figure 4.4: The workflow of the proposed substitute selection approach

ject class. The primary objective behind this process is to extract only the required knowledge instead of loading the entire knowledge base. It is to be noted that the knowledge base is independent of the substitute selection system. It is updated as and when a robot encounters new object classes or object instances. The substitute selection system can be seen as a user of the knowledge base. The knowledge base is generated using the proposed approach detailed in the Chapter 3. An illustration of extracted knowledge about an object class is graphically represented in Fig. 4.5. The example is taken from the knowledge base generated from the RoCS dataset detailed in 2.3. In the figure, knowledge about the class *plastic box* is shown which typically contains qualitative measurements of the physical and functional properties observed in its various instances. The property data is clustered into four clusters leading to four qualitative measurements for each property. As discussed in chapter 3, section 3.3, the numerical values represent the proportion of a qualitative measurement of a property observed in all of the instances of an object class. For instance, *flatness_0* which is a qualitative measurement of a physical property *flatness* is observed in *60%* of the instances of a plastic box. When a qualitative measurement is observed in all the instances of an object class, it is represented by a value *1.0*. For example, in the figure, qualitative measurements *hollowness_0, weight_0, movability_0, blockage_0* have been observed in all of the instances of a plastic box. In such cases, the remaining qualitative measurements of those properties such as *hollowness_1* or *weight_2* are assumed to be absent until new instances are encountered where they may be observed.

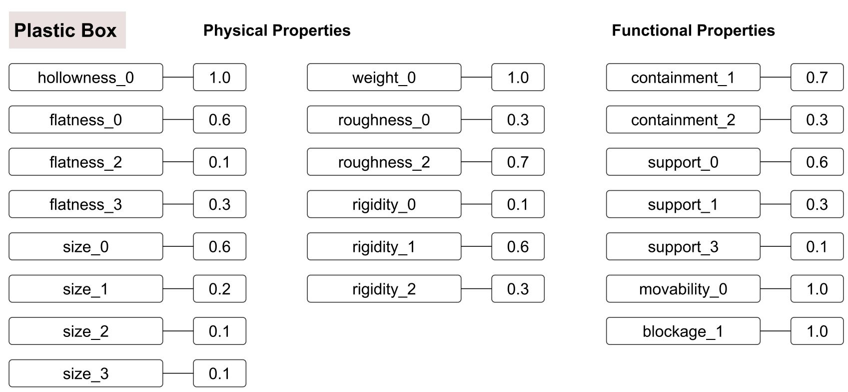


Figure 4.5: The image illustrates how knowledge about an object class typically looks. The example demonstrates the knowledge about the class *plastic box* containing its physical and functional properties.

4.2.2 Generate representative models

We noted in Sec. 4.1.2 that the relevant properties are the essential driving force to enable the designated purpose of a tool. As a result, the relevant properties of an object class corresponding to a tool will be observed across (almost) all the encountered instances of the object class. Consider, however a situation where a *white color* is observed in all of the instances of a cup. One may conclude that the *white color* is a relevant property, however, it is not. On the other hand, consider a qualitative measure *size_2* in Fig. 4.5. It is observed in only *10%* of all the instances which makes it less likely to be a relevant property. Both the examples prompt a means to filter out such properties when determining relevant properties. The *generate representative models* process provides the means to filter out qualitative measures like *size_2*.

The notion of a representative property and a representative model lies in the idea of *stereotypical* properties of an object class [124]. The properties which we usually associate with an object class, for instance, a hammer usually has a wooden stick and a black colored head or a stereotypical shape and size of espresso cups. *A representative property, in this work, is a qualitative measure of a physical or a functional property which is generally observed in the majority of the instances of an object class and provides a stereotypical identity to the respective object class. On the other hand, a representative model of an object class consist of the representative physical properties of the object class.* The generation of representative models is an essential intermediate step towards identifying relevant properties of a tool which is clarified further in the *determine relevant properties* process discussion. We hypothesize that a set of relevant properties is a subset of a set of representative properties and as a result, this process facilitates in ensuring that each relevant property is a representative property, however a representative property may not be necessarily a relevant property.

The process *generate representative models* is a two step process. In the first step, it generates representative models of the available objects and a missing tool wherein it takes the extracted knowledge about the available objects and the missing tool as input and returns representative (physical and functional) properties. In order to determine *representativeness*, we have introduced a *representative model threshold* value that filters the qualitative measures whose proportion value falls below the threshold. For instance, let us assume that the threshold is set at *0.30*, then qualitative measures whose proportion value is higher than the threshold will be considered as representative properties. The underlying idea is that if a qualitative measure is observed more often in the instances of an object class, it is likely that it is a stereotypical qualitative measure of the object class and if a qualitative measure is seldomly observed, then it is not a stereotypical measure of the object class. The *often-ness* or *seldom-ness* of a measure observed in the object class is represented in our work by a proportion value. Therefore, we have used a numerical threshold value to determine the stereotypical-ness or representativeness of a qualitative measure. Fig. 4.6 highlights the qualitative measures whose proportion value is higher than the threshold value of *0.30*. As a result they will be considered as representative properties of a plastic box. Note that, the process returns only the representative physical properties as representative model as illustrated in Fig. 4.6. They are not accompanied by their respective proportional values in the output. The experiment discussed in Sec. 4.3.1 fine tunes the universal value of the *representative model threshold*. The finely tuned threshold is then used in the subsequent experiments related to various substitution selection scenarios.

In the Chapter 3, Sec. 3.3.2, we discussed the notion of function models of qualitative measures of functional properties. A function model of a functional qualitative measure (a.k.a. a qualitative measure of a functional property) consists of proportion of physical qualitative measures (a.k.a. a qualitative measure of a physical property) observed along side the functional qualitative measure across all the functional qualitative measure observations in all the instances of all the object classes. Our primary objective behind a function model will be clarified in the discussion on the process *determine relevant properties*. For the second step, it is sufficient to know that the function models play a vital role in identifying the relevant properties. In the second step, the process extracts the function models of the representative functional properties of a missing tool from the knowledge base (see Fig. 4.6). A typical function model of a functional qualitative measure is illustrated in Fig. 4.7. The example is taken from the knowledge base generated from the RoCS dataset detailed in the Sec. 2.3. The figure illustrates the function model of a representative functional property

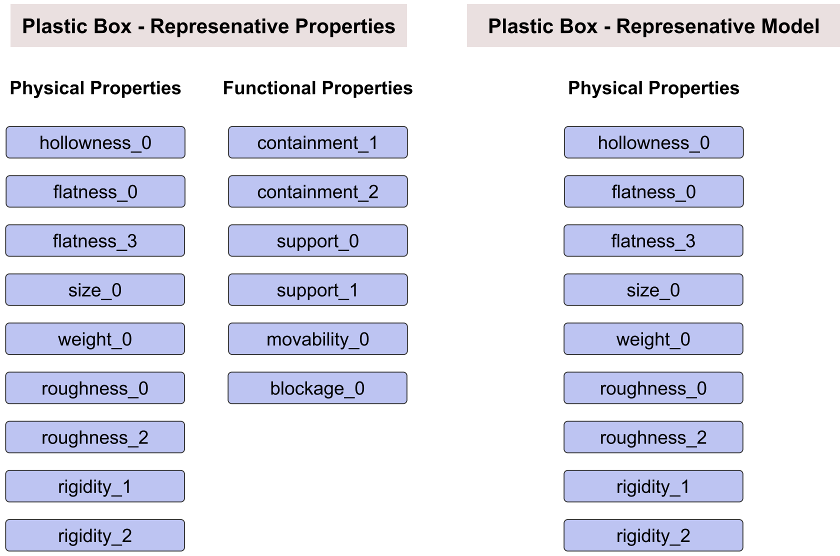


Figure 4.6: The left side of the image highlights the representative physical and functional properties of a plastic box after applying the representative model threshold of *0.30* to filter the proportion values of the qualitative measures (see Fig. 4.5). The right side of the image illustrates the representative model of a plastic box which contains only the representative physical properties.

movability_0 of a plastic box (see Fig. 4.6). For each extracted function model, a representative model is generated that follows the same principle as that for the representative models of available objects. The notion behind a representative model of a representative functional property is that the representative physical properties in the model are generally observed in the objects whenever the representative functional property is also observed in them. In other words, it is likely that the representative physical properties of the model or a subset of them enable the representative functional property in general. In Fig. 4.7, a representative model of *movability_0* is illustrated wherein all the representative physical properties of *movability_0* are listed. At the end of the second step, the process *generate representative models* returns representative models of the available objects, a missing tool and representative functional properties to the next process *determine relevant properties*.

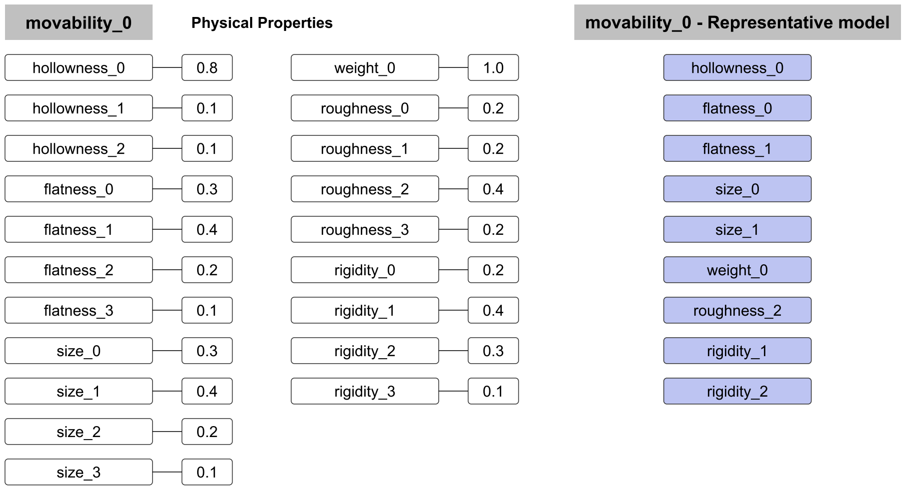


Figure 4.7: On the left side of the image illustrates a function model of *movability_0* which consists of the qualitative measures of physical properties and their corresponding proportion value indicating the proportion of qualitative measures of a physical property observed whenever *movability_0* is observed across all the observation of *movability_0* observed in the instances of all object classes. On the right side is the representative model of *movability_0*.

4.2.3 Determine relevant properties

In Sec. 4.1.2, we explained the notion of relevant properties of a tool. In this section, we discuss how to determine them. *The relevant properties of a tool are qualitative physical and functional properties observed in the tool which enable its designated purpose.* As per suggested in [3] that a certain assemblage of physical properties is essential prerequisite to enable a functionality in a tool, we propose the following. In order to determine the relevant properties of a tool, we hypothesize that there exists a subset of representative physical properties and a subset of representative functional properties of a tool such that the subset of representative physical properties enable the subset of representative functional properties which in turn enable the designated purpose of the tool. Therefore, we can infer that the subset of representative physical properties enable the designated purpose of the tool. In other terms, if the subset of representative physical properties are observed in the tool, then it can be inferred that those properties will enable the designated purpose of the tool. Such subset of representative physical and functional properties are considered as relevant properties of a tool. This hypothesis forms a basis for our approach to determine the relevant properties.

The relevant properties of a tool are determined in two stages. In the first stage, the relevant functional properties of the tool are identified and in the second stage, the relevant physical properties of the tool are determined. As stated in the previous discussion on *generate representative models*, a relevant property, physical or functional, is a representative property. Based on this, in the first stage, we compare the representative model of a representative functional property of a missing tool with a representative model of the missing tool to determine whether the representative functional property is a relevant functional property. The underlying principle is, if both the representative models are similar, then the representative functional property is considered as a relevant functional property and the shared representative physical properties are considered as relevant physical properties. Since the representative models are sets of representative physical properties, we propose the use of set-based similarity measure to determine the similarity between the aforementioned two representative models.

A set based similarity measure is typically used to determine the similarity between two sample sets. While computing a similarity between the two sets in set-based similarity measure, various factors are taken into considerations such as the number of members shared by both the sets and/or the number of members in either set and/or number of members in the union of both the sets. Various set-based similarity measures utilize these factors differently, for instance, in *Jaccard Similarity*, the number of members in the intersection of both the sets is divided by the number of members in the union of both the sets [125] (see Fig. 4.8). In *Overlap coefficient*, also known as *Szymkiewicz–Simpson coefficient*, the number of members in the intersection of the two sets is divided by the number of members in a smaller set of the two sets [126]. On the other hand, in the *Sørensen–Dice coefficient*, twice the number of members in the intersection of the two sets is divided by the sum of the number of members in each set [127].

We are primarily interested in the proportion of the representative physical properties that are shared by both the aforementioned representative models. This makes the Jaccard Similarity a suitable metric to determine the similarity between the two representative models and subsequently,

$A = \{m, a, d, h, u, r\}$

$B = \{m, e, r, a\}$

$A \cap B = \{m, a, r\}$

$A \cup B = \{m, a, d, h, u, r, e\}$

(i)

$$\text{Jaccard Similarity} = \frac{\text{Size of the intersection of two sets}}{\text{Size of the union of two sets}}$$

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|}$$

$$J(A, B) = \frac{3}{7}$$

$$J(A, B) = 0.43$$

(ii)

Figure 4.8: The image illustrates the Jaccard Similarity formulation. It is calculated by dividing the number of members in the intersection of two sets with the number of members in the union of the two sets. The similarity outcome lies between 0 and 1. If the value is 0, it means both the sets are dissimilar as they do not share any common members. If the value is 1, it means both the sets are similar as they have the same members.

whether the representative functional property is relevant to the missing tool. Fig. 4.8 illustrates the the computation for determining the similarity using Jaccard's Similarity metric. In the figure, we have two sets A and B containing alphabet characters. As Jaccard's Similarity determines the proportion of the common members in both the sets, we first need to identify the members shared by both the sets, represented as $A \cap B$. The next step is to identify the total members in both the sets, represented as $A \cup B$. The Jaccard's Similarity is calculated by dividing the size of the intersection of two sets $(A \cap B)$ by the size of the union of two sets $(A \cup B)$. In the figure, the similarity between the sets A and B is 0.43. It should be noted that the Jaccard's Similarity lies between 0 and 1.

When the metric is applied to the representative models, the value 0 would indicate that the representative models of a missing tool and of a representative functional property do not share any representative physical properties, thus, making the representative functional property not a relevant functional property of the missing tool. If, however, the value is 1, then both the representative models contains the same representative physical properties, thus making the representative functional property a relevant functional property of a missing tool and all the representative physical properties of the representative model of the representative functional property are considered as relevant physical properties of the missing tool.

In case the value lies between 0 and 1, we propose a *Minimum Similarity Tolerance* threshold to determine if the representative functional property is a relevant functional property of the missing tool. The *Minimum Similarity Tolerance* embodies the notion that when determining similarity for substitute selection purposes, a minimum similarity is desired as opposed to maximum similarity. There are two primary reasons to aim for minimum similarity. Firstly, it is highly unlikely to always find a substitute that is an exact match to a missing tool. For instance, it is not possible to find a substitute for a hammer which is identical to the hammer structurally. Consider a stone, for instance, which is a substitute for a hammer, however, it is not structurally identical to the hammer. It may not have all the relevant properties of a hammer, but *enough* to be used as a hammer. We have attempted to capture this notion of *enough-ness* of similarity in the *Minimum Similarity Tolerance*. Secondly, by targeting minimum similarity, the scope of the possible choices for a substitute from the available objects becomes wider. This way, the chances of finding a substitute from these choices are also increased. The value for the *Minimum Similarity Tolerance* is determined in the experiment discussed in section 4.3.1 which is set at *0.45*. If the Jaccard Similarity between the representative models of a representative functional property and a missing tool is higher than the *Minimum Similarity Tolerance* threshold then the representative functional property is considered as a relevant functional property of the missing tool. Subsequently, the shared representative physical properties between the two representative models are considered as relevant physical properties of the missing tool.

Fig. 4.9 demonstrates the Jaccard Similarity calculations between the representative models of a plastic box and a representative functional property *movability_0*. The first step for calculating the similarity is to determine the number of representative physical properties shared by both the models which is *7* and the total number of representative physical properties in the union of both the models which is *11*. The similarity between the models accordingly is *0.64* which is higher than the set threshold of *Minimum Similarity Tolerance*. As a result, *movability_0* is considered as

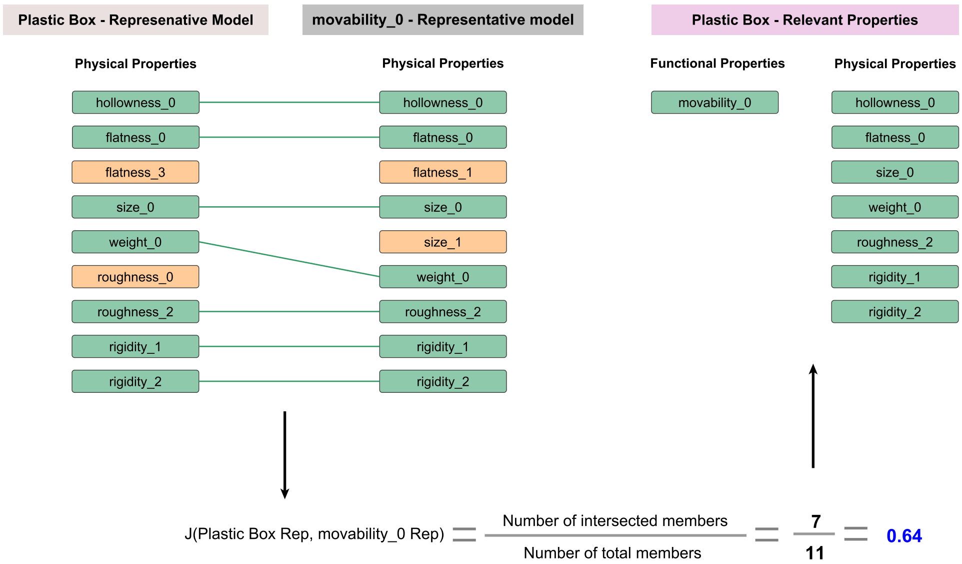


Figure 4.9: The image illustrates the Jaccard Similarity between the representative models of a *Plastic Box* and a representative functional property *movability_0*. The representative physical properties shared by both the models are written in green colored box while the properties that are not shared by them are highlighted in yellow color for clarity. The Jaccard Similarity is calculated to be 0.64. The relevant properties are a plastic box are depicted subsequently.

a relevant functional property the plastic box and the shared representative physical properties are considered as the relevant physical property of the plastic box. In a similar fashion, the Jaccard Similarity is calculated between the representative models of a missing tool and of the remaining representative functional properties and subsequently the relevant physical properties from each similarity calculations are identified. In Fig.4.10, for instance, the similarity between the representative models of a representative functional property *support_1* of a plastic box and that of a plastic box is calculated to be 0.39. Since the similarity is less than the tolerance threshold, *support_1* is not considered as a relevant functional property. However, in Fig. 4.11, the similarity between the representative models of *support_0* and a plastic box is calculated to be 0.78 which is greater than the tolerance threshold. As a result, *support_0* is considered as a relevant functional property. Note that *support_0* and *support_1* are representative functional properties of a plastic box. The final set of relevant physical properties of the missing tool is the union of the relevant physical properties determined from each similarity calculation. In Fig. 4.11, the revised relevant functional properties of a plastic box are shown while the relevant physical properties have remained the same as shown in Fig. 4.9. The process concludes by returning the representative models of the available objects generated in the previous process *Generate representative models* and the relevant physical properties of the missing tool determined in this process. Both the outcomes are inputted to the next process *Determine suitability* wherein the potential substitutes are determined among the available objects.

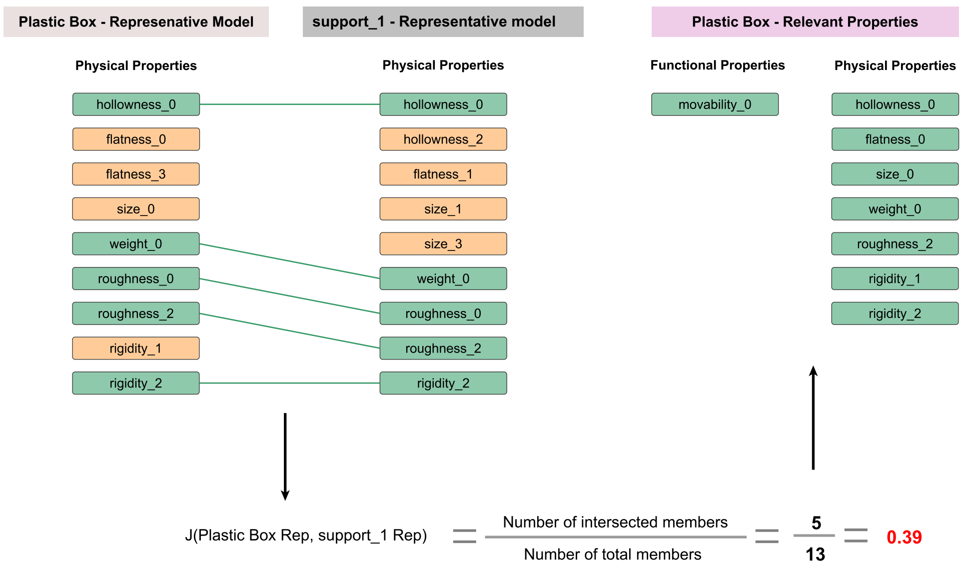


Figure 4.10: The image illustrates the Jaccard Similarity between the representative models of a *Plastic Box* and a representative functional property *support_1*. The representative physical properties shared by both the models are written in green colored box while the properties that are not shared by them are highlighted in yellow color. The Jaccard Similarity is calculated to be 0.39. Since the similarity value is lesser than the threshold, *support_1* is not a relevant functional property of a plastic box.

4.2.4 Determine suitability

We saw in the earlier discussion on relevant properties (see Sec. 4.1.2) that in order to determine the suitability of an object as a possible substitute for a missing tool, it is imperative that the object has the required relevant physical properties of the missing tool. The presence of the relevant physical properties in the object facilitates the selection or rejec-

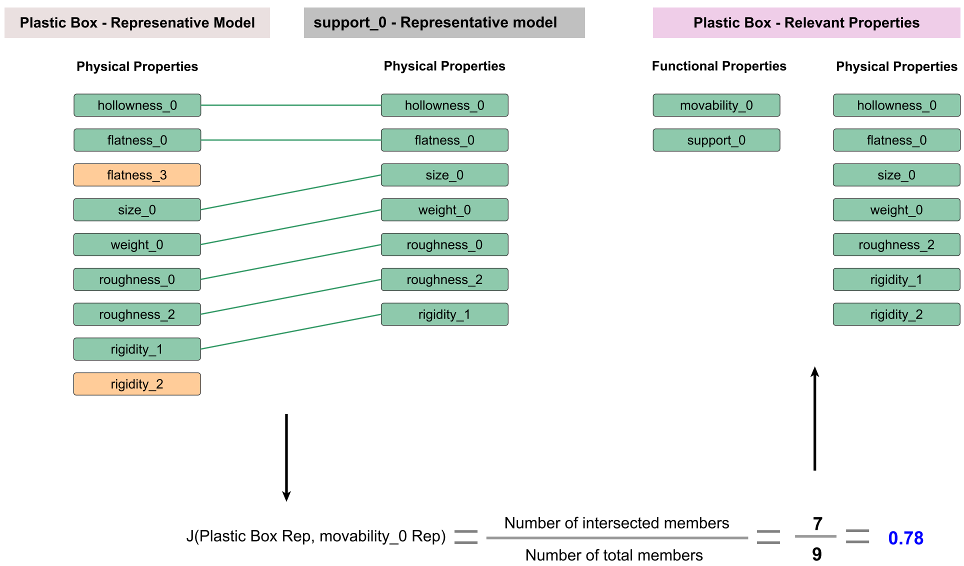


Figure 4.11: The image illustrates the Jaccard Similarity between the representative models of a *Plastic Box* and a representative functional property *support_0*. The representative physical properties shared by both the models are written in green colored box while the properties that are not shared by them are highlighted in yellow color. The Jaccard Similarity is calculated to be 0.78. The relevant properties are a plastic box are depicted subsequently.

tion of the object as a possible substitute for the missing tool. For determining the suitability of an available object, the representative model of the available objects is supplied. The question is why the representative models are required for such determination. Consider a mouse pad as a possible option for a substitute for a tray. Mouse pads *typically* are not rigid, however, there are certain instances where they are rigid enough to be considered as a possible substitute for a tray. Whether the available mouse pad is a suitable substitute for the tray can be determined by estimating the required relevant physical property (qualitative) measurements of the mouse mad using the proposed property estimation and knowledge generation approaches, and compare them with the qualitative measurements of the relevant physical property of the tray. For a single object instance, this may be doable. However, when there are multiple available objects, such instance-level inspection of the relevant property measurements would be time consuming. Note that the available objects are typically instances of various object classes. In such case, it is practical to perform the inspection at a class level. In order to perform such inspection, it is prudent to assume that the physical properties of an available object that are being compared with the relevant physical properties of a missing tool are *commonly observed* or *stereotypical* in the other instances of the class of the available object. We noted in the previous process discussion that representative models manifest the notion of stereotypical understanding of an object class. Therefore, when the relevant physical properties of a tool overlap with the representative model of an available object, it is safe to assume that the overlapped relevant physical properties are also present in the available object. The assumption is necessary as it circumvents the instance-based inspection. In the following, we explain how the representative models of the available objects are compared with the relevant physical properties of a missing tool in order to identify possible substitutes among the available objects.

Figure 4.12: image illustrates the suitability calculations between a plastic box, a metal box and a ball.

In this work, the suitability of an available object as a possible substitute for a missing tool depends on how similar it is to the missing tool with respect to its (missing tool) relevant properties. The process *Determine suitability* computes the similarity between the missing tool and the available object using Jaccard's similarity. The *Minimum Similarity Tolerance* determines whether the similarity leads to the suitability of the available object as a possible substitute for the missing tool. The suitability computation takes place in two steps: In the first step, upon receiving the relevant properties of a missing tool and representative models of the available objects, the process computes similarity between relevant properties of the missing tool and each of the representative models. Each similarity is compared with the tolerance threshold. If it is higher than the threshold, then the corresponding available object is regarded as a potential substitute for the missing tool and if it is lesser than the threshold then it is regarded as a negative substitute for the missing tool. At the end, a set of such potential substitutes is then forwarded to the *Use substitute* process as depicted in figure 4.2.

Fig. 4.12 illustrates the suitability computation. Consider that a plastic box is a missing tool and a metal box and a ball are the available objects. Let, for the sake of the illustration, the relevant properties calculated in the Fig. 4.11 be the relevant properties of the plastic box. Let us consider the representative models of the metal box and the ball as shown in figure 4.12. Fig. 4.12 a) illustrates the similarity calculation between the relevant properties of the plastic box and the representative model of the metal box. The lines highlighted in green color connect the similar properties written inside green box on both the sides while the properties that are dissimilar are written inside orange box. Once the number of similar properties and the size of the union of both the sets is counted, Jaccard's similarity is calculated to be *0.60*. Since the similarity is greater than the *Minimum Similarity Tolerance* threshold which is set to *0.45*, the metal box is considered as a possible substitute for a plastic box. The steps are repeated for the example in the figure 4.12 b) where the similarity between the plastic box and the ball is calculated to be *0.27*. Since it is les than the threshold, the ball is regarded as a negative substitute for the plastic box. The process concludes by sending the labels of the positive substitutes as *potential substitutes* to the *Use substitute* process of the tool substitution workflow shown in the Fig. 4.2.

4.3 Evaluation

In this section, we discuss the experimental evaluation of our proposed substitute selection approach. We have conducted four experiments where each experiment aims for a specific aspect of the substitute selection approach.

The evaluation begins with a parameter tuning experiment to determine the optimal values of the three parameters: *number of clusters* to determine the number of qualitative measures for properties, *representative model threshold* to determine the representative properties of the objects, *minimum similarity tolerance* to determine the suitability of an object as a substitute for a missing tool.

The second experiment deals with a suitability of a substitute of a missing tool. As we discussed earlier in Sec. 4.1.3, a substitutability of a substitute can be assessed after using it in a task in place of a missing tool and examine if the desired result was achieved. Since such assessment is out of the scope of this doctoral work due to the sheer complexity of designing such a validation system, we assessed the *suitability* by comparing the selection of substitutes for various missing tools by our system with the selection of the same by human experts.

The third experiment primarily focuses on the significance of the role relevant properties play in the similarity computation. As we noted earlier, we distinguish between a tool and a substitute on the basis of the relevant properties. The experiment demonstrates why such relevant property based distinction is necessary.

4.3.1 Parameter Tuning

The objective of this experiment is to optimize the following three main parameters which play a vital role in our proposed approaches and can affect substitute selection performance:

Number of clusters The number of clusters refers to the number of qualitative measures of a property. The parameter is required during the sub-categorization process discussed in chapter 3 Sec. 3.3.1. The sub-categorization process generates the qualitative measures for each property from the quantitative measurements of the property using a clustering technique. Our goal is to identify the optimal number of clusters which is applicable for all the properties.

Representative model threshold As we have seen in Sec. 4.2.2, the representative model threshold facilitates the determination of representativeness of a qualitative measure of a physical or a functional property. Similar to the number of clusters, the objective is to identify the optimal value for the threshold which is universal for the properties.

Minimum similarity tolerance The minimum similarity threshold is used to identify the relevant functional and physical properties as detailed in Sec.4.2.3. The threshold is also used to determine whether an available object is a suitable substitute for a missing tool as described in Sec. 4.2.4. The objective of this experiment to identify the optimal value for the tolerance threshold which is universal for the properties.

Experimental Setup

For the experiment, we need the following: 1) knowledge about objects, 2) various missing tool scenarios, 3) substitute selection by the proposed system, and 4) by human experts for each missing tool scenario. The substitute selection by human experts will act as a ground truth against which the substitute selection by the proposed system will be compared. Such validated substitutes will facilitate the determination of the optimal values for the aforementioned parameters.

In order to generate conceptual knowledge about objects using our proposed knowledge generation approach, we need metric data about objects' properties. For this experiment, we generated a dataset of the quantitative measurements of the physical and functional properties whose acquisition is focused on the composite of a machine-centric and a human-centric method. In order to generate conceptual knowledge about objects using our proposed knowledge generation approach, we need metric data about objects' properties. In the classical machine learning setting, such metric data would be called as training data. Since we want to use RoCS dataset for testing our substitute selection approach along with the tuned parameters, it is vital that another dataset is generated on which the parameters can be tuned and the knowledge generation as well as the substitute selection approach can be examined. As a result, for this experiment, we generated a dataset of the quantitative measurements of the physical and functional properties whose acquisition is focused on the composite of a machine-centric and a human-centric method. The reasons to generate the dataset using a different methodology are twofold: 1) to examine how the parameters tuned on this dataset performs on a dataset generated using a different methodology, 2) to examine the robustness of our knowledge generation approach, its application and the substitute selection approach by evaluating how do they perform on two different datasets based on different estimation methodologies.

Dataset Generation: To generate the (training) dataset, we have used household object images from the RGB-D Washington Dataset [128]. The Washington dataset contains total 300 everyday object instances covering 51 object categories where each object instance is captured from multiple view angles that leads to total 250,000 RGB-D images[2]. There are primarily four super object categories: *fruits*, *vegetables*, *devices*, and *containers*. For this experiment, we targeted *devices* and *containers* super categories from which 22 object classes were selected and for each class, we selected random images from all the given object instances of the class leading up to total of 692 images[3]. Table 4.1 illustrates the number of scans per instance and the number of instances selected from each class. In the machine-centric approach, geometrical properties were acquired using a state-of-art non-invasive object shape learning technique [129] which in a data-driven and unsupervised manner learns shape concepts from RGB-D object point clouds as shown in Fig. 4.13. In total, 58 geometrical properties or shape concepts were generated using the unsupervised approach from which four shape concepts were selected using a baseline feature selection technique *Variance Threshold*. We limited the number of shape concepts to four in order not to skew the relevant properties of a missing tool and by extension substitute selection in favor of the shape concepts. The learned shape concepts for the objects are used in the knowledge as machine-generated geometric object properties and are denoted as

[2] Check the link to the dataset: `https://rgbd-dataset.cs.washington.edu/dataset/`

[3] The dataset is available on this link: `https://rgbd-dataset.cs.washington.edu/dataset/rgbd-dataset/`

concept1, *concept2*, etc. However such properties provide a particular facet (object geometry) of the object's physicality which is not sufficient for determining substitutes. In order to enrich the object perception, we also considered non-geometrical physical properties such as *weight*, *rigidity*, *hollowness* and the functional properties like *support*, *blockage* and *containment*. Note that, in general, these properties are challenging and cumbersome to estimate solely from non-invasive visuoperceptual approaches. Consequently, estimating such properties via multi-modal or manipulation capabilities, was beyond the reach at the time of the dataset generation. Therefore, these properties were synthetically acquired by sampling from human expert knowledge. For each object class, based on the selected object instances, the measurement distribution of each property was approximated by drawing random samples from a normal (Gaussian) distribution. The distribution for each property was generated by providing the mean and standard deviation for each class, for the given number of images of each class where the mean and the standard deviation was provided by a human expert. We generated the dataset in Python using NumPy random normal function generator[4] given by,

```
numpy.random.normal(loc=0.0, scale=1.0, size=None)
```

where *loc* is Mean of the distribution, *scale* is Standard Deviation of the distribution and *size* is the shape of the output array, in our case, it will be the number of instances selected for each class. We have provided the numerical dataset in our git repository (see appendix D).

The conceptual knowledge about 22 object categories was generated using proposed knowledge generation approach (see Chapter 3). A baseline clustering technique *k-means* was used to generate qualitative measures of each property. The knowledge base is provided in our git repository (see appendix D). For the experiment, we generated 22 queries based on 22 object categories. The queries are provided in the Table 4.2. Each query consists of a missing tool and 5 randomly selected objects from which a substitute was to be selected. The queries were run on ERSATZ which selected a substitute/s for each query using the generated knowledge about

[4] `https://numpy.org/devdocs/reference/random/generated/numpy.random.normal.html`

[5] The instance labels are same as used in the dataset: `https://rgbd-dataset.cs.washington.edu/dataset/rgbd-dataset/`

Table 4.1: Number of images (#) about the instances (Instance Labels) of each object class ($\Sigma\# = 692$) selected from the Washington RGBD dataset [128].

Class label	**Instance Labels**[5]	**Images per Instance**	#
ball	1-7	5	35
binder	1-3	10	30
bowl	1-6	5	30
cap	1-4	8	32
cereal box	1-5	6	30
coffee mug	1-8	4	32
flashlight	1-5	6	30
food bag	1-8	4	32
food box	1-12	3	36
food can	1-14	2	28
food cup	1-5	6	30
food jar	1-6	5	30
hand towel	1-5	6	30
keyboard	1-5	6	30
kleenex	1-5	6	30
notebook	1-5	6	30
pitcher	1-3	10	30
plate	1-7	5	35
shampoo	1-6	5	30
soda can	1-6	5	30
sponge	1-12	3	36
water bottle	1-9	4	36

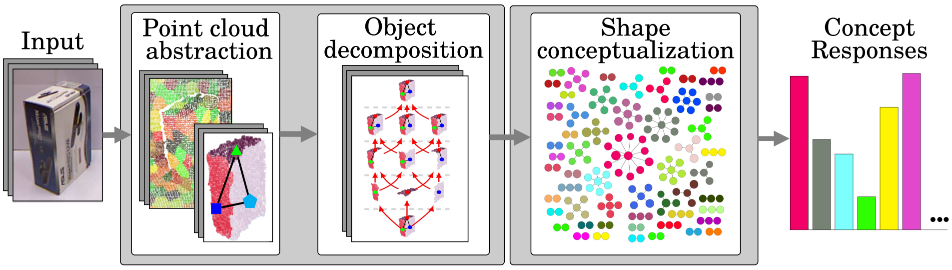


Figure 4.13: Illustration of the object shape conceptualization approach [129]. Concepts (connected components graphs) are randomly colored.

22 object categories. As we needed the *validation of the selected substitutes* for a missing tool in order to assess a parameter's performance, we gave 22 queries to 13 human experts and asked them to select a substitute/s in each query. We noted earlier that humans generally validate a substitute by using it in place of a missing tool, however employing such feedback mechanism in a robot is an extremely complex endeavour and is out of the scope of this work. Therefore, the experts' selection of the substitutes is treated as a ground truth in order to validate the suitability of the selected substitutes by ERSATZ.

The human experts were selected such that they represent diverse background, research experience and age range. The age range of human experts was between 22 and 42 years old. The backgrounds of the experts consisted of doctoral researchers, post-doctoral researchers, students. Only one expert had a background in robotics, the rest of them were from different areas of computer science such as data science, database management systems, swarm intelligence, machine learning, natural language processing and theoretical computer science. We made sure that none of the human experts had any prior research experience in tool use, tool affordances, or tool substitution as we did not want their research experience influence their selection process. We also ensured that the proposed approach to substitute selection was not discussed with them or neither had they read any research papers in similar areas. Each query given to the human experts consisted of a missing tool, its designated purpose, and the available objects. The designated purpose was provided to the experts in order to avoid the multiple interpretations of a missing tool which may affect the substitute selection.

Missing Tool	**Available Objects**				
ball?	coffee_mug	food_cup	cereal_box	keyboard	flashlight
binder?	flashlight	coffee_mug	notebook	water_bottle	bowl
bowl?	hand_towel	ball	shampoo	pitcher	soda_can
cap?	bowl	food_jar	food_box	coffee_mug	notebook
cereal_box?	coffee_mug	food_cup	ball	flashlight	food_jar
coffee_mug?	flashlight	food_can	keyboard	notebook	bowl
flashlight?	food_box	food_cup	ball	water_bottle	plate
food_bag?	food_box	hand_towel	flashlight	coffee_mug	notebook
food_box?	food_jar	food_cup	soda_can	kleenex	cereal_box
food_can?	flashlight	cereal_box	food_cup	food_box	cap
food_cup?	keyboard	pitcher	plate	soda_can	sponge
food_jar?	food_cup	flashlight	notebook	coffee_mug	soda_can
hand_towel?	food_cup	plate	shampoo	food_can	flashlight
keyboard?	bowl	cereal_box	food_can	notebook	food_box
kleenex?	cap	water_bottle	ball	shampoo	flashlight
notebook?	ball	water_bottle	plate	bowl	hand_towel
pitcher?	plate	hand_towel	cereal_box	ball	flashlight
plate?	coffee_mug	food_box	kleenex	pitcher	water_bottle
shampoo?	food_can	food_cup	pitcher	flashlight	food_bag
soda_can?	ball	shampoo	food_box	flashlight	food_bag
sponge?	keyboard	coffee_mug	bowl	flashlight	hand_towel
water_bottle?	bowl	cereal_box	notebook	sponge	soda_can

Table 4.2: The 22 queries generated based on 22 object categories. Each query consists of a missing tool and five available objects

Result

In the experiment, the values of the parameters *number of clusters* were varied between 2 and 8. While, for *representative model threshold* and *minimum similarity tolerance*, the values were varied between 0.25 and 0.50. In order to optimize the given parameters, we targeted three measures: overall substitute selection accuracy, scenario-wise accuracy and overall false positives. The ground truth was provided by the human experts where out of 110 available objects spread across 22 scenarios, they identified 55 objects as substitutes and rest of 55 as not substitutes. We used the binary classification metric *Accuracy* to measure the accuracy of the substitute selection across all the scenarios.The typical binary classification measures are:

- *True Positive (TP):* A positive substitute is considered as a *true positive* if it is selected by ERSATZ and at least by a single expert.
- *True Negative (TN):* A negative substitute is considered as a *true negative* if it is not selected by either ERSATZ nor by any expert.
- *False Positive (FP):* A positive substitute is considered as a *false positive* if it is selected by ERSATZ but not selected by any expert.
- *False Negative (FN):* A negative substitute is considered as a *false negative* if it is not selected by ERSATZ but selected by at least one expert.

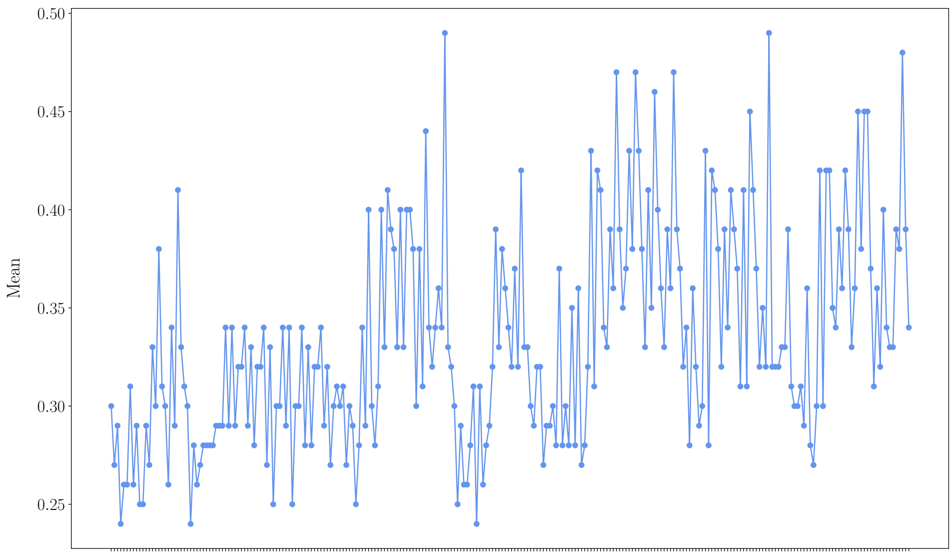
The overall accuracy aims for the accuracy across all the scenarios wherein it considers the sum of all true positives, true negatives, false positives, false negatives and is measured using the formula below:

$$\textit{Overall Accuracy} = \frac{\sum TP + \sum TN}{\sum TP + \sum TN + \sum FP + \sum FN}$$

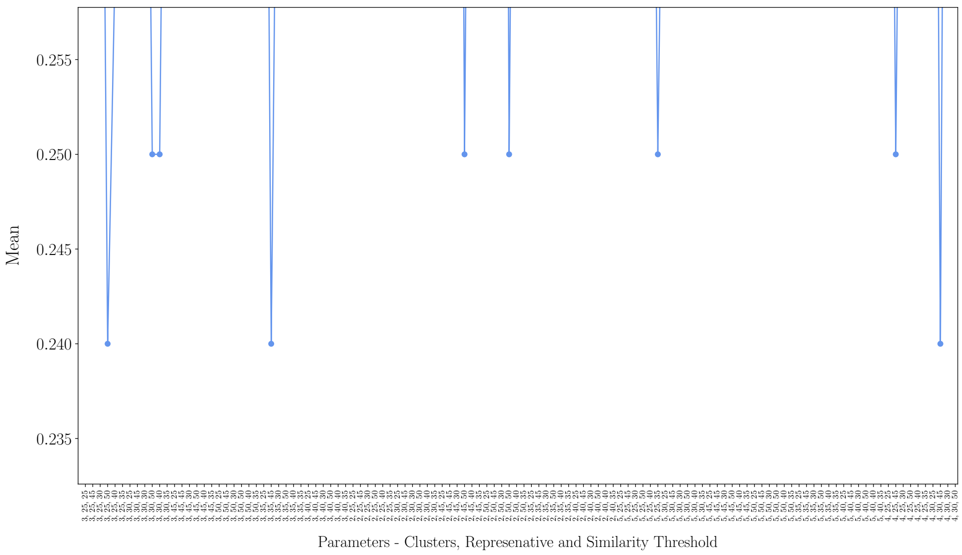
For scenario-wise accuracy measure, we measured accuracy in each scenario wherein we selected those scenarios which contain at least one true positive. The notion behind this measure is to determine the number of scenarios where ERSATZ and the experts have selected the similar substitutes. The false positives, on the other hand, provide the crucial information concerning the selection of the substitutes by ERSATZ but not selected by the experts. In order to determine the optimal values of the three parameters, we calculated overall accuracy, scenario-wise accuracy and false positives for each combination of the values of the parameters. For a combination of the values of the parameters to be chosen as optimal values, the overall accuracy and the scenario-wise accuracy should be maximized and the overall false positives should be minimized. To achieve this target, we first normalized the scenario-wise accuracy and the false positives by dividing them with total number of scenario and the sum of available objects from all scenarios respectively. In the next step, the mean of the distances from the maximum attainable value of the overall accuracy, scenario accuracy and from the minimal attainable value of the false positives was taken. The optimal values of the parameters were considered those which had the lowest mean value. The plot in Fig. 4.14(a) illustrates the mean values calculated for all the 252 combinations of parameters values, whereas the plot in Fig. 4.14(b) illustrates a snap-shot of the mean values on Y-axis at different combination of parameter values on X-axis where mean values are the lowest. Since the combination (4, 30, 45) has the lowest mean value which is calculated to be 0.239, in the subsequent experiments, we used 4 for the number of clusters, 30 for the representative threshold and 45 for the minimum similarity tolerance as the optimal values.

4.3.2 Substitute Validation

We have noted earlier that a substitute is an approximation of a missing tool that has the capacity to achieve a similar result as the missing tool when used in place of the missing tool. When a substitute is selected, one can only infer that it may achieve the similar result based on certain closeness to the missing tool or based on the past experience. As we noted in the Sec. 4.1.3, besides a past experience, the validation of whether it can achieve the similar result as the missing tool can only be substantiated after the usage. As discussed in the Sec. 4.1.3, substitute validation by a robot is a highly complex endeavour and therefore, it is out of the scope of this work. Consequently, we would have to rely on other means for the purpose of evaluating whether ERSATZ can select possible substitutes for a given missing tool successfully. With that aim in mind, we have resorted to determine the validity of a substitute by comparing the substitute selections by our proposed system, ERSATZ, with the selections by human



(a) The mean values calculated for all the 252 combinations of the three parameters.



(b) The snap shot of the plot where mean values are the lowest.

Figure 4.14: As illustrated in the plot, the combination (4, 30, 45) representing the number of clusters, representative and similarity threshold respectively has the lowest mean value which is calculated to be 0.239

experts for various missing tool scenarios. The objective of this experiment is two fold: 1) to examine the *transferability* of the optimized parameters on a different dataset. Therein, the transferability is examined by comparing the performance of ERSATZ on two knowledge bases generated from the two distinct properties measurements datasets; 2) to validate the substitute selections for various missing tool scenarios on two different datasets. It should be noted that, the knowledge used by ERSATZ and by human experts may differ. Additionally, the selection by human experts will be influenced by their own preferences, experiences and/or self-concept knowledge. Therefore, for validation purpose, instead of focusing on overall accuracy, our aim is to find out, in how many scenarios ERSATZ and human experts selected same substitute/s. Additionally, we will also assess the performance of ERSATZ with regard to the frequency distribution of experts' selection of substitutes in each scenario.

For this experiment, we are going to need two distinct knowledge bases about objects. These two knowledge bases will be generated from two distinct datasets about property measurements of objects. One dataset is created from the Washington Dataset discussed in the previous experiment, while the second dataset, known as RoCS Dataset, is acquired using solely machine-centric methods using our property estimation approach where the data is acquired from the real objects. The dataset is discussed in detail in the Chapter 2. Similar to the previous experiment, we will also need for this experiment various missing tool scenarios, substitute selection by ERSATZ and by human experts for each missing tool scenario. The substitute selection by human experts will act as a ground truth against which the substitute selection by ERSATZ will be compared. As we noted in the Sec. 4.1.3, a substitute selection is subjective where the selection can be affected by self-concept knowledge. Therefore, to get a variety of possible selection of substitutes in various missing tool scenarios in our experiment, we invited multiple human experts instead of a single expert to select their choices. Similar to the parameter tuning experiment, the human experts were selected such that they represented diverse background, research experience and age range.

Experimental Setup:

In the RoCS dataset, the metric data related to the physical and functional properties was acquired from 110 objects, comprised of 11 object classes containing 10 instances per class, using our property estimation method. The metric data was used to generate the knowledge about 11 object classes using the baseline clustering technique *k-means*. We have made the knowledge base and the RoCS dataset available in our git repository (see appendix D). Similar to the parameter tuning experiment, we formed 11 queries consisting of one missing tool and five available objects as illustrated in the Table 4.3. The queries were run on ERSATZ and were also given to 21 human experts. Each query given to the human experts consisted of a missing tool, its designated purpose, and the available objects. The designated purpose was provided in order to avoid the multiple interpretations of a missing tool which may affect the substitute selection. Similarly, we used the experimental setup for the parameter tuning experiment with regard to the knowledge about 22 object classes, 22 missing tool scenarios and 13 human experts to select the substitutes in the 22 missing scenarios. The knowledge generation and the substitution selection in both the experiments were carried out using the optimized parameters calculated in the Sec. 4.3.1: 4 for the number of clusters, 30 for the representative threshold and 45 for the minimum similarity tolerance as the optimal values.

Missing Tool	**Available Objects**				
plastic_box?	metal_box	bowl	tray	plate	sponge
bowl?	plastic_box	sponge	cup	to_go_cup	plate
to_go_cup?	tray	cup	book	paper_box	plastic_box
paper_box?	plastic_box	to_go_cup	plate	ball	book
metal_box?	bowl	sponge	cup	book	plastic_box
tray?	cup	plastic_box	book	sponge	plate
plate?	book	metal_box	ball	tray	to_go_cup
cup?	ball	plastic_box	tray	paper_box	plate
sponge?	tray	book	bowl	cup	sponge
ball?	paper_box	sponge	cup	to_go_cup	tray
book?	metal_box	bowl	to_go_cup	plate	paper_box

Table 4.3: 11 queries generated based on 11 object categories. Each query consists of a missing tool and five available objects

Result - Experts Selection

Table 4.4 and Table 4.5 list down the frequency distribution of experts' selections of substitutes for each scenario based on the object categories from Washington Dataset and RoCS Dataset respectively. When observed closely, the distribution listed in the tables can be roughly divided into two kinds of selections: *consensus towards no-selection* (highlighted by orchid colored cells) and *consensus towards selection* (highlighted by green and pink colored cells). In order to understand the distribution, the experts were asked to provide insights into the reasoning that went behind the selections. In the following, we provide the reasoning offered by the experts for their selections.

About consensus towards no-selection

One of the interesting insights we received from the experts are related to *consensus towards no-selection*. In the following, we summarize their reasoning behind their decisions. As specified in Table 4.4, for a ball, whose designated purpose was given as *to-play-with*, the majority of the experts who did not select any substitute had concerns about damaging a substitute if they use it as a ball. In case of the experts who selected a substitute, they assumed that a substitute would be made of a material such as a paper or plastic that would not be damaged when used as a ball. For a flashlight, whose designated purpose was given as *to-illuminate*, we got surprising results. We assumed that none of the experts would select any substitute, however, five experts selected a water bottle as a substitute while the rest (8 experts) did not select any. When the experts who selected a water bottle were asked about the reasoning, they stated that they would put a tiny lamp inside an empty water bottle. They justified this use by stating that since a lamp resides inside the body of a flashlight and as a result, the body of the flashlight protects the user from the hot lamp, it is reasonable to use the water bottle as a body inside which a small lamp can be placed. In other terms, the five experts imagined manufacturing of a substitute and a water bottle as a constituent. For a hand towel and Kleenex, the designated purpose was given as *to-clean*. In both the scenarios, the overwhelming majority of the experts did not select any substitute as they did not think that any choice would fulfill the designated purpose. Besides, even though the designated purpose was stated as *to-clean*, these experts extended it further as *to dry hands* in case of a hand towel and *to wipe dirty hands* in case of the Kleenex which according to them affected their selection process.

Missing Tool	Available Objects					None
ball? *Experts*	coffee_mug 2	food_cup 1	cereal_box 1	keyboard 0	flashlight 2	None 7
binder? *Experts*	flashlight 1	coffee_mug 0	notebook 9	water_bottle 0	bowl 0	None 3
bowl? *Experts*	hand_towel 0	ball 0	shampoo 0	pitcher 10	soda_can 2	None 1
cap? *Experts*	bowl 5	food_jar 2	food_box 0	coffee_mug 0	notebook 2	None 4
cereal_box? *Experts*	coffee_mug 0	food_cup 2	ball 0	flashlight 0	food_jar 8	None 3
coffee_mug? *Experts*	flashlight 0	food_can 1	keyboard 0	notebook 0	bowl 11	None 1
flashlight? *Experts*	food_box 0	food_cup 0	ball 0	water_bottle 5	plate 0	None 8
food_bag? *Experts*	food_box 11	hand_towel 1	flashlight 0	coffee_mug 0	notebook 0	None 1
food_box? *Experts*	food_jar 4	food_cup 2	soda_can 0	kleenex 0	cereal_box 6	None 1
food_can? *Experts*	flashlight 1	cereal_box 2	food_cup 4	food_box 4	cap 0	None 2
food_cup? *Experts*	keyboard 0	pitcher 5	plate 3	soda_can 3	sponge 0	None 2
food_jar? *Experts*	food_cup 8	flashlight 0	notebook 0	coffee_mug 3	soda_can 1	None 1
hand_towel? *Experts*	food_cup 0	plate 2	shampoo 0	food_can 0	flashlight 0	None 11
keyboard? *Experts*	bowl 0	cereal_box 1	food_can 0	notebook 7	food_box 0	None 5
kleenex? *Experts*	cap 1	water_bottle 1	ball 0	shampoo 1	flashlight 0	None 10
notebook? *Experts*	ball 0	water_bottle 0	plate 6	bowl 0	hand_towel 1	None 6
pitcher? *Experts*	plate 2	hand_towel 0	cereal_box 4	ball 0	flashlight 1	None 6
plate? *Experts*	coffee_mug 1	food_box 7	kleenex 0	pitcher 1	water_bottle 0	None 4
shampoo? *Experts*	food_can 3	food_cup 0	pitcher 1	flashlight 1	food_bag 0	None 8
soda_can? *Experts*	ball 0	shampoo 2	food_box 4	flashlight 1	food_bag 2	None 4
sponge? *Experts*	keyboard 0	coffee_mug 0	bowl 0	flashlight 0	hand_towel 12	None 1
water_bottle? *Experts*	bowl 1	cereal_box 0	notebook 0	sponge 0	soda_can 11	None 1

Table 4.4: The 22 queries generated based on 22 object categories from Washington Data set. Each query consists of a missing tool and five available objects. Below each query, the frequency distribution of 13 experts answers are provided. The last column represents a *None* option when an expert does not select any substitute from the available objects

Missing Tool	Available Objects					None
plastic_box? *Experts*	metal_box 17	bowl 2	tray 1	plate 0	sponge 0	None 1
bowl? *Experts*	plastic_box 15	sponge 0	cup 1	to_go_cup 0	plate 5	None 0
to_go_cup? *Experts*	tray 0	cup 20	book 0	paper_box 0	plastic_box 1	None 0
paper_box? *Experts*	plastic_box 19	to_go_cup 1	plate 1	ball 0	book 0	None 0
metal_box? *Experts*	bowl 0	sponge 0	cup 0	book 0	plastic_box 21	None 0
tray? *Experts*	cup 0	plastic_box 1	book 0	sponge 0	plate 20	None 0
plate? *Experts*	book 8	metal_box 0	ball 0	tray 12	to_go_cup 0	None 1
cup? *Experts*	ball 0	plastic_box 14	tray 0	paper_box 0	plate 1	None 6
sponge? *Experts*	tray 0	book 2	bowl 0	cup 0	sponge 19	None 0
ball? *Experts*	paper_box 2	sponge 11	cup 0	to_go_cup 2	tray 0	None 6
book? *Experts*	metal_box 3	bowl 0	to_go_cup 1	plate 2	paper_box 1	None 14

Table 4.5: 11 queries generated based on 11 object categories from RoCS Data set. Each query consists of a missing tool and five available objects. Below each query, the frequency distribution of 13 experts answers are provided. The last column represents a *None* option when an expert does not select any substitute from the available objects

In case of the hand towel, the two experts who selected a plate assumed it to be a paper plate which can be used to dry hands. Interestingly, one expert assumed the hand towel is made up of a paper which allowed him to select the plate which is also made up (according to him) of a paper. As for the Kleenex, one expert who selected a cap as a substitute stated that she can use a cap to clean since it is made of a garment. The expert who selected a water bottle assumed that it contains water and thus can be used to clean, while the expert who selected a shampoo stated that a shampoo can be used to clean. In case of a shampoo, whose designated purpose was stated as *to-wash-hair*, the majority of the experts who did not select any substitute stated that none of the available objects can be used to wash hair. Interestingly the five experts who did select a substitute, focused on the container into which a shampoo is usually stored. As a result, the substitute selection by them was focused on an object that can be used as a container. The expert who selected a flashlight stated that by removing all the components inside the flashlight, the empty body of the flashlight can be used as a container. When the five experts were questioned about a food cup and a food bag as they both are able to contain, they mentioned that the material of both the objects played a role in their decision. In case of the distribution given in the Table 4.5, a scenario involving a book has a majority of the experts rejecting all the options. The designated purpose of a book was stated as *to-read-from* and to our surprise the seven experts who selected substitutes stated that their substitute will contain some text on their outer surface which can be read as well.

About consensus towards selection

In case of the consensus towards selection of substitute, roughly two kinds of selections can be observed: 1) the majority of the experts converge on a specific substitute and 2) lack of consensus in substitute selection. Table 4.4 reflects both kinds of selection, while Table 4.5 reflects only the convergence on a specific substitute. There were some interesting insights we received from the experts about their selection reasoning. In case of a binder, for instance, whose designated purpose was given as *to-file-papers*, the majority of the experts selected a notebook. When we enquired about it, they stated that since the notebook is also capable of storing data, one can use the electronic formats of the papers to store them in the notebook. In other words, they stated that they did not see a notebook as a substitute but an improved version of storing equipment. While the three experts who did not select any substitute, did not think that an adequate choice was available to store the papers. When we asked to provide a possible example of an adequate choice, they stated: a bag or a box. For a bowl, a coffee mug, a food bag, a food jar, a sponge, and a water bottle, the experts who selected a substitute in overwhelming numbers stated that it was an easy choice from the given available objects. The similar reason was provided by the experts in case of Table 4.5. For the same set of tools, the experts who did not select any substitute stated that the size of the possible substitute primarily affected their decision: a substitute is either small compared to the missing tool or big. In many instances, especially in the case of *lack of consensus*, the experts mentioned that while there were more than one choice for a substitute available, their preference drove the selection decision. For instance, in case of a cap in Table 4.4 whose designated purpose was given as *to-cover-head*, the experts who selected a notebook agreed that a bowl or a food jar could also have been a possible choice, however, they preferred notebook since it covers the head better than a bowl or a food jar.

We also noted some additional insights concerning the objects which appear in both the tables. In case of a ball, the majority of the experts in case of the Washington Dataset table did not select a substitute while in case of the RoCS Dataset table the majority of the experts selected a sponge. The experts who selected a paper box and a to-go-cup in the second table stated that the material was a primary reason to select them. The number of experts who did not select any substitute in both the tables is almost the same. However, in case of the second table, the experts stated that none of the available objects could be use for playing purpose as opposed to the experts in the first table who stated that they did not select any objects due to the fear of damaging the objects. In case of a bowl, in both the tables, the experts did select a substitute in overwhelming numbers, while the number of experts who did not select any substitute in both the tables are almost none. A cereal box in the first table, which is typically made up of a paper, is equivalent to a paper box in the second table. In both the tables, majority of the experts converge on a specific substitute. A coffee mug in the first table is equivalent to a to-go-cup and a cup in the second table. In all the three scenarios, the majority of the experts converge on a single substitute. In case of a coffee mug and a to-go-cup, only one expert did not select any substitute, however, in case of a cup, six experts did not select any substitute. Their primary reason was the size of a plastic box even though they agreed that it could have been a plausible substitute. We also asked the experts who selected a plastic box about not selecting a paper box. They stated that since it is made up of a paper, it would not have been an ideal material to hold liquid. The similar reason was given by those who did not select any substitute. On the other hand, when we asked the experts except the one who selected a plate, about selecting a plate, they stated that although it would have been a plausible substitute, the plate would not be easy to lift with drinks in it. In other words, the plate would not be easy to use for drinking liquid from it. In case of a plate in both the tables, the majority of the experts agree on a single substitute. Interestingly though, the majority of the experts in the first table selected a paper box, none of the experts in the second table selected a metal box. Instead a book was selected which is made up of the same material as a paper box. When we asked the experts who selected a paper box or a book about this, we were surprised that they all stated that they felt safer to use a paper than a metal for food. In case of a sponge in both the tables, whose designated purpose was given as to-clean, almost all of the experts selected a specific substitute. One of the available objects in the second table for a sponge included a sponge and surprisingly, even there two experts diverged from the majority by selecting a book. When we asked the two experts about their choice, they stated they would use a page from the book to clean. When we asked them why did not they select a sponge, they clarified that since the task was to select a substitute, they went for a book instead of a sponge.

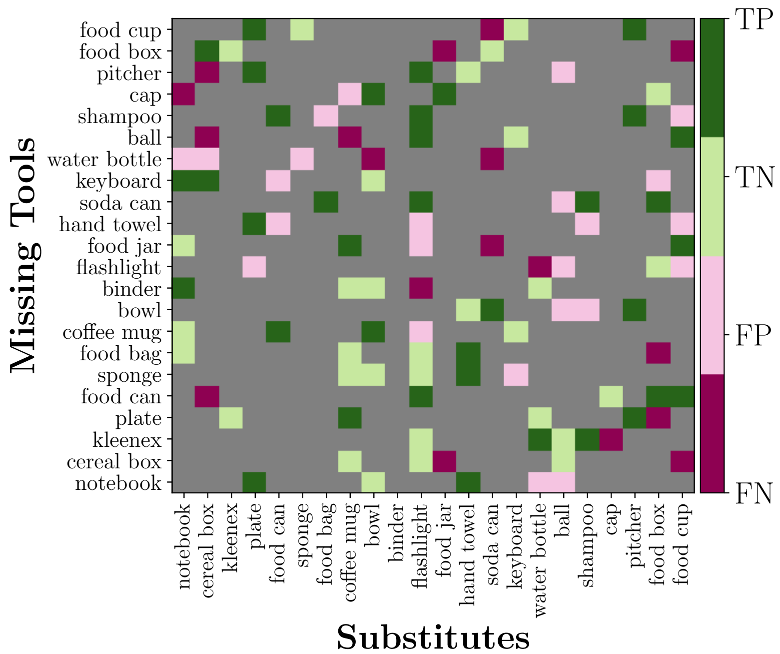
Summary

We can observe here that there is no specific pattern in the experts selection that can determine what drives an expert to select or not select an object as a possible substitute for a given missing tool. In some cases, the majority of the experts do not select any substitute, however there are still a minority who do select one. In some cases the majority of the experts select a specific substitute, however there are still a minority of the experts who do select a different substitute or do not select any substitute. And in other cases, there is no consensus among the experts in their selection choices: experts will select different objects from each other or will not even select any. In the cases where a minority of the experts makes a different decision than the majority, one can not simply invalidate their selection choice. This became apparent when we asked them about their choice and showed them that the majority had selected different object, they defended their selection and in some cases, such as a flash light or Kleenex, vehemently disagreed with the majority. It is also worth noting that when the majority converges on a specific selection or no-selection, the reasoning provided by the experts is related to the striking similarity or dissimilarity between a missing tool and a selected substitute with respect to their physical properties. However, in cases where there is no majority or where minority diverges from the majority, the reasoning is more related to the personal preferences. In either case, we noted that the knowledge used by the experts when they reason about their selection primarily consists of personal preferences (self-concept knowledge), common sense knowledge such as naive physics, causal relationships. In both the experiments: Washington Dataset based and RoCS Dataset based queries, we did not observe any specific factor that drives a decision to select or not select a substitute among the experts.

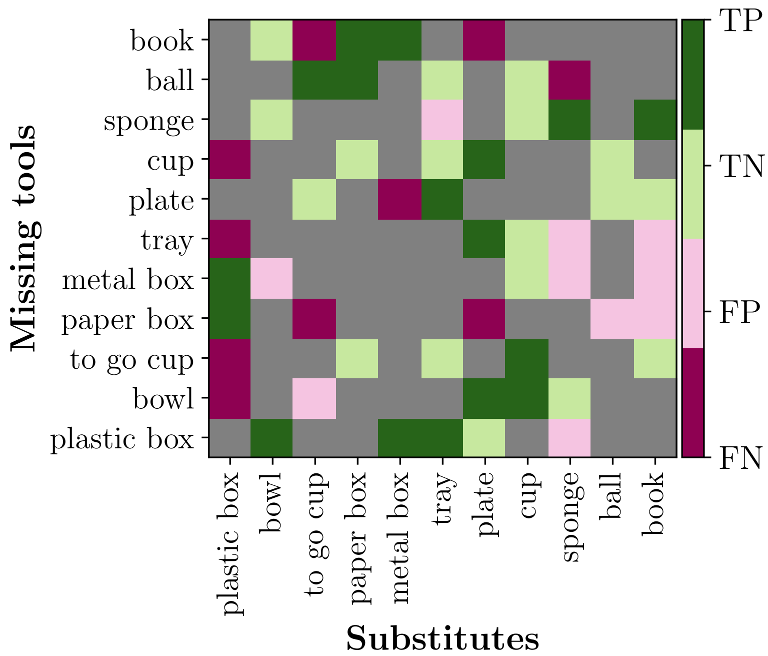
Result - ERSATZ Selection

As we have noted earlier, the objective of this experiment is two fold: 1) to examine the transferability of the optimized parameters on a different dataset. Therein, the transferability is examined by comparing the performance of ERSATZ on two knowledge bases generated from the two distinct properties measurements datasets; 2) to validate the substitute selections for various missing tool scenarios on two different datasets. For the validation assessment, we use the typical binary classification measures *True Positives*, *True Negatives*, *False Positives*, and *False Negatives*. We have seen earlier that the frequency distribution of substitute selection among the experts varies from scenario to scenario: in some cases a majority of the experts select one substitute, in some cases the majority do not select any substitute and in some cases no substitute is selected by majority. This makes it difficult to assess the validity of a substitute in every scenarios. As a result, our intent is to treat ERSATZ as an artificial expert and compare its selection with other experts' selection. We are primarily interested in comparing the selections where at least one expert selects the same substitute as ERSATZ. In this context, *True Positive* can be interpreted as at least one expert and ERSATZ selected a same substitute; *True Negative*: the experts and ERSATZ did not select an available object as a possible substitute; *False Positive*: the experts did not select a substitute but is selected by ERSATZ; *False Negative*: at least one expert selected a substitute but is not selected by ERSATZ. Note that True Positives and True Negatives demonstrate the consensus between ERSATZ and the experts. On the other hand, False Positive demonstrates the disagreement between ERSATZ and the experts. The interesting case is the False Negative, as it is the disagreement between not only ERSATZ and the expert/s but also between the expert/s and expert/s. In a sense, ERSATZ and the expert/s who did not select a substitute in the case of False Negatives agree with each other. As a result, False Negatives can not be considered in validating a substitute. For the validation purpose, we are primarily interested in the *True Positives* as it focuses on whether ERSATZ and at least one expert agree on a substitute.

In order to validate the substitutes, we have plotted heat plots to highlight the *True Positives*, *True Negatives*, *False Positives*, and *False Negatives*. As we are primarily interested in the *True Positives*, our discussion will follow accordingly. The heat plot in Fig. 4.15(a) illustrates the substitute selection by the 13 experts and ERSATZ respectively in 22 scenarios based on Washington Dataset. Similarly, the substitute selection by the 21 experts and ERSATZ in 11 scenarios based on RoCS Dataset are plotted as a heat map in Fig. 4.15(b). The grayed cells in the plots mean the corresponding object categories were not included in the available objects in the corresponding query. In Fig. 4.15(a), we can notice that out of 22 scenarios, ERSATZ selected True Positives in 19 scenarios. On the other hand, the second experiment in Fig. 4.15(b) had 11 scenarios where ERSATZ selected True Positives in all 11 scenarios. Additionally, we can also observe that in the first heat plot, out of 22 scenarios, ERSATZ and the experts did not select any substitute (True Negatives) in 18 scenarios. In the second heat plot, out of 11 scenarios, ERSATZ and the experts did not select any substitute (True Negatives) in 10 scenarios. In other terms, with the help of the parameters tuned on Washington Dataset, in 86% scenarios ERSATZ and at least one expert agreed on a substitute. These tuned parameters were used by ERSATZ on RoCS Dataset wherein in 100% scenarios ERSATZ and at least one expert agreed on a substitute. We see clearly that the performance of ERSATZ selection improved on RoCS Dataset based scenarios where RoCS Dataset was generated solely using machine-centric property estimation methods proposed in this work. These substitution selection results allow us to infer that the parameters tuned on the Washington-Dataset (containing 692 data points) were successfully transferred on a real-world dataset (containing 110 data points).



(a) ERSATZ Selection on Washington Dataset



(b) ERSATZ selection on RoCS Dataset

Figure 4.15: ERSATZ Performance: the distribution of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) in each substitution scenario using (a) Washington dataset and (b) RoCS dataset.

4.3.3 Similarity Comparison - ERSATZ vs WordNet

We have hypothesized that the distinction between a tool and a substitute is important when we determine similarity between them. For instance, if an object *A* can be replaced by an object *B* due to the desired similarity then *B* can be seen as a substitute for a tool *A*. However, the vice-versa need not be true, that is, an object *B* can not necessarily be replaced by an object *A*. Within the context of a designated purpose, the substitutability relationship between a tool and a substitute is symmetric, for instance, *for hammering*, a hammer can be replaced by a heeled shoe and vice versa. However, it is not the case once you step outside the context, for instance, a hammer can not be used as a heeled shoe for walking. In this experiment we examine this very aspect and in order to accomplish that we have pitted our proposed knowledge about objects and our proposed similarity computation against WordNet [119], a large lexical database and its associated similarity measures. WordNet consists of nouns, verbs, adverbs and adjectives which are grouped together in sets called synsets, where a typical synset would consist of synonymous words that represents a specific concept. The primary objective of this experiment is to substantiate the distinction between a tool and a substitute. In that regard, in this experiment, we compare the similarity among different objects determined by the similarity measures used in WordNet, and relevant-properties driven Jaccard Index-based similarity proposed in this work.

Experimental setup

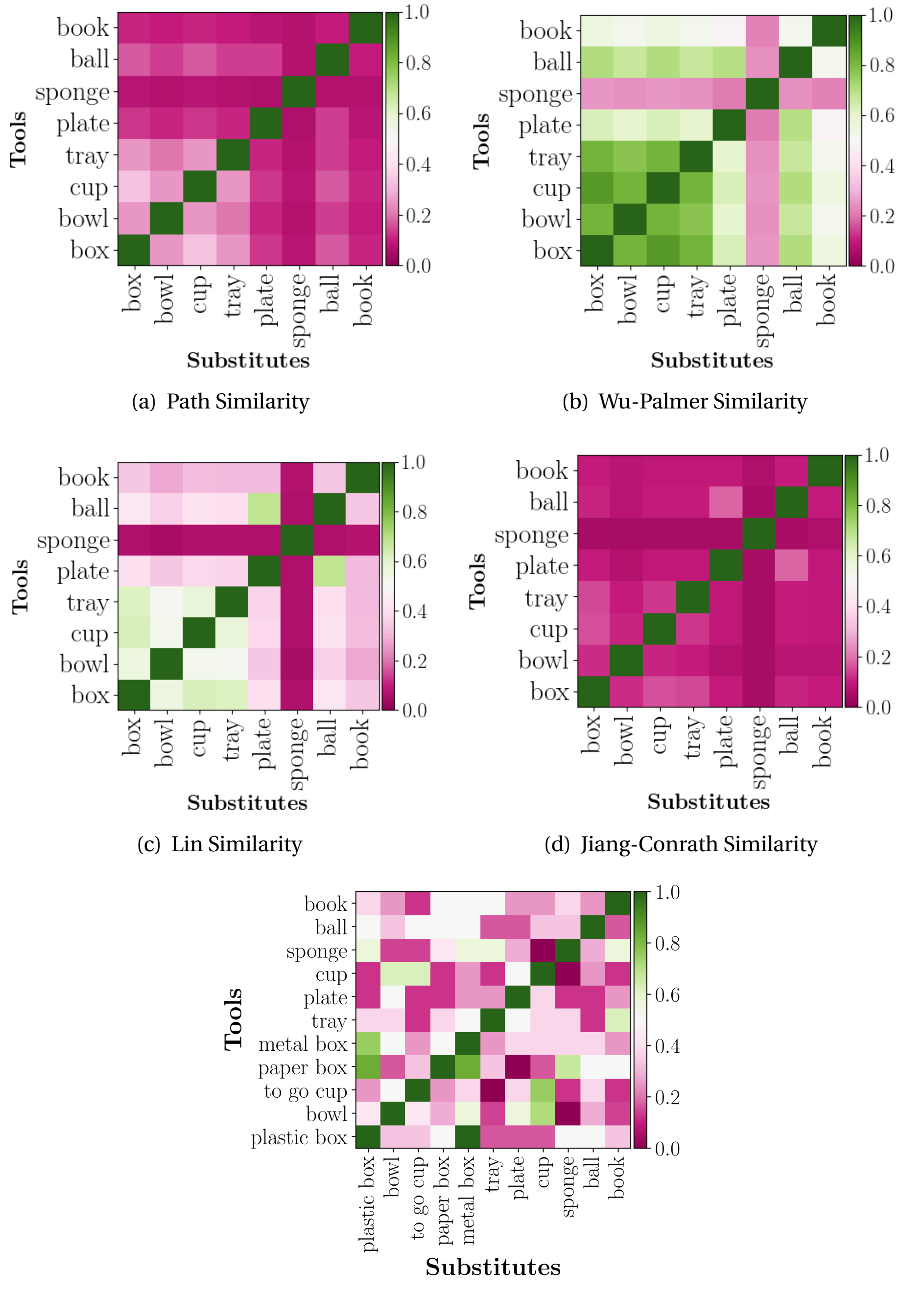
The similarity measures used in WordNet are typically meant to determine semantic similarity between two words or two sentences. In WordNet there are two types of similarity measures: *path-length based measure* and *information content based measure*. In the path-based similarity measure, WordNet is viewed as an undirected graph and the similarity between two concepts is computed by measuring the distance between them [130]. As stated by P. Resnik in [131]: "the shorter the path from one node to another, the more similar they are". On the other hand, the information content based similarity measure between the two concepts focuses on the contents of the information shared by both the concepts. According to

P. Resnik in [131]: "the more information two concepts share in common, the more similar they are".

For this experiment, for the path-length based similarity measure, we considered *Path Similarity* [132] and *Wu-Palmer Similarity* [133] and for information content based measures, we considered *Lin Similarity* [134] and *Jiang-Conrath Similarity* [135]. We used the WordNet interface provided by the Natural Language Toolkit (NLTK) platform [136] wherein an interface to compute the aforementioned similarities is provided. We used the object labels from the RoCS Dataset, however some labels were adapted while using WordNet to compute the similarity. For instance, in the RoCS Dataset we have a to-go-cup and a cup, however in WordNet there is no to-go-cup, therefore we use only a cup. Similarly, as WordNet does not differentiate between a plastic, a metal and a cardboard box, we considered only box for WordNet comparison. After adapting the labels, we had altogether 8 object labels for computing similarities in WordNet. We used RoCS Dataset to generate the knowledge base and used the 11 object labels from the dataset to computer the similarities between them. The similarities were computed for each pair of object labels for each similarity label.

Result

The similarity between each pair of object labels using aforementioned similarity measures was expressed as heat plots shown in the Fig. 4.16 where each color coded cell represents the similarity value between 0 and 1. The closer the value is to 1, the more similar are the objects. In the figure, the first row consists of the heat plots of path-based similarity computations, the second row consists of information-content based similarity computations and the last row illustrates the relevant-properties based Jaccard's Similarity computation proposed in this work. What interesting is, each WordNet-based similarity measures produces different similarity values for the same pair except any object that is paired with a sponge. In *Path Similarity*, the objects tray, cup, bowl and box when paired among each other are more similar to each other than when paired with rest of the objects or when rest of the objects paired among each other. A similar pattern can be observed in *Wu-Palmer Similarity* and *Lin Similarity*.



(a) Path Similarity

(b) Wu-Palmer Similarity

(c) Lin Similarity

(d) Jiang-Conrath Similarity

(e) Jaccard Similarity

Figure 4.16: The similarity comparison between the Path Similarity, Wu-Palmer Similarity, Lin Similarity and Jiang-Conrath Similarity measure used in WordNet and relevant property driven Jaccard based similarity measure.

While in *Jiang-Conrath Similarity* all of the objects are less similar to each other, in *Wu-Palmer Similarity* all of the objects are more similar to each other than in the rest of the WordNet-based similarity measures.

As can be observed in the figure, the resulting heat plots of the similarity between different objects computed using WordNet-based similarity measures is symmetric in nature while the relevant property driven Jaccard's Similarity based heat plot is non-symmetric. By symmetric in nature we mean that the similarity between objects, say A and B is same as the similarity between B and A. This is visible in all the WordNet-based similarity computations in the figure. In contrast, the similarity between objects being non-symmetric means the similarity between, say, objects A and B, and, B and A may not be same which is visible in relevant properties based Jaccard's similarity computation. This discrepancy is caused by the way objects are treated by WordNet and ERSATZ. WordNet does not make any distinction between a tool and a substitute. For WordNet, two object labels are two concepts in its database. Therefore when the similarities between, say objects A and B, and between B and A are computed the contents (path-length or information content) considered during the computation remain unchanged. Note that, in path-based similarity measure, the similarity is the distance between concepts A and B in an *undirected* graph. On the other hand, in information-content based similarity measure, the similarity is the information contents shared by the concepts A and B. In both the cases, the direction does not play any role, as in, there is no distinction between path from A to B and from B to A. On the contrary, our proposed approach distinguishes between a tool and a substitute. As a consequence the contents considered to determine a similarity between a tool A and a substitute B differs from the contents considered when computing a similarity between a tool B and a substitute A. In other terms, in order to compute the similarity between objects A and B, in our approach, the direction matters and therefore the similarity computed between two objects is not symmetric. As discussed in Sec.4.1.2, when an object is a tool, its relevant properties are first identified and the similarity with a possible substitute is determined on the basis of the relevant properties of a tool and the representative properties of a substitute. Consequently, the properties used to determine similarity when an object A is a tool can be different from the properties when the object A is a possible substitute.

Such non-symmetric relation is a necessity in the tool substitutions since it can not be assumed that if A is a substitute of a tool B, then B is a substitute of a tool A. Such assumption due to the symmetric relation may lead to an inadequate selection of a substitute. For instance, if we use WordNet as a knowledge source about objects and use *Wu-Palmer Similarity* measure to determine similarity for substitute selection, then according to the similarity computation a hammer and a shoe will be substitute of each other.

4.4 Related Work

We noted in the literature on substitute selection that, similar to our approach, a substitute for a missing tool is determined by means of knowledge about object, and the knowledge-driven similarity between a missing tool prototype and a potential substitute. In the following, we have summarized the approach proposed in each related work. We conclude the section by providing the insights we have drawn from the related work and the differences we have noted with our proposed approach.

The research work discussed in [62] proposes a neural model which is trained by user demonstrations where the primary focus is on a robot learning common sense knowledge about using a tool in a task instructed by human teachers. The proposed approach is provided with metric data about position, orientation, size; semantic relations such as On top, Inside, Connected to, Near; and symbolic knowledge about hand-picked relations such as *similar-to* and *capable-of* extracted from ConceptNet. The approach is trained for eight household tasks involving household objects and eight factory tasks involving factory objects. One of the problems the approach deals with is to find an alternative tool when the available tool is missing. The notable difference to our approach is that they do not use a substitute as defined by our approach. An alternative tool in their work still has a similar purpose as the missing tool and such similarity is determined by *similar-to* and *capable-of* relations. Some of the notable examples of alternate tools stated in the paper are: a ladder instead of a stool for elevating oneself, a box instead of a tray for transporting objects.

In [137], supervised learning with dual neural network based approach is proposed which uses shape and material similarity to determine a substitute. As specified in [137], "dual neural networks consist of two identical networks, each accepting a different input, combined at the end with a distance metric." The proposed approach trains shape based model and material based model separately. The models are trained for six actions where each action is trained on a separate network. For training the shape based models, the network is provided with positive and negative pairing of tools based on their shapes which are created randomly. The positive pairing refers to the applicability of both the tools for a given action and the negative pairing refers to the opposite. For training the material based models, five materials are considered and for each action, the network is provided with a set of materials from which an ideal tool is created. The similarity is determined on the basis of shape or material or both. In this work, a substitute is required to have a similar shape and material to the missing tool as opposed to our approach.

The approach proposed in [63] learns a visual predictive model which is trained using visual data collected from the demonstrations via kinesthetic teaching by humans and multi-object interactions with diverse objects. The main goal is to use such predictive models to perform the tasks involving previously unseen tools. The approach focuses on the tasks related to sweeping, wiping, and hooking. The notable difference to our approach is that this work does not perform substitute selection. Instead the object that is to be used in a unconventional manner (specified as unseen object) is already given, the proposed approach using the learned visual predictive models determines how to use it. In other words, our focus is on how to select a substitute as opposed to how to use it.

The approach proposed in [64] learns a model that has twofold objectives: 1) to assess if an object is a possible substitute for a missing tool on the basis of a score for effectiveness of it in the task; 2) to provide cues for manipulation which includes the geometric models related to the positioning and orientation of grasping and end effector. In this approach a tool is represented in terms of 21 parameters which consist of geometric models of a tool related to grasping and action; the relationship between the features which enable the applicability of the tool in a task; moment of inertia and mass; and orientation and positioning of a tool in a task. Additionally, the training data also includes a human-labeled affordance score where the score label indicates how good the tool is in a task. A machine learning technique Gaussian Process Regression is used to train a task function using the aforementioned parameters for learning a score for the effectiveness of a tool in the given task. There are four tasks considered in this work: hammering nail; lifting pancake; rolling dough; cutting lasagne; and scooping grains and for each task, the relationship between the features is learned during a tool use in the task in a simulation. When selecting a substitute, the approach looks for a candidate that has similar features. The proposed system does not perform any selection, instead the model is tested on three vision based datasets about objects where each object is assessed for its *substitutability* for a missing tool in a given task.

The approach discussed in [74] is based on random forest, a supervised learning based classifier, to classify whether an object is a valid substitute in the given task. The approach uses WordNet and ConceptNet to extract the potential candidates if they share the same parent with a missing tool for the *predetermined* relations: *has-property*, *capable-of* and *used-for*. The similarity between each candidate and a missing tool for a given task is calculated using three similarity metrics: WordNet path similarity, Divisi pairwise similarity, and Semantic Similarity Engine's analogical similarity. To train the classifier, each candidate is labeled by two experts as suitable or unsuitable substitutes for the missing tool in the given task. While in total, nine tasks were considered, it should be noted that the approach uses all the objects given in the ConceptNet for candidate extraction. However, the objects space can be reduced based on the objects in the environment a robot has access to.

The work discussed in [65] proposes a vision based estimation of affordances in the objects which is used to determine a substitute for a missing tool. The substitute is determined on basis of desired affordances shared by a missing tool where desired affordances in a substitute are estimated along with a confidence value. The approach uses object-wise global features and a multi-label learning method called JointSVM. For training, three different benchmark datasets containing point clouds of 1) 85 objects 2) 125 objects 3) 100 objects were used where each object is labeled with 12 affordances. The training is performed on each dataset separately as well as by combining all three datasets. The motivation is to examine which dataset is suitable for tool substitution, and whether training with more data leads to the improved performance and reliability of the prediction confidence.

The approach in [75] follows a similar modular based system as our proposed approach where knowledge acquisition is decoupled from substitute selection. For substitute selection, the approach makes use of an external relational database called ROAR (Repository of objects & attributes with roles). ROAR contains object labels and its associated affordances which are hand coded or can also be created from experience. In order to determine the similarity between a missing tool and an available object, the approach extracts all the affordances associated with the available object and check if these affordances are similar to those of the missing tool. In [76], a substitute for a missing tool is selected in a similar manner as above. The knowledge about objects is modeled manually after the dictionary definitions of the objects. The knowledge consists of object labels, their affordances, and inheritance and equivalence relations among the objects. A substitute for a missing tool is inferred on the basis of inheritance and equivalence relations between the substitute and the missing tool.

Most of the approaches discussed above are data-intensive and machine learning based. Consequently the approaches require training of a model which in turn need training examples. These training examples are needed to be substantial in numbers and also require features which are carefully selected. Note that any additional feature would require re-training of the models which would add additional computational efforts. In case of substitute selection, it is likely that new features will have to be added if the existing features would not suffice. All the approaches require labeling the training examples which would allow them to learn about what to expect about a potential substitute for a missing tool. The labeling of training examples takes different form in each approach: in some cases, objects are labeled with the associated affordances and in some cases they are labeled as valid or invalid substitute. Moreover, the models are trained for certain number of tasks. We have noted that none of the approaches is able to generalize outside of the tasks for which they are trained. The approaches that do not use machine learning, use carefully selected hand-coded knowledge about objects and matches a substitute on the basis of shared affordances or shared relations. In either case, human experts play a vital role in learning about a substitute or determining a substitute. Our approach, on the other hand, does not require any training since it is not based on any machine learning technique. It determines the relevant properties of a missing tool and selects a substitute on the basis of the relevant properties based similarity measure. This allows us to generalize the approach to other tasks as well.

5
Discussion

5.1 Recap

When a robot is operating in a dynamic environment, it can not be assumed that a particular tool required to solve a task will always be available. In such scenarios, capabilities are required to mitigate the consequences of the absence of a tool by finding an alternative as humans do. This skill is significant when operating in a dynamic, uncertain environment because it allows a robot to adapt to unforeseen situations. The question is: how can a robot determine which object in the environment is a viable candidate for a substitute? This is the challenge we have addressed in this thesis. Our thesis work on substitute selection is inspired by the way humans select a substitute and it led to the investigation of the following research questions:

1. What is the nature of the conceptual knowledge about objects desired in substitute selection?
2. How to acquire such conceptual knowledge?
3. What is a substitute?
4. How to determine a substitute for a missing tool?

We regarded the research questions (1) and (3) as conceptual questions while the questions (2) and (4) as procedural questions.

Tool use or tool substitution in robots is relatively a new research area in robotics compared to other research areas such robot vision, navigation or manipulation. As a result, we turned our attention to research on tool use in humans and animals for insights and cues as it is extensively investigated in cognitive science, psychology, neuroscience. Consequently, our research work has leaned on substantially on the theories and viewpoints provided by the literature on tool use in animals and humans and have been an inspiration behind the conceptual understanding of tool substitution for robots, substitute selection, various workflows presented in this work and proposed approaches. Our research has led to the following proposals with respect to the aforementioned research questions:

- A tool substitution is a highly complex, multi-layered integrated system which requires various functionalities in robotic system such as object perception, object grasping, object manipulation, localization, navigation, fault diagnosis and recovery, knowledge acquisition etc. to work together seamlessly (Sec. 4.1.3). It primarily consists of four processes: Search available objects, Select a substitute, Use a substitute, Validate a substitute where each process requires multiple inputs from various functionalities of a robot (Fig. 4.2).

- A tool is distinct from a substitute (Sec. 4.1.1). A tool has been defined in this work as *a tool is foremost a physical object and is manufactured artificially for a designated purpose.* In contrast, *a substitute is either a tool which is used for an unconventional purpose for which a conventional tool exists or it is a naturally occurring object.* This distinction is essential when we determine a suitability between them as it indicates that the suitability between two objects depends on which object is a tool and which is a substitute.

- The notion of the *relevant properties of a tool* plays a central role when selecting a substitute (Sec. 4.1.2). *The relevant properties of a tool in this work are considered as those properties which enable a tool's designated purpose.* As stated in [2], *a characteristic of a well-designed tool is that it feels comfortable and balanced when held*, it is paramount that, when a tool is designed for its designated purpose, each relevant physical property needs to be present to a certain degree such that a user can use the tool comfortably. The similar notion needs to be followed when selecting a substitute. It means, a substitute should be selected not only on the basis of the relevant physical properties of a tool but also the degree with which they are expected to be present in a substitute.

- For substitute selection, three kinds of knowledge are desired: *conceptual knowledge about objects* which include knowledge concerning objects such as physical and functional properties of objects, temporal and spatial properties of object; *self-concept knowledge* which focuses on a user's knowledge about its own physical and perception capabilities; and *common-sense knowledge* which usually consists of naive physics based rules that are intuitive and commonly held (Sec. 4.1.3). These three kinds of knowledge influence the selection of a substitute from the available objects for a missing tool.

- The main constituents of conceptual knowledge about objects required for a substitute selection are physical and functional properties observed in objects (Sec. 3.2.1). The ideal characterization of such conceptual knowledge is knowledge that is *generalized* where knowledge is about an object class as opposed to an object instance; *relative* where knowledge about an object class is derived from its instances that have been encountered by a robot and is subject to change as more instances are encountered; *subjective* where knowledge is derived from a robot's sensory experiences and interaction with the object's instances as opposed to hand-coded by experts based on their experiences with the objects; *qualitative* where knowledge about objects is not merely represented in terms of their properties but also to what degree they are present in the objects.

- While commonsense knowledge is concerned with commonly known knowledge by most people, subjective knowledge is concerned with knowledge held by an individual (Sec. 3.2.2). We term such subjective knowledge as robot-centric which states that robot-centric knowledge should be acquired from a *first-person-perspective.* In order to capture the robot-centricness, the knowledge should be grounded in robot's own sensory perception of objects' properties [87] and therefore we propose that knowledge about properties should be generated from the sensory measurements of the properties.

We have proposed the following approaches which are built upon the aforementioned proposals:

- An approach to identify relevant properties of a tool (Chapter 4, Sec. 4.2)

- An approach to select a substitute for the missing tool using conceptual knowledge about objects on the basis of the relevant properties of the tool (Chapter 4, Sec. 4.2)

- An approach to generate and represent *generalized, relative, robot-centric* and *qualitative* conceptual knowledge about objects from the quantitative measurements of the physical and functional properties of objects (Chapter 3, Sec. 3.2.3, Sec. 3.3)

- An approach to estimate the measurements of physical and functional properties observed in the various instances of objects (Chapter 2, Sec.2.2)

In this work, we have presented a proof of concept of the proposed approaches. Our proof of concept includes an extensible property estimation framework called **Ro**bot-**C**entric data**S**et (RoCS) framework which consists of light-weight estimation methods requiring minimal experimental set-up to obtain the quantitative measurements of physical properties (rigidity, weight, etc.) and functional properties (containment, support, etc.) from household objects. We acquired a dataset of 110 household objects comprising six physical properties: hollowness, size, flatness, roughness, rigidity, heaviness and four functional properties: support, containment, movability, blockage. The experimental evaluation on the dataset have revealed the *stability* as well as the *inter-class generality* of the proposed object property estimation methods. To generate the conceptual knowledge about objects from the property measurements, we employed unsupervised clustering methods to transform quantitative measurements into qualitative measurements followed by a step where Bi-variate Joint Frequency Distributions and Sample Proportion was used to generate the desired conceptual knowledge about objects. For our substitute selection approach, termed as ERSATZ, we used Jaccard's Similarity to identify the relevant properties of a missing tool and to determine similarity between a potential substitute and a tool on the basis of the relevant properties.

5.2 ERSATZ Integration With Object Perception

We noted in Sec. 4.1.3, a tool substitution system requires integration of various functionalities working together seamlessly in real-time. In that regard, we proposed a workflow for a tool substitution system (see Fig.4.2) where we focused on the substitute selection process for our research work. In this experiment, we have taken a baby-step towards an integrated system where we integrated our substitute selection system with an object perception system as illustrated in Fig. 5.1. The substitute selection system, in this work, receives three kinds of input: label of a missing tool, labels of available objects, and knowledge about the available objects and the missing tool. The objective of this experiment is to receive labels of available objects from an object perception system and select a possible substitute among them for a given missing tool in real-time. We intend to examine whether such real-time execution of object perception system and the proposed substitute selection system is realizable.

Experimental Setup

The underlying object perception system mainly consists of an RGB-D based object localization [138], categorization system [129] and ERSATZ. Detected and categorized objects in form of object position and category label are populated and subsequently received by ERSATZ. We selected five object categories from RoCS dataset: paper box, cup, plate, to go cup and book. We used real objects belonging to these five categories and created in total 10 scenarios for object perception system. In each scenario, three objects were placed in the scene which were localized and subsequently categorized by the perception system. Upon categorization the system sends the labels of the categorized objects to ERSATZ. ERSATZ is given 10 tool missing scenarios as illustrated in Table 5.1 where each scenario consists of a missing tool label and three available objects' labels. ERSATZ sends the selected substitutes' labels and rejected substitutes' labels back to perception system which then displays the result on the monitor by highlighting the selected substitutes' labels in blue color and rejected substitutes' labels in red color along with their corresponding similarity values.

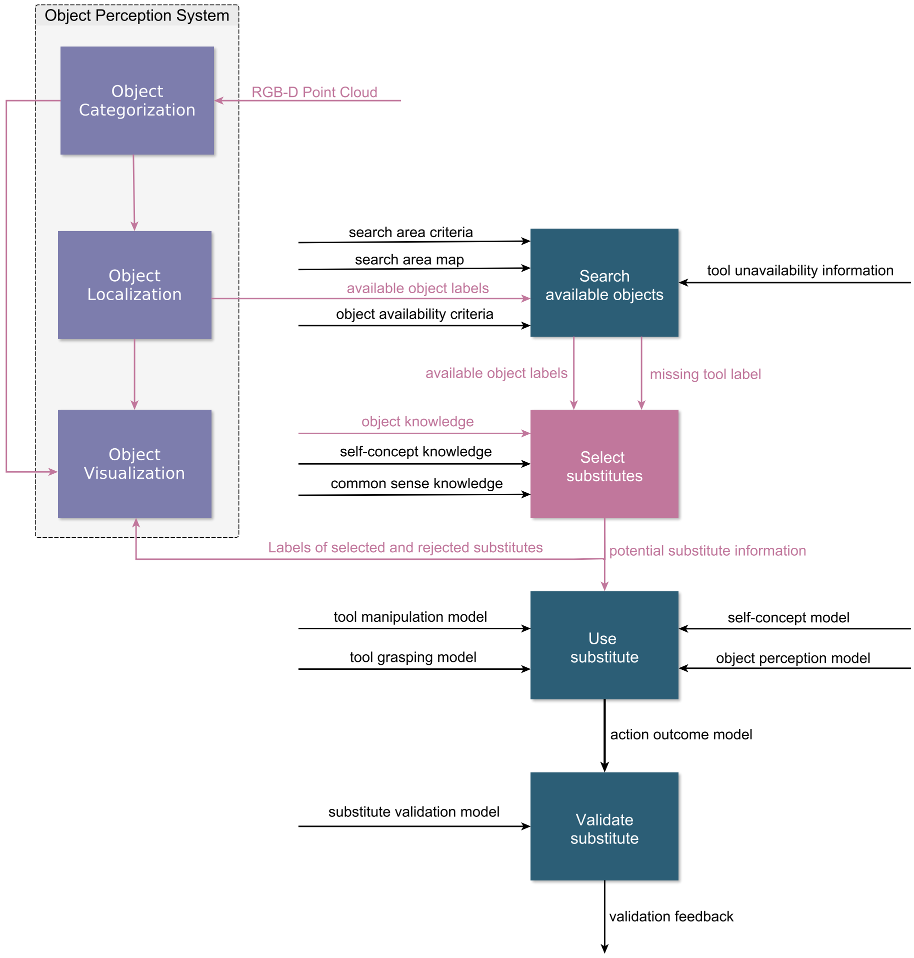


Figure 5.1: We have integrated an object perception system which accepts RGB-D point cloud of objects in the environment as inputs and outputs the labels of the categorized objects. These labels are forwarded to the substitute selection system ERSATZ which identifies a substitute among them for a given missing tool.

Table 5.1: 10 real-world scenarios.

#	Missing tool	Available objects (with similarity, r : indicates rejection)		
1	paper box	**paper box (1.0)**	to go cup (0.75)	book (r)
2	paper box	plate (r)	cup(0.625)	**to go cup** (0.75)
3	cup	**to go cup** (0.625)	plate (r)	**paper box** (0.625)
4	cup	paper box (0.625)	book(r)	**cup**(1.0)
5	plate	paper box (r)	to go cup (r)	**book**(0.75)
6	plate	cup (r)	book (0.75)	**plate**(1.0)
7	to go cup	plate(r)	**paper box**(0.857)	cup(0.714)
8	to go cup	**to go cup**(1.0)	paper box(0.857)	book(0.429)
9	book	**plate**(0.75)	to go cup (r)	paper box(r)
10	book	**book**(1.0)	plate(0.75)	cup(r)

Result

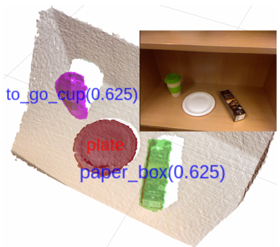
The outcome of the experiment is given in Table 5.1 where the selected substitutes along with their substitutes are highlighted by bold text. Additionally, the outcome is also highlighted by the perception system as illustrated in Fig. 5.2. For instance, a scenario shown in Fig. 5.2(e) in which ERSATZ has to select a substitute for a missing *plate* from the available objects in the environment. As a result three objects are detected in the environment and classified in the scene as *to go cup*, *book* and *paper box*. ERSATZ's responses show, in blue, the selected substitute, whereas in red, the rejections. ERSATZ correctly identifies the *book* as an optimal substitution, while *paper box* and *to go up* are rejected as substitutes. We can notice that in all of the scenarios where a missing tool is in the available objects, it was correctly selected. In general, it is observable that ERSATZ generally identifies reasonable substitutes along with their similarity value for a missing tool, in our experiment of 10 scenarios.

5.3 Open Questions

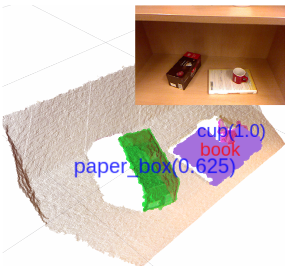
Our research in substitute selection as well as in tool substitution has made us realized that the seemingly simple sounding problem of finding a substitute for a missing tool and using it in the task is markedly a complex problem. Our research work has exposed myriad number of challenges, view points and open questions along the way, and not all can be listed here. In the following, we have listed open questions, that we deemed as significant, related to the factors we encountered during our research. It should be noted that this is a non-exhaustive list and the open questions are raised with a focus on robotic applications.

(a) scene #1

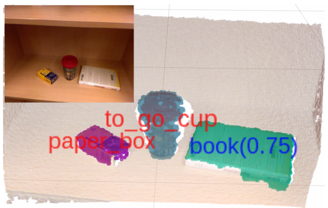
(b) scene #2



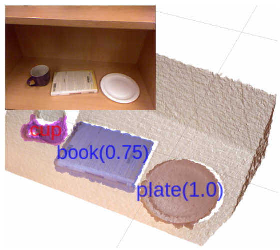
(c) scene #3



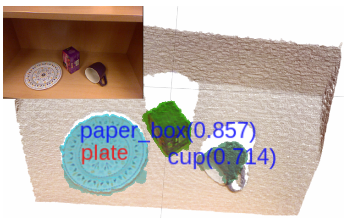
(d) scene #4



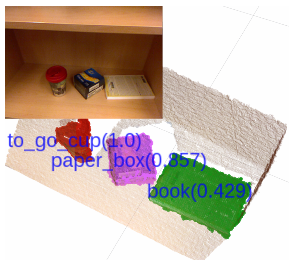
(e) scene #5



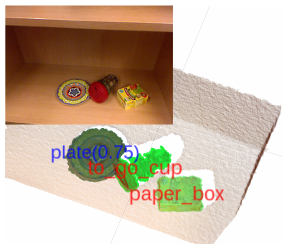
(f) scene #6



(g) scene #7



(h) scene #8



(i) scene #9

(j) scene #10

Figure 5.2: Corresponding scenes of scenarios shown in Table 5.1.

About substitute selection

We have stated in Chapter 4 that a substitute selection is a subjective choice. While we have identified three kinds of knowledge bases that influence the selection of a substitute, it is necessary to investigate further what are the other factors that may influence the selection. While in this work, we have selected a substitute in a non-invasive manner, an invasive method in a limited capacity would also be a possible approach. In that case, the challenge would be to devise an invasive method such that it can be used in multiple scenarios involving distinct tools. Additionally, defining the scope of the *limited capacity* in itself is a complicated task. Moreover, in this work, we have assumed a specific view on a substitute where an object is used as a substitute without altering its form. It would be worth investigating what are the other forms of substitutes such as binding multiple objects together in a certain manner or altering the object's physical form etc.

Conceptual knowledge about objects

Our thesis focuses on the conceptual knowledge that involves a generalized, relative, robot-centric and qualitative knowledge about physical and functional properties of objects. One of the challenge we faced in the early research was to determine the granularity of such knowledge. More specifically, the challenge was to determine how much detailed the knowledge should be and how does the detailedness affect a substitute selection. In the proposed approach the granularity of the knowledge is determined empirically and is uniform for all substitute selection scenarios. However, it is possible that different missing-tool scenarios may require varying degree of detailedness of knowledge. In that case the question is, how to determine the degree of detailedness in each missing-tool scenario. Besides physical and functional properties, another issue to consider is what other properties should be included in the conceptual knowledge, for instance, the knowledge about parts of the objects, relationship between the parts of the objects, spatial and temporal properties of the objects etc.

Role of self-concept and common sense knowledge

We noted in the Chapter 4, Sec. 4.1.3 the role of self-concept and the common sense knowledge, and the influence they exert over the selection. But the question is what constitutes self-concept knowledge. Ideally, the self-concept knowledge is about one's physical strengths as well as the limitations, perception capabilities, personal preferences etc. However, the challenge here is, how do we determine and formalize the contents that can represent these factors? Another issue is acquiring such knowledge, in other words, how does a robot acquire knowledge about itself? The same inquiry holds true for common sense knowledge or naive physics knowledge as it also plays role in rejecting the substitutes. We also believe that the selection takes place in multiple phases where the initial selection may be performed on the basis of conceptual knowledge about objects, while the rejection of the substitutes selected in the initial phase is performed on the basis of self-concept knowledge.

Robot-centric

We stand by our robot-centric aspect towards knowledge as given the varying capabilities of sensors and manipulation, we believe that relying on common sense knowledge bases such as WordNet and ConceptNet do not provide a multi-dimensional view of an object which is vital in the case of tool substitution. Therefore, more efforts are needed to devise approaches to acquire knowledge from a robot's perspective. Any additional knowledge that is needed, whether it is self-concept knowledge or naive physics knowledge or even more expressive conceptual knowledge involving spatial and temporal properties of objects or part-relationship in the objects, the challenge is how to capture the robot-centric aspect when acquiring such knowledge in a bottom-up fashion as proposed in this work.

Property estimation

One of the pressing questions we faced during the property estimation development was, assuming that an estimation framework supports *n* number of property estimation methods, is it necessary to estimate each property from each instance. For our work, we created a dataset of property measurements in a bulk, but when a robot is exploring the environment, it will not process all the objects in the environment at once. It will be a gradual exploration, and in that case, is it necessary to estimate every single property in an object and in what order the properties should be estimated in an object? The second issue is, what kind of properties should be considered? Our property estimation methods are superficial, in a sense that they do not separate different parts of the object. For instance, a plastic box is made up of a lid and a container OR that a plastic box can be viewed from different angles which may change the property estimation for certain properties as illustrated in the Fig. 5.3 while in our work, we considered only a single pose: a natural pose for an object.

Figure 5.3: A plastic box is viewed from different angles which may change the property estimations for properties such as size, shape, hollowness, support etc.

Summary

Tool substitution is a cognitively demanding activity which is distinct from a tool use and we propose that such differentiation needs to be highlighted as they both require different form of knowledge, representation and reasoning. We noted that, generalization is a central aspect of tool substitution. What is interesting to note is that while a tool can have multiple substitutes, an object can be a substitute for multiple tools. This is where generalization plays a key role as the system has to generalize over not only the physical structure of a tool but also how is it to be grasped, manipulated, and transfer that generalization to a substitute. A transfer of a grasping model or a manipulation model from a tool to a substitute without making any alterations will lead to a failure. The question is how to attain such generalization as any form of generalization in robotics is one of the open research questions and a solution still eludes the researchers. We also realized that there are valuable cues and theoretical knowledge offered in cognitive science, neuroscience, psychology on tool use in animals and humans, and robotic researchers should lean on these disciplines. Given the complexity of a tool substitution, we believe that tool substitution requires collective multidisciplinary efforts in order to perform a more theoretical research work to get a better conceptual understanding of the various processes and parameters that are involved. The insights we gained have made us realize that tool substitution in general and substitute selection in particular is not a trivial problem and developing such system for a robot would be a giant leap towards building an intelligent robot.

A
List of Figures

B

List of Tables

C

Knowledge Bases Links

In the following, we have provided the webpage links to the knowledge bases which we have reviewed in the Chapter 3, Sec. 3.4.2. We were unable to find any online presence for the following knowledge bases: OMICS, OMRKF and OUR-K.

1. **KnowRob:**
   `http://knowrob.org/`
2. **MLNKB:**
   `https://web.stanford.edu/~yukez/eccv2014.html`
3. **NMKB:**
   `https://tinyurl.com/y9uboh62`
4. **ORO:**
   `https://www.openrobots.org/wiki/oro-server`
5. **PEIS:**
   `http://www.aass.oru.se/Research/Robots/projects.html`
6. **Robobrain:**
   `http://robobrain.me/about.html`
7. **Robobrain Source code:**
   `https://github.com/RoboBrainCode`

D
Software Repository

All the three approaches: property estimation, knowledge generation and substitute selection were implemented in Python programming language. In the following we provide the git repository links to the source code, datasets, and knowledge bases for our approaches.

Source Code

1. **Property Estimation -** `https://gitlab.com/rock_paper_scissors/property_estimation`
2. **Knowledge Generation -** `https://gitlab.com/rock_paper_scissors/knowledge-generation`
3. **Substitute Selection -** `https://gitlab.com/rock_paper_scissors/substitute-selection`

Dataset

1. **RoCS Dataset -** `https://gitlab.com/rock_paper_scissors/dataset/-/tree/main/RoCS`
2. **Objects used in RoCS dataset -** `https://gitlab.com/rock_paper_scissors/dataset/-/tree/main/RoCS_Objects`
3. **Dataset using Washington Dataset -** `https://gitlab.com/rock_paper_scissors/dataset/-/tree/main/Washington_Dataset-based`

Knowledge base

1. **Based on Washington dataset -** `https://gitlab.com/rock_paper_scissors/knowledge-base/-/tree/main/Washingtondatasetbased`
2. **Based on RoCS dataset -** `https://gitlab.com/rock_paper_scissors/knowledge-base/-/tree/main/RoCSbased`

E
Bibliography

[1] C. Baber, “Introduction,” in *Cognition and Tool Use.* Taylor and Francis, 2003, ch. 1, pp. 1–15.

[2] ——, “The Design of Tools,” in *Cognition and Tool Use.* Taylor and Francis, 2003, ch. 6, pp. 69–80.

[3] ——, *Cognition and Tool Use*, 2003.

[4] C. M. Sanz, J. Call, and C. Boesch, Eds., *Tool Use in Animals: Cognition and Ecology.* Cambridge University Press, 2013.

[5] S. Butler, “Lord, What is a Man: Tools,” in *The Notebooks of Samuel Butler*, 10th ed., 2012, ch. 1, pp. 1–26.

[6] M. Thosar, C. A. Mueller, G. Jaeger, J. Schleiss, N. Pulugu, R. M. Chennaboina, S. V. Jeevangekar, A. Birk, M. Pfingsthorn, and S. Zug, “From Multi-modal Property Dataset to Robot-Centric Conceptual Knowledge About Household Objects,” *Frontiers in Robotics and AI - Computational Intelligence in Robotics, Special Issue: Robots that Learn and Reason: Towards Learning Logic Rules from Noisy Data (Under Review)*, vol. 8, p. 87, 2021.

[7] G. Jäger, C. A. Mueller, M. Thosar, S. Zug, and A. Birk, “Towards Robot-Centric Conceptual Knowledge Acquisition,” in *Robots that Learn and Reason Workshop in IEEE/RSJ International Conference on Intelligent Robots and Systems*, Madrid, 2018.

[8] M. Thosar, S. Zug, A. M. Skaria, and A. Jain, “A Review of Knowledge Bases for Service Robots in Household Environments,” in *6th Inter-*

national Workshop on Artificial Intelligence and Cognition, Palermo, Italy, 2018.

[9] M. Thosar, C. Mueller, S. Zug, and M. Pfingsthorn, "Towards a Prototypical Approach to Tool-Use Improvisation - Extended Abstract," in *International Conference on Autonomous Agents and Multiagent Systems*, Montreal, 2019.

[10] M. Thosar, C. Mueller, and S. Zug, "What Stands-in for a Missing Tool?: A Prototypical Grounded Knowledge-based Approach to Tool Substitution," in *11th International Workshop on Cognitive Robotics in 16th International Conference on Principles of Knowledge Representation and Reasoning*, Tempe, Arizona, 2018.

[11] M. Thosar, C. A. Mueller, G. Jäger, M. Pfingsthorn, M. Beetz, S. Zug, and T. Mossakowski, "Substitute Selection for a Missing Tool Using Robot-Centric Conceptual Knowledge Of Objects," in *Knowledge Representation and Reasoning Track in 35th ACM/SIGAPP Symposium On Applied Computing*, Brno, 2020.

[12] A. Stoytchev, "Robot tool behavior: A developmental approach to autonomous tool use," Ph.D. dissertation, College of Computing, Georgia Institute of Technology, 2007.

[13] S. Brown and C. Sammut, "A relational approach to tool-use learning in robots," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 7842 LNAI, no. September, pp. 1–15, 2013.

[14] V. Gallese and G. Lakoff, "The brain's concepts: The role of the sensory-motor system in conceptual knowledge," *Cognitive Neuropsychology*, vol. 22, no. 3-4, pp. 455–479, 2005.

[15] M. M. Louwerse and P. Jeuniaux, "The linguistic and embodied nature of conceptual processing," *Cognition*, vol. 114, no. 1, pp. 96–104, 2010.

[16] J. Feldman and S. Narayanan, "Embodied meaning in a neural theory of language," *Brain and Language*, vol. 89, no. 2, pp. 385–392, 2004.

[17] R. Pfeifer and J. Bongard, *How the Body Shapes the Way We Think - A New View of Intelligence Rolf*. Cambridge, Massachusetts: The MIT Press, 2007.

[18] ——, “Intelligence, Thinking, and Artificial Intelligence,” in *How the Body Shapes the Way We Think - A New View of Intelligence Rolf*. Cambridge, Massachusetts: The MIT Press, 2007, ch. 1, pp. 5–23.

[19] G. Metta, L. Natale, F. Nori, G. Sandini, D. Vernon, L. Fadiga, C. von Hofsten, K. Rosander, M. Lopes, J. Santos-Victor, A. Bernardino, and L. Montesano, “The iCub humanoid robot: An open-systems platform for research in cognitive development,” *Neural Networks*, vol. 23, no. 8, pp. 1125–1134, 2010. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0893608010001619

[20] “PR2 User Manual,” Willow Garage, Tech. Rep., 2012.

[21] A. van Casteren, “Tool Use: Crows Craft the Right Tool for the Job,” *Current Biology*, vol. 27, no. 24, pp. R1314–R1316, 2017. [Online]. Available: https://doi.org/10.1016/j.cub.2017.10.022

[22] H. M. Manrique, G. Sabbatini, J. Call, and E. Visalberghi, “Tool choice on the basis of rigidity in capuchin monkeys,” *Animal Cognition*, vol. 14, no. 6, pp. 775–786, 2011.

[23] H. M. Manrique, A. N. M. Gross, and J. Call, “Great Apes Select Tools on the Basis of Their Rigidity,” *Journal of Experimental Psychology: Animal Behavior Processes*, vol. 36, no. 4, pp. 409–422, 2010.

[24] E. Visalberghi, G. Sabbatini, A. H. Taylor, and G. R. Hunt, “Cognitive insights from tool use in nonhuman animals.” *APA handbook of comparative psychology: Perception, learning, and cognition.*, no. September 2018, pp. 673–701, 2017.

[25] A. Seed and R. Byrne, “Animal tool-use,” *Current Biology*, vol. 20, no. 23, pp. R1032–R1039, 2010. [Online]. Available: http://dx.doi.org/10.1016/j.cub.2010.09.042

[26] J. Wu, J. J. Lim, H. Zhang, J. B. Tenenbaum, and W. T. Freeman, “Learning Physical Object Properties from Unlabeled Videos,” in *British Machine Vision Conference*, 2016. [Online]. Available: https://jiajunwu.com/papers/physics_bmvc.pdf

[27] J. Sinapov, M. Wiemer, and A. Stoytchev, “Interactive Learning of the Acoustic Properties of Household Objects,” in *IEEE International Conference on Robotics and Automation*, 2009.

[28] A. J. Spiers, M. V. Liarokapis, B. Calli, and A. M. Dollar, “Single-Grasp Object Classification and Feature Extraction with Simple Robot Hands and Tactile Sensors,” *IEEE Transactions on Haptics*, vol. 9, no. 2, pp. 207–220, 2016.

[29] J. Kim, H. Lim, S. C. Ahn, and S. Lee, “RGBD Camera Based Material Recognition via Surface Roughness Estimation,” in *Proceedings - 2018 IEEE Winter Conference on Applications of Computer Vision, WACV 2018*, 2018.

[30] D. Kraft, N. Pugeault, E. Baseski, M. Popovic, D. Kragic, S. Kalkan, F. WÖRGÖTTER, and N. KRÜGER, “Erratum: "Birth of the Object: Detection of Objectness and Extraction of Object Shape Through Object–Action Complexes",” *International Journal of Humanoid Robotics*, vol. 06, no. 03, 2009.

[31] S. Takamuku, G. Gómez, K. Hosoda, and R. Pfeifer, “Haptic discrimination of material properties by a robotic hand,” in *2007 IEEE 6th International Conference on Development and Learning, ICDL*, 2007.

[32] K. Vaesen, “The cognitive bases of human tool use,” *Behavioral and Brain Sciences*, vol. 35, no. 04, pp. 203–218, 2012.

[33] D. Biro, M. Haslam, and C. Rutz, “Tool use as adaptation.” *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, vol. 368, no. 1630, 2013.

[34] J. Vauclair and J. A. Anderson, “Object Manipulation, Tool Use, and The Social Context in Human and Non-Human Primates,” *Techniques and Culture*, vol. 23-24, p. 121–136, 1994.

[35] T. Susi and T. Ziemke, “On the subject of objects: Four views on object perception and tool use,” *tripleC-Cognition, Communication, Co-operation*, vol. 3, no. 2, p. 6–19, 2005. [Online]. Available: http://triple-c.at/index.php/tripleC/article/viewArticle/19

[36] M. Hernik and G. Csibra, "Functional understanding facilitates learning about tools in human children," *Current Opinion in Neurobiology*, vol. 19, no. 1, pp. 34–38, 2009.

[37] R. Bischoff, U. Huggenberger, and E. Prassler, "Kuka youbot - a mobile manipulator for research and education," in *2011 IEEE International Conference on Robotics and Automation*, May 2011, pp. 1–4.

[38] D. M. Swoboda, "A comprehensive characterization of the asus xtion pro depth sensor," 2014.

[39] N. J. Emery, "Insight, imagination and invention: Tool understanding in a non-tool-using corvid," in *Tool Use in Animals: Cognition and Ecology*, C. M. Sanz, J. Call, and C. Boesch, Eds. Cambridge University Press, 2013, ch. 4, pp. 67–88.

[40] C. Boesch, "Ecology and cognition of tool use in chimpanzees," in *Tool Use in Animals: Cognition and Ecology*, J. B. C. Sanz, Crickette M. Call, Ed. Cambridge University Press, 2013, ch. 2, pp. 21–47.

[41] S. Garrido-Jurado, R. Munoz-Salinas, F. Madrid-Cuevas, and M. Marin-Jimenez, "Automatic generation and detection of highly reliable fiducial markers under occlusion," *Pattern Recognition*, vol. 47, no. 6, pp. 2280–2292, 2014.

[42] M. A. Fischler and R. C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 24, no. 6, pp. 381–395, 1981.

[43] J. J. Gibson, *The Theory of Affordances*. Psychology Press, Taylor & Francis Group, 1986.

[44] C. Baber, "Working With Tools," in *Cognition and Tool Use*. Taylor and Francis, 2003, ch. 5, pp. 51–68.

[45] R. Hartson, "Cognitive, physical, sensory, and functional affordances in interaction design," *Behaviour & Information Technology*, vol. 22, no. 5, pp. 315–338, 2003.

[46] F. Osiurak, Y. Rossetti, and A. Badets, “What is an affordance? 40 years later,” *Neuroscience and Biobehavioral Reviews*, vol. 77, no. August 2016, pp. 403–417, 2017.

[47] W. Kuhn, “An Image-Schematic Account of Spatial Categories,” *Spatial Information Theory*, pp. 152–168, 2007.

[48] M. Johnson, *The Body in the Mind. The Bodily Basis of Meaning, Imagination, and Reasoning.* The University of Chicago Press, 1987.

[49] G. Lakoff, *Women, Fire, and Dangerous Things. What Categories Reveal about the Mind.* The University of Chicago Press, 1987.

[50] G. Lakoff and R. Núñez, *Where Mathematics Comes from: How the Embodied Mind Brings Mathematics Into Being.* New York: Basic Books, 2000.

[51] B. Bennett and C. Cialone, “Corpus Guided Sense Cluster Analysis: a methodology for ontology development (with examples from the spatial domain),” in *8th International Conference on Formal Ontology in Information Systems (FOIS)*, ser. Frontiers in Artificial Intelligence and Applications, P. Garbacz and O. Kutz, Eds., vol. 267. IOS Press, 2014, pp. 213–226.

[52] J. M. Mandler and C. Pagen Canovas, “On defining image schemas,” *Language and Cognition*, vol. 6, no. 04, pp. 510–532, 2014.

[53] P. Gärdenfors, “Cognitive semantics and image schemas with embodied forces,” *Image (Rochester, N.Y.)*, pp. 1–16, 1987.

[54] L. Jamone, E. Ugur, A. Cangelosi, L. Fadiga, A. Bernardino, J. Piater, and J. Santos-Victor, “Affordances in Psychology, Neuroscience, and Robotics: A Survey,” *IEEE Transactions on Cognitive and Developmental Systems*, vol. 10, no. 1, pp. 4–25, 2016.

[55] H. Min, C. Yi, R. Luo, J. Zhu, and S. Bi, “Affordance research in developmental robotics: a survey,” *IEEE Transactions on Cognitive and Developmental Systems*, vol. 8, no. 4, pp. 237–255, 2016.

[56] A. Koubâa, *Robot operating system (ros): The complete reference.* Springer, 2017, vol. 2.

[57] A. B. Sharma, L. Golubchik, and R. Govindan, "Sensor faults: Detection methods and prevalence in real-world datasets," *ACM Transactions on Sensor Networks*, vol. 6, no. 3, 2010.

[58] J. Benesty, J. Chen, Y. Huang, and I. Cohen, "Pearson Correlation Coefficient," in *Noise Reduction in Speech Processing*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, pp. 1–4.

[59] N. Saito, K. Kim, S. Murata, T. Ogata, and S. Sugano, "Tool-Use Model Considering Tool Selection by a Robot Using Deep Learning," *IEEE-RAS International Conference on Humanoid Robots*, vol. 2018-Novem, pp. 814–819, 2019.

[60] Y. Wu, S. Kasewa, O. Groth, S. Salter, L. Sun, O. P. Jones, and I. Posner, "Imagine that! Leveraging emergent affordances for tool synthesis in reaching tasks," *arXiv*, pp. 1–12, 2019.

[61] V. Tikhanoff, U. Pattacini, L. Natale, and G. Metta, "Exploring affordances and tool use on the iCub," *IEEE-RAS International Conference on Humanoid Robots*, pp. 130–137, 2015.

[62] R. Bansal, S. Tuli, R. Paul, and Mausam, "TOOLNET: Using Commonsense Generalization for Predicting Tool Use for Robot Plan Synthesis," *arXiv*, 2020.

[63] A. Xie, F. Ebert, S. Levine, and C. Finn, "Improvisation through physical understanding: Using novel objects as tools with visual foresight," *arXiv*, 2019.

[64] P. Abelha and F. Guerin, "Learning how a tool affords by simulating 3D models from the web," *IEEE International Conference on Intelligent Robots and Systems*, vol. 2017-Septe, pp. 4923–4929, 2017.

[65] W. Mustafa, M. Wächter, S. Szedmak, A. Agostini, D. Kraft, T. Asfour, J. Piater, F. Wörgötter, and N. Krüger, "Affordance Estimation For Vision-Based Object Replacement on a Humanoid Robot," in *47th International Symposium on Robotics*, vol. 2016, 2016, pp. 164–172.

[66] M. Kaboli, D. Feng, and G. Cheng, "Active Tactile Transfer Learning for Object Discrimination in an Unstructured Environment using Multimodal Robotic Skin," *International Journal of Humanoid Robotics*, 2017.

[67] S. Brown and C. Sammut, "A Relational Approach to Tool-use Learning in Robots," *Inductive Logic Programming - 22nd International Conference*, pp. 1–15, 2012.

[68] K. Takahashi, T. Ogata, and H. Tjandra, "Tool – body Assimilation Model Based on Body Babbling and Neuro-dynamical System," *International Conference on Artificial Neural Networks*, 2014.

[69] J. Stückler and S. Behnke, "Adaptive Tool-Use Strategies for Anthropomorphic Service Robots," *14th IEEE-RAS International Conference on Humanoid Robots (Humanoids)*, no. November, 2014. [Online]. Available: http://www.ais.uni-bonn.de/papers/Humanoids_2014_Stueckler.pdf

[70] W. Li and M. Fritz, "Teaching robots the use of human tools from demonstration with non-dexterous end-effectors," *IEEE-RAS International Conference on Humanoid Robots*, vol. 2015-Decem, pp. 547–553, 2015.

[71] H. Wicaksono and C. Sammut, "Tool use learning for a real robot," *International Journal of Electrical and Computer Engineering*, vol. 8, no. 2, pp. 1230–1237, 2018.

[72] M. Toussaint, K. R. Allen, K. A. Smith, and J. B. Tenenbaum, "Differentiable physics and stable modes for tool-use and manipulation planning - Extended abstract," *IJCAI International Joint Conference on Artificial Intelligence*, pp. 6231–6235, 2019.

[73] P. Abelha, F. Guerin, and M. Schoeler, "A model-based approach to finding substitute tools in 3D vision data," *Proceedings - IEEE International Conference on Robotics and Automation*, 2016.

[74] A. Boteanu, A. St. Clair, A. Mohseni-Kabir, C. Saldanha, and S. Chernova, "Leveraging Large-Scale Semantic Networks for Adaptive Robot Task Learning and Execution," *Big Data*, vol. 4, no. 4, pp. 217–235, 2016.

[75] A. Agostini, M. J. Aein, S. Szedmak, E. E. Aksoy, J. Piater, and F. Worgotter, "Using structural bootstrapping for object substitution in

robotic executions of human-like manipulation tasks," *IEEE International Conference on Intelligent Robots and Systems*, vol. 2015-Decem, pp. 6479–6486, 2015.

[76] I. Awaad, G. K. Kraetzschmar, and J. Hertzberg, "Challenges in Finding Ways to Get the Job Done," *Planning and Robotics (PlanRob) Workshop at 24th International Conference on Automated Planning and Scheduling*, 2014.

[77] S. H. Johnson-Frey, "The neural bases of complex tool use in humans," *Trends in Cognitive Sciences*, vol. 8, no. 2, pp. 71–78, 2004.

[78] F. Osiurak and A. Badets, "Tool use and affordance: Manipulation-based versus reasoning-based approaches," *Psychological Review*, vol. 123, no. 5, pp. 534–568, 2016.

[79] A. M. Ruiz and L. R. Santos, "Understanding differences in the way human and non-human primates represent tools: The role of teleological-intentional information," in *Tool Use in Animals: Cognition and Ecology*, C. M. Sanz, J. Call, and C. Boesch, Eds. Cambridge University Press, 2013, ch. 6, pp. 119–133.

[80] A. Rand, "Concept-Formation," in *Introduction to Objectivist Epistemology*, H. Binswanger and L. Peikoff, Eds. Plume Books, 1990, ch. 2.

[81] P. Hemeren and P. Gärdenfors, "A framework for representing action meaning in artificial systems via force dimensions," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 7716 LNAI, pp. 99–106, 2012.

[82] J. R. Hodges, J. Spatt, and K. Patterson, ""What" and "how": Evidence for the dissociation of object knowledge and mechanical problem-solving skills in the human brain," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 96, no. 16, pp. 9444–9448, 1999.

[83] G. A. Miller, R. Beckwith, C. Fellbaum, D. Gross, and K. J. Miller, "Introduction to wordnet: An on-line lexical database," *International Journal of Lexicography*, vol. 3, no. 4, pp. 235–244, 1990.

[84] H. Liu and P. Singh, “Conceptnet - a practical commonsense reasoning tool-kit,” *BT Technology Journal*, vol. 22, no. 4, pp. 211–226, 2004.

[85] F. Ilievski, A. Oltramari, K. Ma, B. Zhang, D. L. McGuinness, and P. Szekely, “Dimensions of commonsense knowledge,” *Knowledge-Based Systems*, vol. 229, 2021. [Online]. Available: https://doi.org/10.1016/j.knosys.2021.107347

[86] J. Feldman, “What is a visual object?” *Trends in Cognitive Sciences*, vol. 7, no. 6, pp. 252–256, 2003.

[87] I. H. Suh, G. H. Lim, W. Hwang, H. Suh, J. H. Choi, and Y. T. Park, “Ontology-based multi-layered robot knowledge framework (OMRKF) for robot intelligence,” *IEEE International Conference on Intelligent Robots and Systems*, no. October, pp. 429–436, 2007.

[88] R. Davis, H. Shrobe, and P. Szolovits, “What Is a Knowledge Representation ?” *AI Magazine*, vol. 14, pp. 17–33, 1993.

[89] A. Sloman, “Why We Need Many Knowledge Representation Formalisms,” *Proceedings BCS Expert Systems Conference*, pp. 163–183, 1984.

[90] S. Coradeschi and A. Saffiotti, “An introduction to the anchoring problem,” *Robotics and Autonomous Systems*, vol. 43, no. 2-3, pp. 85–96, 2003.

[91] S. Harnad, “The Symbol Grounding Problem,” *Physica D*, vol. 42, pp. 335–346, 1990.

[92] L. A. Zadeh, “Fuzzy sets,” *Information and Control*, vol. 8, no. 3, pp. 338–353, 1965.

[93] S. Lemaignan, M. Warnier, E. A. Sisbot, A. Clodic, and R. Alami, “Artificial cognition for social human–robot interaction: An implementation,” *Artificial Intelligence*, vol. 247, pp. 45–69, 2017. [Online]. Available: http://dx.doi.org/10.1016/j.artint.2016.07.002

[94] A. Haidu and M. Beetz, “Action recognition and interpretation from virtual demonstrations,” in *IEEE International Conference on Intelligent Robots and Systems*, 2016, pp. 2833–2838.

[95] J. Winkler, G. Bartels, L. Mösenlechner, M. T. De, and M. Beetz, "Knowledge Enabled High-Level Task Abstraction and Execution," in *Advances in Cognitive Systems*, vol. 2, Palo Alto, CA, 2012, pp. 131–148.

[96] W. M. Toro, F. G. Cozman, K. Revoredo, and A. H. R. Costa, "Probabilistic relational reasoning in semantic robot navigation," *10th International Workshop on Uncertainty Reasoning for the Semantic Web (URSW) in 13th International Semantic Web Conference (ISWC)*, 2014.

[97] M. Tenorth and M. Beetz, "KNOWROB- Knowledge Processing for Autonomous Personal Robots," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2009, pp. 4261–4266.

[98] Y. Zhu, A. Fathi, and L. Fei-Fei, "Reasoning About Object Affordance in a Knowledge Based Representation," *European Conference on Computer Vision*, no. 3, pp. 408–424, 2014.

[99] L. A. Pineda, A. Rodríguez, G. Fuentes, C. Rascón, and I. Meza, "A light non-monotonic knowledge-base for service robots," *Intelligent Service Robotics*, vol. 10, no. 3, pp. 159–171, 2017.

[100] R. Gupta and M. J. Kochenderfer, "Common Sense Data Acquisition for Indoor Mobile Robots," in *Proceedings of the Nineteenth National Conference on Artificial Intelligence, Sixteenth Conference on Innovative Applications of Artificial Intelligence*, San Jose, California, USA, 2004, pp. 605–610.

[101] S. Lemaignan, R. Ros, L. Mösenlechner, R. Alami, and M. Beetz, "ORO, a knowledge management platform for cognitive architectures in robotics," *IEEE/RSJ 2010 International Conference on Intelligent Robots and Systems, IROS 2010 - Conference Proceedings*, no. April, pp. 3548–3553, 2010.

[102] G. H. Lim, I. H. Suh, and H. Suh, "Ontology-based unified robot knowledge for service robots in indoor environments," *IEEE Transactions on Systems, Man, and Cybernetics Part A:Systems and Humans*, vol. 41, no. 3, pp. 492–509, 2011.

[103] M. Daoutis, S. Coradeshi, and A. Loutfi, "Grounding commonsense knowledge in intelligent systems," *Journal of Ambient Intelligence and Smart Environments*, vol. 1, no. 4, pp. 311–321, 2009.

[104] A. Saxena, A. Jain, O. Sener, A. Jami, D. K. Misra, and H. S. Koppula, "RoboBrain: Large-Scale Knowledge Engine for Robots," *arXiv*, pp. 1 – 11, 2014.

[105] M. Tenorth, L. Kunze, D. Jain, and M. Beetz, "KNOWROB-MAP – Knowledge-Linked Semantic Object Maps," in *10th IEEE-RAS International Conference on Humanoid Robots*, Nashville, TN, USA, 2010, pp. 430–435.

[106] M. Tenorth and M. Beetz, "KnowRob: A knowledge processing infrastructure for cognition-enabled robots," *International Journal of Robotics Research*, vol. 32, no. 5, pp. 566–590, 2013.

[107] ——, "Representations for robot knowledge in the KnowRob framework," *Artificial Intelligence*, vol. 247, pp. 151–169, 2017.

[108] R. Gupta and V. Calais Pedro, "Knowledge Representation and Bayesian Inference for Response to Situations," 2005.

[109] R. Gupta and K. Hennacy, "Finite state grammar transduction from distributed collected knowledge," in *Proceedings of the 7th international conference on Computational Linguistics and Intelligent Text Processing*. Mexico City, Mexico: Springer-Verlag, 2006, pp. 343–354.

[110] E. Wang, Y. S. Kim, H. S. Kim, J. H. Son, S. Lee, and I. H. Suh, "Ontology Modeling and Storage System for Robot Context Understanding," *Knowledge-Based Intelligent Information and Engineering Systems*, vol. 3683, pp. 922–929, 2005. [Online]. Available: http://link.springer.com/10.1007/11554028

[111] W. Hwang, J. Park, H. Suh, H. Kim, and I. H. Suh, "Ontology-Based Framework of Robot Context Modeling and Reasoning for Object Recognition," in *Fuzzy Systems and Knowledge Discovery. FSKD 2006. Lecture Notes in Computer Science*, vol. 4223. Springer Berlin Heidelberg, 2006, pp. 596–606.

[112] S. Lallée, S. Lemaignan, A. Lenz, C. Melhuish, L. Natale, S. Skachek, T. Van Der Zant, F. Warneken, and P. F. Dominey, “Towards a platform-independent cooperative human-robot interaction system: I. Perception,” *IEEE/RSJ 2010 International Conference on Intelligent Robots and Systems, IROS 2010 - Conference Proceedings*, no. October 2010, pp. 4444–4451, 2010.

[113] R. Ros, S. Lemaignan, E. A. Sisbot, R. Alami, J. Steinwender, K. Hamann, and F. Warneken, “Which one? Grounding the referent based on efficient human-robot interaction,” *Proceedings - IEEE International Workshop on Robot and Human Interactive Communication*, pp. 570–575, 2010.

[114] M. Daoutis, S. Coradeschi, and A. Loutfi, “Cooperative Knowledge Based Perceptual Anchoring,” *International Journal on Artificial Intelligence Tools*, vol. 21, no. 03, p. 1250012, 2012. [Online]. Available: http://www.worldscientific.com/doi/abs/10.1142/S0218213012500121

[115] L. A. Pineda, I. V. Meza, H. Avilés, C. Gershenson, C. Rascon, M. Alvarado, and L. Salinas, “IOCA: An Interaction-Oriented Cognitive Architecture,” *Journal Research in Computing Science*, vol. 54, pp. 273–284, 2010.

[116] M. Daoutis, A. Loutfi, and S. Coradeschi, “Knowledge Representation for Anchoring Symbolic Concepts to Perceptual Data,” *Bridges between the methodological and practical work of the robotics and cognitive systems communities*, no. January, 2012.

[117] A. Olivares-Alarcos, D. Beßler, A. Khamis, P. Goncalves, M. K. Habib, J. Bermejo-Alonso, M. Barreto, M. Diab, J. Rosell, J. Quintas, J. Olszewska, H. Nakawala, E. Pignaton, A. Gyrard, S. Borgo, G. Alenyà, M. Beetz, and H. Li, “A review and comparison of ontology-based approaches to robot autonomy,” *Knowledge Engineering Review*, vol. 34, 2019.

[118] H. Liu and P. Singh, “ConceptNet — A Practical Commonsense Reasoning Tool-Kit,” *BT Technology Journal*, vol. 22, no. 4, pp. 211–226, 2004.

[119] C. Fellbaum, Ed., *WordNet: An Electronic Lexical Database*. Cambridge, MA ; London: The MIT Press, 1998.

[120] D. B. Lenat, "Cyc: A large-scale investment in knowledge infrastructure," *Commun. ACM*, vol. 38, no. 11, pp. 33–38, 11 1995.

[121] A. T. Gavin R. Hunt, R. Gray, "Why is Tool Use Rare in Animals?" in *Tool Use in Animals: Cognition and Ecology*, J. B. C. Sanz, Crickette M. Call, Ed. Cambridge University Press, 2013, ch. 5, pp. 89–118.

[122] C. R. Rogers, "A Theory of Therapy and Personality Change: As Developed in the Client-Centered Framework," *Psychology: A Study of a Science*, vol. 3, pp. 184–256, 1959.

[123] C. Baber, "How tool use breaks down," in *Cognition and Tool Use*. Taylor and Francis, 2003, ch. 8, pp. 89–103.

[124] A. Augello, S. Gaglio, G. Oliveri, and G. Pilato, "Concepts, protoconcepts, and shades of reasoning in neural networks," in *Proceedings of the 6th International Workshop on Artificial Intelligence and Cognition (AIC)*, vol. 2418, Palermo, 2018, pp. 111–124.

[125] P. Jaccard, "The Distribution Of The Flora In The Alpine Zone.1," *New Phytologist*, vol. 11, no. 2, pp. 37–50, 1912.

[126] M. K. Vijaymeena and K. Kavitha, "A Survey On Similarity Measures In Text Mining," *Machine Learning and Applications: An International Journal*, vol. 3, no. 2, pp. 19–28, 2016.

[127] T. Sørenson, *A Method of Establishing Groups of Equal Amplitude in Plant Sociology Based on Similarity of Species Content and Its Application to Analyses of the Vegetation on Danish Commons*. I kommission hos E. Munksgaard, 1948.

[128] K. Lai, L. Bo, X. Ren, and D. Fox, "A Large-Scale Hierarchical Multi-View RGB-D Object Dataset," in *IEEE International Conference on Robotics and Automation (ICRA)*, Shanghai, China, 2011, pp. 1817–1824.

[129] C. A. Mueller and A. Birk, "Conceptualization of Object Compositions Using Persistent Homology," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Madrid, 2018.

[130] A. Budanitsky and G. Hirst, “Evaluating wordnet-based measures of lexical semantic relatedness,” *Computational Linguistics*, vol. 32, no. 1, pp. 13–47, 3 2006.

[131] P. Resnik, “Using Information Content to Evaluate Semantic Similarity in a Taxonomy,” in *Proceedings of the 14th International Joint Conference on Artificial Intelligence*, Montreal, 8 1995, pp. 448–453. [Online]. Available: http://arxiv.org/abs/cmp-lg/9511007

[132] R. Rada, H. Mili, E. Bicknell, and M. Blettner, “Development and Application of a Metric on Semantic Nets,” *IEEE Transactions on Systems, Man and Cybernetics*, vol. 19, no. 1, pp. 17–30, 1989.

[133] Z. Wu and M. Palmer, “Verb semantic and lexical selection,” *In 32nd annual Meeting of the Association for Computational Linguistics*, pp. 133–138, 1994.

[134] D. Lin, “An Information-Theoretic Definition of Similarity,” in *Proceedings ofthe International Conference on Machine Learning.*, 1998, pp. 296–304.

[135] J. J. Jiang and D. W. Conrath, “Semantic Similarity Based on Corpus Statistics and Lexical Taxonomy,” *Proceedings on International Conference on Research in Computational Linguistics*, p. 19–33, 8 1997.

[136] E. Loper and S. Bird, “NLTK: The Natural Language Toolkit,” in *Proceedings of the ACL Workshop on Effective Tools and Methodologies for Teaching Natural Language Processing and Computational Linguistics*. Philadelphia: Association for Computational Linguistics, 2002, pp. 62–69. [Online]. Available: http://arxiv.org/abs/cs/0205028

[137] N. Shrivatsav, L. Nair, and S. Chernova, “Tool substitution with shape and material reasoning using dual neural networks,” *arXiv*, 2019.

[138] C. A. Mueller and A. Birk, “Hierarchical Graph-Based Discovery of Non-Primitive-Shaped Objects in Unstructured Environments,” in *International Conference on Robotics and Automation*, 2016.

Ehrenerklärung

Ich versichere hiermit, dass ich die vorliegende Arbeit ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe; verwendete fremde und eigene Quellen sind als solche kenntlich gemacht. Insbesondere habe ich nicht die Hilfe eines kommerziellen Promotionsberaters in Anspruch genommen. Dritte haben von mir weder unmittelbar noch mittelbar geldwerte Leistungen für Arbeiten erhalten, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen.

Ich habe insbesondere nicht wissentlich:

- Ergebnisse erfunden oder widersprüchliche Ergebnisse verschwiegen
- statistische Verfahren absichtlich missbraucht, um Daten in ungerechtfertigter Weise zu interpretieren,
- fremde Ergebnisse oder Veröffentlichungen plagiiert,
- fremde Forschungsergebnisse verzerrt wiedergegeben.

Mir ist bekannt, dass Verstöße gegen das Urheberrecht Unterlassungs- und Schadensersatzansprüche des Urhebers sowie eine strafrechtliche Ahndung durch die Strafverfolgungsbehörden begründen kann. Die Arbeit wurde bisher weder im Inland noch im Ausland in gleicher oder ähnlicher Form als Dissertation eingereicht und ist als Ganzes auch noch nicht veröffentlicht.

Date: ....

Madhura Thosar