

Why is That a Good or Not a Good Frying Pan? – Knowledge Representation for Functions of Objects and Tools for Design Understanding, Improvement, and Generation

Seng-Beng Ho
Institute of High Performance Computing (IHPC)
Agency for Science, Technology and Research (A*STAR)
Singapore
hosb@ihpc.a-star.edu.sg

Abstract—The understanding of the functional aspects of objects and tools is of paramount importance in supporting an intelligent system in navigating around in the environment and interacting with various objects, structures, and systems, to help fulfil its goals. A detailed understanding of functionalities can also lead to design improvements and novel designs that would enhance the operations of AI and robotic systems on the one hand, and human lives on the other. This paper demonstrates how a particular object – in this case, a frying pan – and its participation in the processes it is designed to support – in this case, the frying process – can be represented in a general function representational language and framework, that can be used to flesh out the processes and functionalities involved, leading to a deep conceptual understanding with explainability of functionalities that allows the system to answer “why” questions – why is something a good frying pan, say, or why a certain part on the frying pan is designed in a certain way? Or, why is something *not* a good frying pan? This supports the re-design and improvement on design of objects, artifacts, and tools, as well as the potential for generating novel designs that are functionally accurate, usable, and satisfactory.

Keywords—functional understanding, functional representation, knowledge representation for function and affordance, artifact functionality, frying pan function representation, design understanding, design generation

I. INTRODUCTION

While an intelligent autonomous system (IAS) navigates around and interacts with the environment and the objects within it, there are at least two aspects of intelligent processing that feature prominently in its activities. Firstly, it has to recognize the structures and objects in the environment, and secondly, it has to understand what they *afford* it – i.e., the functions of these structures and objects that may support its activities and problem solving processes. Currently, the field of computer vision concerns primarily the former – i.e., object or structure recognition: labeling an object (e.g., a cup) or structure (e.g., a wall) as belonging to a known class [1]. Much less work in computer vision in particular and in AI in general has been devoted to the second aspect – function understanding [2]–[9]. This paper proposes a general knowledge representation language and framework that allows an IAS to characterize and understand the functions of

objects, structures, and systems, hence supporting its intelligent activities such as problem solving processes to satisfy its goals and needs.

There have been some attempts at recognizing the functionalities of objects and tools [4], [6]–[9]. Functionalities can be represented in a shallow [8] or a deeper level. In Ho [4] and Wu [9], the detailed constructions of the objects involved are represented, and interactions between objects and humans are considered and simulated to elicit the functionalities involved, going one step further than [8] in characterizing the understanding of the functions. However, there is still no deeper level of understanding such as being able to explain *why* certain constructions are present in the object/tool, and there is also no general framework proposed to unify all kinds of situations involving functioning and functions of objects/artifacts/tools.

Ho [10] laid out a unified and general framework for characterizing functions, using a representational language that employs a fixed, limited, and general set of concepts and links for representing functional concepts in general situations (the CD+ representational framework). In this paper, this framework is applied to answer specific “why” questions concerning the functionalities of a cooking tool – the frying pan.

Currently there are generative AI systems that can generate many novel variations of objects, structures, and designs [11], [12]. Our proposed method would take these generated forms as starting point and evaluate the suitability of these novel forms for certain functionalities. As a result, the design of objects could be improved accordingly, and novel objects could be identified as well.

II. MOTIVATION AND BACKGROUND

Unlike the process of recognition, where the end result is usually a label attached to a certain structure or object, the end result of the process of function understanding involves some descriptions of the function and functioning of the system, structure, and object involved. For example, to characterize the *functioning* of a kettle, we might say, “a kettle works by employing some electrical heating elements placed in its interior, which, when electricity is passed through them, they

heat up, and the heat is transferred to the water that is also placed in the kettle’s interior in contact with these heating elements. After some time of heating the water, it reaches the boiling point and boils.”

In order for a machine or IAS to really understand the function and functioning involved, the “meaning” or the deep structures of the sentences above have to be encoded, elucidated, and represented in some machine processable computational structures. Meaning representation is usually addressed by the fields of linguistic semantics [13]–[16] and computational linguistics [17]–[19].

Among the various attempts in linguistics and computational linguistics, it has been identified in [10] that conceptual dependency (CD) theory of Schank [20]–[23] is the most suitable deep meaning representational framework to be applied to the characterization of functionality. The main reason for this is that an adequate functional characterization of any system, structure, or object calls for a representational systems that can deal with specifications of complex causalities between events, objects, and structures, and CD provides just such a causal framework. Ho [10] extended the original CD framework to CD+, incorporating new constructs that provide for a more complete framework for the representation of functionality.

Another relevant domain of AI research is the recent work on recognizing and characterizing human activities in various environments [24]–[31]. These works can be supplemented by CD+ to capture the complex causal structures inherent in the characterizations of many complex functional concepts.

III. THE REPRESENTATIONAL FRAMEWORK

It is instructive that we first describe the basic representational constructs of CD+ [10].

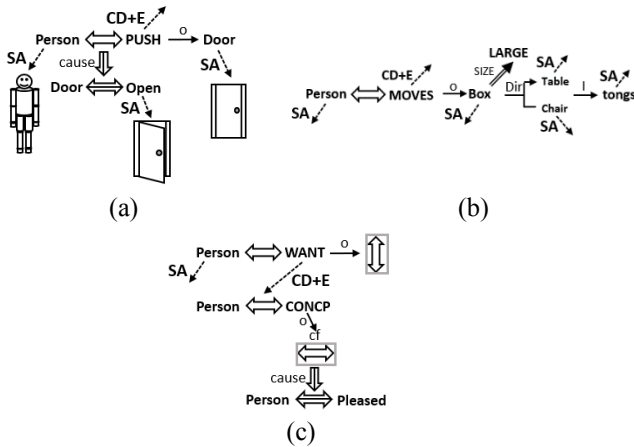


Fig. 1. CD+ representations for illustrating the basic CD+ constructs. (a) “Person pushes the door open.” (b) “Person moves the large box from the chair to the table with a pair of tongs. (c) Person wants something.

Fig. 1 provides three examples of CD+ representations. In Fig. 1(a), the event represented is “Person pushes the door open.” In CD+, an event is a “conceptualization” linking a subject and an action. In this event, “Person” is the subject, and “PUSH” is the action. The link linking them is a “conceptualization” link (a double arrow). The object of PUSH in this case is Door - an object (“o”) link is used. There

is actually a causality present in this statement that is not explicit, that is, the pushing of the door *causes* the door to acquire the *state* of “Open.” A causal link is the downward arrow with a vertical line going down its center. The double arrow with a line going down its center is the “state” link.

In addition, there are two kinds of constructs that elaborate on the symbols in this representation. The Structure Anchor (SA) is a structural specification of symbols that are objects. Person and Door are both symbols pointing to certain real world physical constructs through SA. These structures are detailed models that could be represented by using, say, vectoral constructs. Another kind of elaboration is called the CD+ Elaboration (CD+E). This kind of link elaborates actions, and points to a specification of how the action involved is executed. This elaboration could in turn be specified in the form of a CD+ representation.

Fig. 1(b) is a representation of the event “Person moves the large box from the chair to the table.” The Dir link specifies the direction of movement. The SIZE of the box, LARGE, can be specified using a “property” link. “I” is an Instrument (I) link.

Fig. 1(c) is the representation of the conceptualization “Person wants something.” When a person wants something, the person CONCePtualizes (CONCP) that if that something, which could be an event, represented as another conceptualization (shown as the usual double arrow bounded by a box), were to come true in the future, the person is “Pleased.” In the figure, “c” stands for “if” (conditional), and “f” stands for “future.”

Having described the basic representational constructs, we next turn to the fleshing out of the function of a frying pan.

IV. THE REPRESENTATION OF FRYING AND THE FUNCTION OF A FRYING PAN

In this section, we describe the processes of cooking and frying using CD+.

Given any activity, whether it is those of more immediate current concern such as baking, frying, cooking, or all other activities that are associated with human day-to-day functioning, there is usually background knowledge associated with it. The background knowledge “sometimes people like to eat cooked food” puts cooking in the context of certain human needs. We begin with a discussion on this.

A. Basic Functioning and Process of Cooking and Eating – Background Knowledge

Fig. 2(a) shows the representation for the conceptualization of “Person wants to eat well-cooked food.” (Of course, a person may also want to eat raw food at some other times.) This is the same representational structure as that of Fig. 1(c). The state of “well-cookedness” is represented as COOKED(WELL). Note that the state of “well-cookedness” (W) determines in a fuzzy way, Person’s satisfaction (S). This in turn determines the degree of being “Pleased” (P).

In a number of places in this paper we use predicate-like representations such as COOKED(WELL), In(Spatula, Frying-Pan), Centered-On(Frying-Pan, Burner), etc. and we often omit the first argument if it is clear, such as Centered-

On(Burner) in Fig. 3(a) because this is part of the elaboration of the main concept Frying-Pan.

For a human being to determine that something has been “cooked” or “well-cooked”, she has internal sensing criteria in her gustatory system to define what is meant by something being “cooked” or “cooked-well.” Humans also often experience and learn the association of certain visually identifiable attributes (externally or internally) and the degree of being cooked. Thus COOKED(WELL) can be determined by a human or AI through vision.

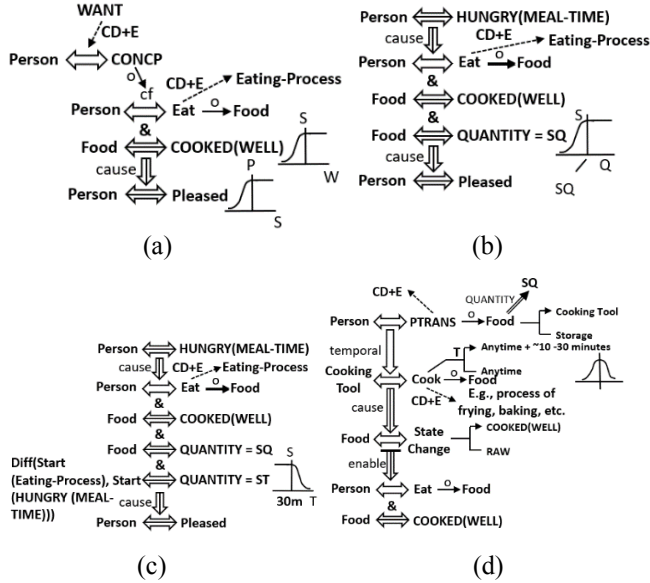


Fig. 2. Background knowledge on cooking and eating. (a) Person wants to eat well-cooked food. (b) Person wants to eat a Satisfactory Quantity (SQ) of well-cooked food when hungry at mealtime. (c) When hungry at mealtime, Person wants to eat a Satisfactory Quantity (SQ) of food, and Person also wants to start eating within a Satisfactory amount of Time (ST, typically, say, about 30 minutes) of feeling hungry. (d) A typical cooking process. Here we introduce an “enable” link – the *cause* arrow topped with a horizontal line [10].

Fig. 2(b) shows the representation for the conceptualization of “Person wants to eat a Satisfactory Quantity (SQ) of well-cooked food when hungry at mealtime.” Given a certain quantity of food (Q), the degree of satisfaction (S) is a fuzzy function of Q, and HUNGRY(MEAL-TIME) is also a fuzzy function – above a certain degree of hunger, the *want* to eat food is (strongly) triggered (the *cause* link on top). Pleased is now a function of COOKED(WELL) and SQ.

There could be situations in which a smaller quantity of food is first cooked and served, and then as the person is eating, more is cooked and served (e.g., the Japanese dining style known as Omakase). But the above is *one* satisfactory situation.

Fig. 2(c) shows the representation for the conceptualization of “When hungry at mealtime, Person wants to eat a Satisfactory Quantity (SQ) of food, and Person also wants to start eating within a Satisfactory amount of Time (ST, typically about 30 minutes) of feeling hungry.”

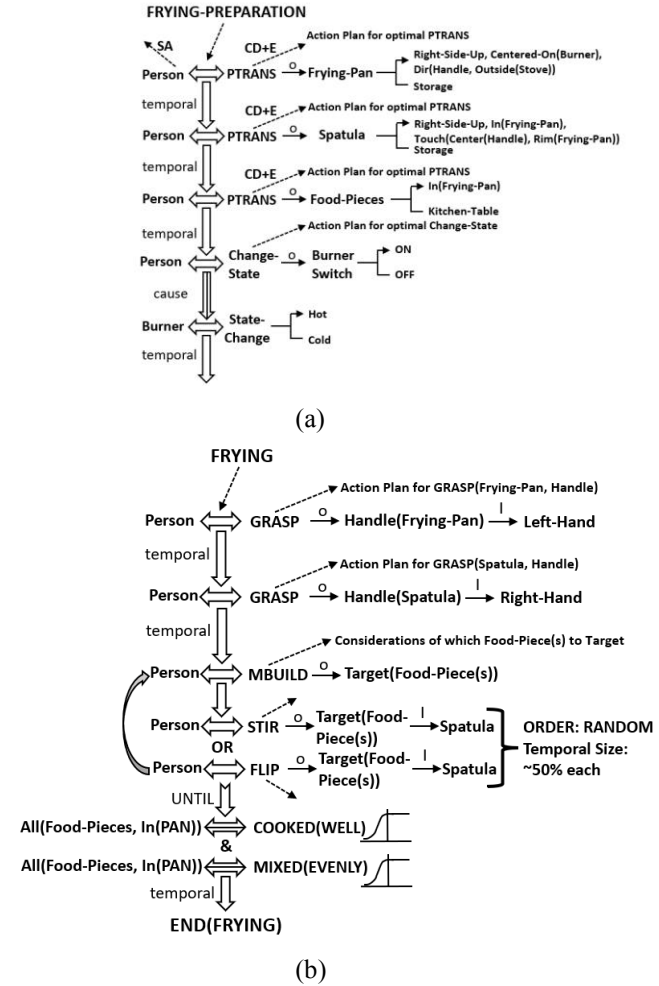
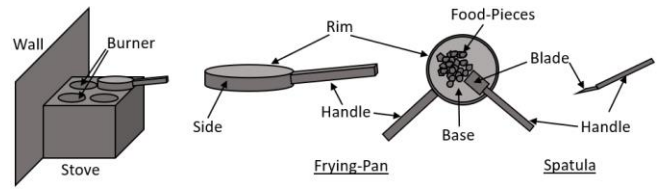


Fig. 3. (a) FRYING PREPARATION process. (b) FRYING process.

In Fig. 2(d) we show the representation for a typical cooking process. The first step of the process is Physically TRANSferring (PTRANS) the food to be cooked from the Storage area to the Cooking Tool (Frying pan for frying, oven for baking, etc.). A new kind of link, “temporal,” is introduced here in which the first conceptualization flows to the next one after some time.

How are the various causal links in Figs. 2(a) – (d) learned? A causal learning process which is not explored in this paper but which has been extensively studied elsewhere [32]–[37] identifies certain *temporal* associations (first, put well-cooked food in mouth, then experience pleasure) as *causal* relations.

B. Observed Functioning and Process of Frying

Assume that there is a process similar to that described in prior work such as that in [24]–[26], in which human daily activities in a house, say, are observed, learned, and encoded.

A possible CD+ encoding of the frying process is shown in Fig. 3.

In Fig. 3, we divide the process into FRYING-PREPARATION and FRYING proper (in reality these stages are contiguous). Fig. 3(a) shows that the FRYING PREPARATION process consists of first PTRANSing the Frying-Pan from the Storage to the Burner on a Stove. It is positioned Right-Side-Up, Centered-On the Burner, and is oriented with the Handle facing the Outside of the Stove. The meaning of these spatial terms are pre-defined as grounded concepts or near-ground concepts as articulated in [10], [38].

In this first step, the details of the PTRANS action is described in a CD+E which is the “Action Plan for (Optimal) PTRANSing of Frying-Pan from Storage to the top of the Burner, positioned accordingly.” This Action Plan can be learned from observation, or obtained from a motion planning process, such as described in [39], [40].

The other steps are self-explanatory, and the process ends with Burner changing state from Cold to Hot.

After the process of FRYING-PREPARATION, the process of FRYING begins (Fig. 3(b)). Typically, for a right-hander, Person would use Left-Hand to GRASP Handle(Frying-Pan) and Right-Hand to GRASP Handle(Spatula).

Also, typically, when a person is frying something, she would sometimes *stir* the food pieces around to mix them and sometimes *flip* the food pieces so that different sides of the food pieces get cooked. These two actions may take place in any sequence, but they each may be executed, say, roughly equally often. At the beginning of each action, the person usually forms some thoughts about which food-piece should be “operated on” next (e.g., “that piece looks uncooked, let’s target to flip it next...”, or “these pieces are not well mixed, let’s target to stir them around,” etc.). This is an MBUILD (Mental reasoning to BUILD some internal mental representations) process in CD+ [10], [20], which could determine whether STIR or FLIP is to be executed as shown in Fig. 3(b).

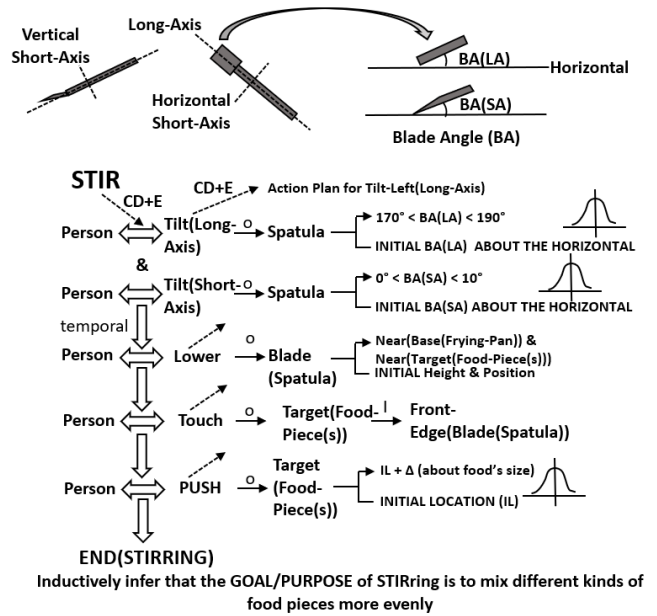
A typical end-process test would be when “all food pieces in the frying pan are cooked AND all food pieces in the frying pan are evenly (well) mixed.” Having the food well mixed is a basic “want” of a typical person.

The criteria of “COOKED(WELL)” and “MIXED(EVENLY)” are also fuzzy functions. The definition of “MIXED(EVENLY)” is described in Section 4.7. These terminating conditions could be inductively derived through many instances of observation, or through background knowledge such as that in Fig. 2(a).

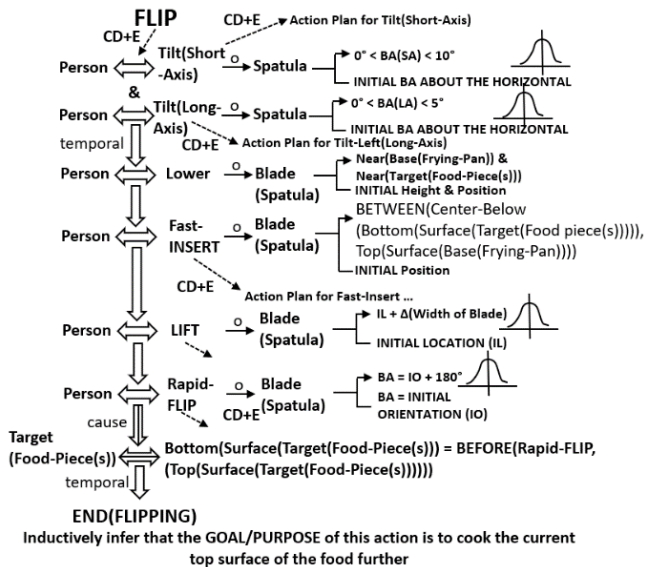
Fig. 4 shows the processes of STIRring and FLIPping. These are self-explanatory. Longish objects have a long axis and two short axes [38]. (Fig. 6(a) illustrates the positioning of the spatula in the STIRring mode.)

One thing to note is that in the FLIPping process, the spatula has to be inserted underneath the food pieces in a “fast” insertion process, otherwise, Blade(Spatula) would just push the target food-pieces around and not be positioned for the FLIPping step. Also, after LIFTing the food, a Rapid-

FLIPping action is executed, otherwise, instead of flipping the food-piece(s), they may just fall back to the frying pan with the respective faces still facing in the same directions. We made assumptions on the various angles indicated in Fig. 4 but they are supposed to be derived from real-world observations.



(a)



(b)

Fig. 4. (a) STIRring process. (b) FLIPping process.

There could yet be other ways to STIR and FLIP Food-Piece(s) and these could be added to the processes in Fig. 4 as disjunctive processes to achieve the same ends.

C. Cooking/Frying Success Conditions

There are two further conditions that must be tested before the success of the cooking/frying process could be ascertained.

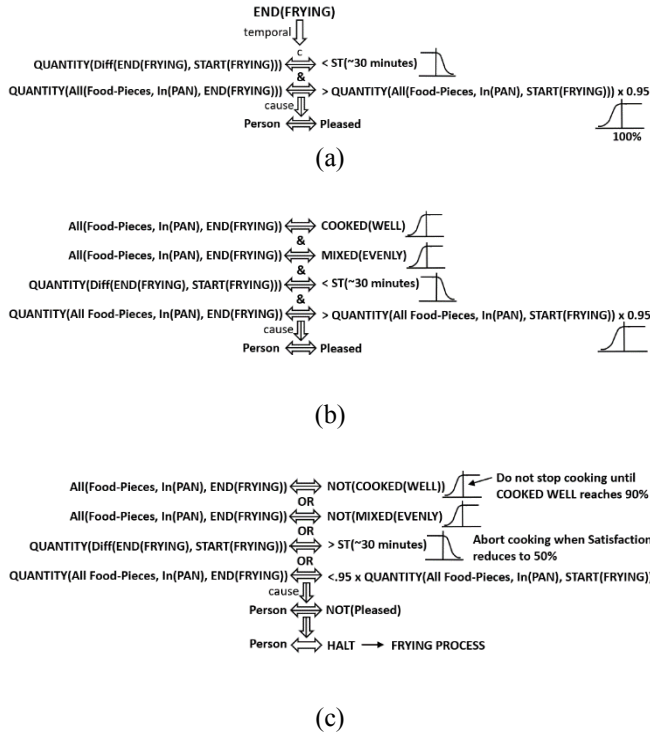


Fig. 5. (a) Satisfaction from the time taken to cook and the amount of food that is retained in the frying pan. (b) Conditions in (a) plus the conditions in Fig. 3(b). (c) The contrapositive of (b).

As stated in Fig. 2(c), a person typically wants to start eating a sufficient quantity (SQ) of food within a certain time (ST) of beginning to feel hungry. Let's assume that this time is about 30 minutes. The system could infer that completing each frying or whatever kind of cooking process within ST is a condition of "satisfaction," and that Person will be Pleased. This is stated in Fig. 5(a).

Now, in whatever cooking process, especially in a frying process, food pieces could be lost due to excessive STIRring or LIFTing and FLIPping. Hence, another satisfaction condition is that the quantity of food pieces at the end of the frying process should be, perhaps, typically larger than 95% (say, through observation what is typically acceptable) of what the frying process started with. The satisfaction with regard to this is also a fuzzy function.

Fig. 5(b) combines all the 4 conditions, including COOKED(WELL) and MIXED(EVENLY).

Fig. 5(c) is a contrapositive of Fig. 5(b), which is that if any of the 4 conditions is not met, Person would not be Pleased and it would cause Person to HALT the process.

Later in Section 4.4, Fig. 5(c) would be used to judge if certain frying tools (Frying-Pan and Spatula) with certain designs can adequately function to support frying.

D. Understanding the Frying Process

With these representations, the temporal links can be used to answer "what" questions. If the question is "What happens after Food-Pieces are placed inside Frying-Pan?" The answer would be "Person then turns on Burner." (Fig. 3(a)). Causal links could be used to answer "why" questions. For example, if the question is "Why Person is not Pleased with the frying process?" The answer could be "Because the frying process took more than 30 minutes." (Fig. 5(c)). At this stage, the system does not seem to possess the representations for answering questions related to the causal functions of the frying pan and the spatula, and for that matter, the burner, in supporting the cooking process of turning the food pieces from uncooked to well-cooked, and also well-mixed in the frying pan. If we look at Fig. 3(a), it does show that Burner is turned on, and it changes from Cold to Hot, but there is no knowledge that it is the heat that is transmitted through the frying pan placed on top of Burner, that in turn gets transmitted to the food placed in the frying pan that causes the food to be cooked.

Similarly, in Fig. 3(b), it is the stirring and flipping actions that resulted in the well-mixing and well-cookedness of the food pieces, but there is no explicit causal links here.

There needs to be further background knowledge of physics of heat and heat transfer across metal, as well as physics of the more mechanical aspects of things such as pushing, lifting, flipping, etc. that will elicit the causal understanding.

In the interests of paper length, we assume that this knowledge of physics can be learned and encoded in CD+ to support reasoning to be discussed.

E. The Function of the Frying Pan's Side

In this section we begin with the understanding of how the frying pan functions to support cooking in general and frying in particular, before we present the processes that identify what is wrong with certain other designs that cannot fulfil the function in a satisfactory manner.

The title of this section seems a little narrow – we are only looking at the function of one part of the frying pan - its Side, shown in Fig. 3(a). This is mainly because we will only be considering a small sets of physical knowledge related to basic physical interactions. For example, to understand the function of the base of the frying pan, which is partly to transmit heat from Burner to Food-Pieces, we need background knowledge on heat. However, the consideration of the function of the frying pan's side provides the paradigm for reasoning about the other parts of the frying pan, or for that matter, other objects and tools.

In Fig. 6(a) we show the cross-section of the stirring process, which, according to the encoding in Fig. 4(a), involves flipping the spatula "upside-down" as shown and the edge of the blade is used to touch and push the food pieces involved.

Now, we can set this Fig. 6(a) configuration up in a simulator such as the Unity engine (www.unity.com) and have it simulate what would happen after executing some stirring action. The processes outlined in Fig. 3(b) (of the entire frying process) and Fig. 4(a) are used to direct the simulation

process. Naturally, the simulator would generate the expected consequences: the food pieces staying within the confines of the frying pan.

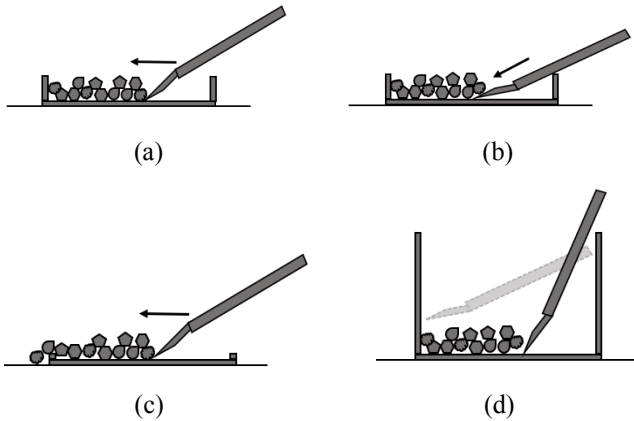


Fig. 6. (a) The STIRring situation. (b) The FLIPping situation. (c) A “frying pan” with a very low side. (d) A “frying pan” with an excessively high side. The handle is omitted here.

However, the physics reasoning engine of Unity has all the physics rules built-in and they are hence implicit and cannot be tapped directly for reasoning about causality. This is when the explicitly encoded rules physics becomes useful (which we have mentioned at the end of Section IV(D) that can in principle be encoded in CD+).

Physics rules dictate that the pushing action would transmit forces over some of the food pieces, and ultimately the forces would reach some of the food pieces right next to and are in contact with the side. Now, as the side is immobile, the food pieces right next to it and are contacting it would be immobile too and stay within the frying pan. Hence the following conclusion can be derived: “The side of the frying pan *causes* the food pieces *not* to go beyond the confine of the frying pan.”

Ho [10] describes a counter-factual process that can also be executed to infer the functions of the various parts of an object. In this case, suppose the AI system takes a mental action of *removing* the side of the frying pan, and then execute a simulation of the stirring process of Fig. 4(a). If the Unity engine is then used to simulate the consequences, it would find that the food pieces would spill out of the frying pan’s confine. (These processes of mental reasoning and activating simulation, termed “MENTAL EXPERIMENTS,” can be represented in CD+ format, as shown in (Ho 2022). Therefore, the AI system observes that “had the side of the frying pan *not* been there, the food would spill outside the frying pan during the stirring process,” (representable in CD+) and concludes that “the function of the side of the frying pan is to *enable* the food pieces to stay within the base of the frying pan during the stirring process” (also representable in CD+).

This function of the Side can also be revealed in the flipping process shown in Fig. 6(b).

F. Why is that a Good or Not a Good Frying Pan?

We are now armed with the devices that allow an AI system to reason about whether certain tool/artifact/object can satisfy certain functional requirements, and if not, *why* not (i.e., providing explanations).

Fig. 7 shows the overall process of functional reasoning. Given an object to be tested for frying function, the FUNCTIONAL REASONER deposits the object involved into a SIMULATION ENVIRONMENT (a buffer of sort). (The structure of Fig. 7 had been earlier proposed in [4], [10].) Then the reasoner follows the processes dictated by Fig. 3(a) for FRYING-PREPARATION (this representation is stored in the CONCEPTUAL CORE). There is a module called PHYSICAL REASONER that contains both Unity’s implicit knowledge (or for that matter, the knowledge in any simulation engine) as well as possibly explicitly stated knowledge of physics (say, in CD+). All these modules and processes are subparts of a more general Unified General Autonomous and Language Reasoning Architecture (UGALRA) described in [10].

Next, the FRYING processes described in Figs. 4(b) and 5 are executed accordingly. A slightly tricky situation is, a usual simulator may not be able to simulate how well-cooked certain piece of food may look after certain amount of heat is added to it. For just this purpose, we could build in a rule that says that each piece of food has to have each of its various faces touching Top(Surface(Base)) of Frying-Pan for X minutes in total. In a typical, say, 20-30-minute frying process, X could be in the order of 3-5 minutes.

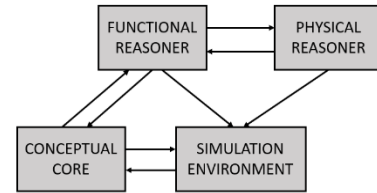


Fig. 7. The various modules of FUNCTIONAL REASONER, PHYSICAL REASONER, SIMULATION ENVIRONMENT, and CONCEPTUAL CORE interacting with each other for functional reasoning. See text for explanations.

The MIXED(EVENLY) condition can be defined as follows: the entire area of Food-Pieces is first divided into, say, 10 or more subareas, and if the number of different kinds of food pieces in each of the subareas are all about equal, then the MIXED(EVENLY) condition is met.

The FUNCTIONAL REASONER could first end the FRYING process through the two conditions stated in Fig. 3(b) – COOKED(WELL) and MIXED(EVENLY), and then check if the other two conditions, “< ST” and “>95%”, stated in Fig. 5(a) are met. If so, “Person is Pleased” and the frying pan fulfils its function.

At this point, the system can assume that all the individual steps of the entire frying process must have been carried out correctly, thereby leading to the final positive result. There are two aspects involved: the activities carried out at each step, and the tools involved in supporting those activities. Therefore, Burner must have fulfilled its function, so did Frying-Pan and Spatula. At this stage, the system can respond with “it is a good Burner,” “it is a good Frying-Pan,” “it is a good Spatula,” “it is a good Stove,” etc. (“etc.” because the list could go on and include “it is a good heat source supplied to the burner” and so on.)

Suppose the system is asked a further detailed question such as “WHY is that a good frying pan?” The system would consider the 4 conditions in Fig. 5(b) each in turn to see how the frying pan and its construction cause the condition to be satisfied.

Suppose the system considers the condition “leftover quantity of food pieces in frying pan must be $> 95\%$ of what was there in the beginning of the frying process,” and tries to identify how the frying pan may cause that to happen, it could carry out a MENTAL EXPERIMENT like mentioned in the previous section, in which the functioning of the frying pan in the process is dissected. At this stage, the system may not know that it is the frying pan and its construction that allow this to happen, and it has to consider all the tools involved and their parts.

Because the system removes ONE factor in each step of its consideration, it can conclude that the observed unsatisfactory consequent is *cause* by THAT factor. Therefore, when the Side of the pan is being considered, the reasoning is, “had Frying-Pan *not* have Side, Food-Pieces will not stay in Frying-Pan and the $>95\%$ condition fails.”

The system could respond with an output such as “That is a good frying pan because its side prevents food pieces from spilling outside of the frying pan in the frying process.”

Likewise, there are other aspects of the frying pan’s function that can be deduced: “That is a good frying pan because its base is not too small to allow stirring and mixing of food evenly,” “That is a good frying pan because its base is made of metal which allows heat to transfer from the heat source to the food,” etc.

Suppose in another mental experiment, the system removes Blade of Spatula instead (Frying-Pan and other things stay constant), then the consequence would be that the STIRring process is very inefficient and the FLIPping process is nearly impossible. Food-Pieces will still get pushed around, resulting in some mixing, and some may occasionally roll over, resulting in all sides of Food-Pieces get cooked, but by the time the process finishes, it would be much longer than ST. In this experiment, the $>95\%$ rule may still be satisfied. The system can continue to drill down further to understand WHY a Spatula without Blade or with a poorly designed Blade cannot function well to support the process of frying.

In Fig. 6(c), we show a frying pan that has a very low side. A similar process of reasoning will lead to the conclusion that “That is NOT a good frying pan,” and “That is not a good frying pan BECAUSE its side is too low and in the process of frying food-pieces will spill outside the pan and much more than 95% of food-pieces will be lost.”

Fig. 6(d) shows a situation in which the side of the frying pan is way too high. In simulation, it is likely that the time taken for “COOKED(WELL)” will be too long because the spatula cannot effectively flip the food pieces around so that all the sides of the food pieces can be properly cooked within a reasonable time ($<ST$), though in principle the sides could still ultimately be cooked as the pushing action may still flip the pieces occasionally. Fig. 6(d) also shows that even if the entire Spatula is lowered into Frying-Pan to obtain a good angle for the Blade (the “dotted” Spatula), the FLIPping

action still cannot be properly executed. The MIXED(EVENLY) condition could probably still be met within ST.

In a similar reasoning process as discussed above, it is identified that it is the height of the side of the frying pan that causes the inability in positioning Blade(Spatula) in a correct way as stipulated in Fig. 4(b) for the effective executing of the Fast-INSERT step, as a result the FLIPping act cannot be successfully carried out (i.e., no Action Plan exists for FLIPping to be executed properly). Thus, this is WHY the frying pan is not a good frying pan.

A verbal explanatory output of the system could be something like “This frying pan is not a good frying pan because its side is too high, which prevents the spatula from being able to adopt an angle to be able to flip the food pieces, and this leads to a much longer time than that of ST to reach the well-cooked states of all the food pieces involved, and this in turn leads to the person involved not being pleased.”

With other knowledge such as heating and burning effects on various materials and the human body, we can present utensils like a frying pan or a spatula made of paper or porcelain, or with inadequate handle length, etc., and have the system explain why they are not good frying pans, or spatulas. (If the handle of the frying pan or spatula is too short, the human or robot’s hand could be burned in the process of frying – i.e., in the process of carrying out the frying actions as stipulated in the representation of Fig. 3(b).)

V. CONCLUSIONS AND FUTURE WORK

This paper demonstrates that the representation and understanding of functionalities involve several aspects of events taking place in the environment and requires the use of a representation language. These include the relevant contexts in which certain processes and functioning take place (in our particular example, it would be cooking and frying), and in which certain objects and tools participate in the causal processes that support the functionalities involved. There is also usually some background knowledge involved (in our particular case, this would be the eating and cooking background knowledge of Fig. 2, and the physics knowledge discussed). And we have successfully demonstrated that whether it is background knowledge, knowledge about the particular causal effects that a certain object or tool have in its interacting with other objects to elicit certain functionalities, or the knowledge about the entire context in which the functioning takes place, in which the adequacy of the objects and tools involved in supporting the desired functionalities is evaluated and understood, they can all be represented using the CD+ representational language and framework [10].

This proposed method and framework could work with certain current generative AI systems [11], [12] in evaluating the novel designs they generate, thus identifying which novel design would work well functionally in certain given environments to serve certain purposes.

Future work includes investigating how this methodology could be extended to many other daily-encountered tools and objects, as well as more specialized tools and objects found in factories and other business environments such as offices, restaurants, gas stations, etc. to represent and understand the

functionalities involved so as to be able to judge how well they can fulfil certain desired functionalities as well as how new and better designs can be concocted. Future work will also flesh out the detailed mental experimental processes in CD+ form for reasoning about design improvement and novel design generation. The other main work to devote future effort to would be a complete program implementation of the representational constructs and the attendant reasoning processes.

REFERENCES

- [1] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2006.
- [2] P. Freeman and A. Newell, "A model for functional reasoning in design," in *Proceedings of the International Joint Conference on Artificial Intelligence*, London, England, 1971, pp. 621–640.
- [3] J. J. Gibson, *The Ecological Approach to Visual Perception*. Hillsdale, New Jersey: Lawrence Erlbaum Associates, 1979.
- [4] S.-B. Ho, "Representing and Using Functional Definitions for Visual Recognition," Ph.D. Thesis, University of Wisconsin-Madison, U.S.A., 1987.
- [5] M. Dimanzo, E. Trucco, F. Giunchiglia, and F. Riccit, "FUR: Understanding functional reasoning," *International Journal of Intelligent Systems*, vol. 4, no. 4, pp. 431–457, 1989, doi: 10.1002/int.4550040403.
- [6] L. Stark and K. Bowyer, "Achieving generalized object recognition through reasoning about association of function to structure," *IEEE Trans Pattern Anal Mach Intell*, vol. 13, no. 10, pp. 1097–1104, 1991, doi: 10.1109/34.99242.
- [7] H. Grabner, J. Gall, L. van Gool, and E. K. Zurich Leuven, "What makes a chair a chair?," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2011, pp. 1529–1536. doi: 10.1109/CVPR.2011.5995327.
- [8] B. Yao, J. Ma, and L. Fei-Fei, "Discovering object functionality," in *Proceedings of the IEEE International Conference on Computer Vision*, Institute of Electrical and Electronics Engineers Inc., 2013, pp. 2512–2519. doi: 10.1109/ICCV.2013.312.
- [9] H. Wu, D. Misra, and G. S. Chirikjian, "Is that a chair? Imagining affordances using simulations of an articulated human body," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2020, pp. 7240–7246. doi: 10.1109/ICRA40945.2020.9197384.
- [10] S.-B. Ho, "A general framework for the representation of function and affordance: A cognitive, causal, and grounded approach, and a step toward AGI," 2022. doi: 10.48550/arXiv.2206.05273.
- [11] E. Kolve *et al.*, "AI2-THOR: An Interactive 3D Environment for Visual AI," 2022.
- [12] M. Deitke *et al.*, "ProcTHOR: Large-Scale Embodied AI Using Procedural Generation," 2022.
- [13] L. Talmy, *Toward a Cognitive Semantics Volume I and II*. Cambridge, Massachusetts: The MIT Press, 2000.
- [14] V. Evans and M. Green, *Cognitive Linguistics: An Introduction*. Mahwah, New Jersey: Lawrence Erlbaum Associates, 2006.
- [15] R. W. Langacker, *Cognitive Grammar: A Basic Introduction*. Oxford: Oxford University Press, 2008.
- [16] D. Geeraerts, *Theories of Lexical Semantics*. Oxford: Oxford University Press, 2010.
- [17] J. van Eijck and C. Unger, *Computational Semantics with Functional Programming*. Cambridge: Cambridge University Press, 2010.
- [18] R. Mitkov, Ed., *The Oxford Handbook of Computational Linguistics*. Oxford: Oxford University Press, 2005.
- [19] A. Clark, C. Fox, and S. Lappin, Eds., *The Handbook of Computational Linguistics and Natural Language Processing*. Hoboken, New Jersey: Wiley-Blackwell, 2012.
- [20] R. C. Schank, "Identification of conceptualizations underlying natural language," in *Computer Models of Thought and Language*, R. C. Schank and K. M. Colby, Eds., San Francisco: W. H. Freeman and Company, 1973, pp. 187–247.
- [21] R. C. Schank and R. P. Abelson, *Scripts, Plans, Goals, and Understanding: An Inquiry into Human Knowledge Structure*. Mahwah, New Jersey: Lawrence Erlbaum Associates, 1977.
- [22] R. C. Schank, *Conceptual Information Processing*. Amsterdam: North-Holland Publishing Company, 1975.
- [23] R. C. Schank and C. J. Rieger III, "Inference and the Computer Understanding of Natural Language," Stanford, California, May 1973.
- [24] X. Puig *et al.*, "Virtualhome: Simulating household activities via programs," in *IEEE International Conference on Computer Vision and Pattern Recognition*, 2018.
- [25] X. Gao, R. Gong, T. Shu, X. Xie, S. Wang, and S.-C. Zhu, "Vrkitchen: an interactive 3d virtual environment for task-oriented learning," 2019.
- [26] Y. Yang, Y. Li, C. Fermuller, and Y. Aloimonos, "Robot learning manipulation action plan by 'watching' unconstrained videos from the world wide web," in *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*, 2015, pp. 3686–3692.
- [27] C. Li *et al.*, "igibson 2.0: Objectcentric simulation for robot learning of everyday household tasks," in *Conference on Robot Learning*, 2021.
- [28] H. Fu *et al.*, "Rfuniverse: A physics-based action-centric interactive environment for everyday household tasks," 2022.
- [29] F. Xiang *et al.*, "SAPIEN: A simulated part-based interactive environment," in *IEEE International Conference on Computer Vision and Pattern Recognition*, 2020.
- [30] T. Inamura and Y. Mizuchi, "Sigverse: A cloud-based VR platform for research on multimodal human-robot interaction," *Front Robot AI*, 2021.
- [31] C. Gan *et al.*, "Threedworld: A platform for interactive multimodal physical simulation," in *Conference on Neural Information Processing Systems*, 2021.
- [32] S.-B. Ho and F. Liasuvia, "A ground level causal learning algorithm," in *Proceedings of the IEEE Symposium Series on Computational Intelligence for Human-like Intelligence*, 2016, pp. 110–117. doi: 10.1109/SSCI.2016.7850025.
- [33] X. Yang and S.-B. Ho, "Learning correlations and causalities through an inductive bootstrapping process," in *Proceedings of the IEEE Symposium Series on Computational Intelligence*, 2018, pp. 2270–2277. doi: 10.1109/SSCI.2018.8628932.
- [34] S.-B. Ho, X. Yang, and T. Quieta, "Achieving human expert level time performance for Atari games - A causal learning approach," in *Proceedings of the IEEE Symposium Series on Computational Intelligence for Human-like Intelligence*, 2020, pp. 449–456. doi: 10.1109/SSCI47803.2020.9308301.
- [35] S.-B. Ho, M. Edmonds, and S.-C. Zhu, "Actional-perceptual causality: Concepts and inductive learning for AI and robotics," in *Proceedings of the IEEE Symposium Series on Computational Intelligence for Human-like Intelligence*, 2020, pp. 442–448. doi: 10.1109/SSCI47803.2020.9308168.
- [36] A. Fire and S.-C. Zhu, "Using causal induction in humans to learn and infer causality from video," in *Proceedings of the 2013 Annual Meeting of the Cognitive Science Society*, 2013.
- [37] A. Fire and S.-C. Zhu, "Learning perceptual causality from video," *ACM Trans Intell Syst Technol*, vol. 7, no. 2, pp. 1–22, 2016, doi: 10.1145/2809782.
- [38] S.-B. Ho and Z. Wang, "Language and robotics: Complex sentence understanding," in *Proceedings of the 12th International Conference on Intelligent Robotics and Applications*, 2019, pp. 641–654. doi: 10.1007/978-3-030-27529-7_54.
- [39] S. Ruan, K. L. Poblete, H. Wu, Q. Ma, and G. S. Chirikjian, "Efficient path planning in narrow passages for robots with ellipsoidal components," *IEEE Transactions on Robotics*, vol. 39, no. 1, pp. 110–127, Feb. 2023.
- [40] S. Ruan, X. Wang, and G. S. Chirikjian, "Collision Detection for Unions of Convex Bodies With Smooth Boundaries Using Closed-Form Contact Space Parameterization," *IEEE Robot Autom Lett*, vol. 7, no. 4, pp. 9485–9492, Oct. 2022.