Something Something Game Something: A Visual Approach to Definitions in Games

Johnathan Harrington
City University of Hong Kong

Introduction

In this paper, I will be looking at the viability of using games, or playful interactive artifacts as a basis of conducting philosophical research with valid scientific output. By conducting a side by side non-ludic experiment, we can evaluate what the advantages and disadvantages of using a more visually inclined, ludic experiment could be over more traditional textual experiments.

Background

Before I start discussing the philosophical output, or perhaps even the philosophical questions raised by this paper, I think it is important to give a brief introduction to the game, and to why I am standing up here talking about it.

Something Something Soup Something (SSSS) is a game designed by Stefano Gualeni, a person who most of you probably know from previous papers in this conference series. As many of you also know, he is usually the person talking about how game design can be a transformative practice, in many different senses, such as game design as critical reflexive practice (2016), game design as situated knowledge (2016), game design as self transformation (2015), game design as self liberation (2014) and as his most recent monograph discusses, game design (and virtual worlds at large) as tools for philosophy (2016).

SSSS continues to work on this idea, except this time I was also involved in it, which is why I am the one standing up here presenting this paper, as opposed to Stefano. SSSS is a game where you play as a kitchen handler, receiving food items from a distant space colony, populated by aliens who do not have the same conceptions of language that we do. The concept of soup is alien to them, so humans before you collated formal properties as a basis for their definition. However, both due to them still not sharing a similar cognitive framework as well as space interference in the radio signal, the soup delivered can be sometime less than ideal. The player's job, as the player, is to determine whether their delivery is soup or not soup, using solely your personal judgement.

This is how Stefano originally pitched the idea to me, and how I think it faithfully stands. Some things here and there have changed (we discussed making it about sandwiches, briefly), but the general gist of the game remains. We wanted to explore a host of philosophical ideas, such as post-colonial structures within video games. However, the most important point that we wanted to tackle was definitional fuzziness as presented in a digital game, as opposed to an academic paper, a pertinent question even to out our field, with recent papers such as Arjoranta (2015) and Aarseth and Calleja (2014) still tackling the issue, just to name a couple.

To name just one more paper exploring this definition fuzziness, in this game's inception, I had just submitted my masters thesis, using cognitive linguistics, specifically Rosch and Mervis' work, as a basis to discuss why the definition of games is still very fuzzy. So I had Very Strong Opinions on Stefano's game, specifically about how he approached the methodology of determining soupness, and our lack of understanding of it. So he recruited me as a Field Researcher (which is a title I am very fond of; it makes me sound like I went around eating soup, for science). More importantly, the title puts methodology approaches squarely on two entities: time constaints, and me.

Since then, the game has garnered considerable mainstream attention, with outlets such as *Kotaku*, *Vice Waypoint*, and *Atlas Obscura* picking it up and discussing it in detail. More importantly, it has led a lot of people to discuss soupness in the way we expected them to - with a lot of confusion. For example, Kotaku's article written by Nathan Grayson had this image, accompanied by the subtext "I'll go back and forth on whether or not this one is technically soup until the day I die." There was a disparity about what constituted a soup and what he felt was a soup.

This disparity is not an unfamiliar one in philosophical discussions around definitions. There are often discrepancies between real definitions and nominal definitions, and soup does not present any exception. Our question then becomes whether this discrepancy manifests itself differently when the method of definition forming changes from a textual one to a visual one.

Theoretical Background

To do this, we started off by running experiments based off of Rosch and Mervis' definitions work. Eleanor Rosch, later accompanied by Carolyn Mervis, applied Wittgenstein's work on definitions within cognitive linguistics. Apart from field appropriation, their work also led to creating some more scientific clarity to Wittgenstein's work. With Wittgenstein's family resemblances, we understand that despite a family member not having the distinctive Roman nose, they can still be biological members of the family. However, with Wittgenstein, we perhaps cannot fully understand why the Roman nose is a distinctive family feature. We cannot say with full certainty what the Harrington family looks like, even though we full well know what the Harrington family generally looks like (they look like me). We just know that despite my sister being darker than the average Harrington, she is still my sister.

Rosch and Mervis help clarify this with prototype theory. In this theory, they explain how when we are judging whether a new subconcept fits within our cognitive model of its family concept, we are basing it off of a prototype, a cognitive median of all the things we associated with that concept before we were introduced to this new subconcept.

For example, let us consider being asked about the 101st bird (assuming birdness is set in stone, thanks to biology) that we would see in our life. When we are being asked what a bird is, we may be very tempted to say that a bird is an animal that flies, despite not all of them flying. If we are presented with a non-flying bird for the first time (such as an emu, a kiwi or a penguin), we might choose to dismiss these as non-birds, and this would not be necessarily false within our cognitive model. If we are presented with a non-flying bird for the second time, we might still choose to dismiss these as non-birds, due to their low-level membership within our cognitive model – non-flying birds would have a sub 1% incidence. However, this would no longer be true within our cognitive model, as we have had at least one incidence of a non-flying bird. Yet Rosch and Mervis explain that we would likely still dismiss it as not a bird, as we form definitions based on our prototype, which aggregates the features most often present, rather than the features that might be present. If we were presented with our 500th non-flying bird (and no new flying birds), we would very likely say the opposite - that birds tend to not fly.

If flight is the defining feature within the bird word-family, something such as a Sparrow, an Eagle, or a Hawk would have a high-membership rate, as these are often perceived to be flying. Something such as a Dodo, a Kiwi, or a Kakapo, would have a low-membership rate, as they don't fly. Birds that can fly, but are not associated with flying, such as perhaps Chickens, Peacocks or Turkeys would perhaps mid-level members. However, this cognitive mapping of birds is purely speculative - Rosch and Mervis ran studies about furniture, vehicles, fruit, weapons, vegetables and clothing in perhaps their most influential paper. Based on their study, we ran one on soup.

Another concept which they explain which I will be making ample use of within the experiment discussion section of this paper is word categories. They discuss how a word can belong into three different types of word categories: superordinate categories, basic level categories, and subordinate categories. Definitional fuzziness tends to occur in superordinate categories, as this category contains a high level of generability and a wide degree of membership. This means that when we discuss superordinate category words, we are not very exclusive. On the other hand, subordinate categories would have a low level of generability and a narrow degree of membership. This means that when we discuss subordinate categories, we are very exclusive.

If we were to put this on a spectrum, we could argue that the concept games would lean towards superordinate categories, while Super Mario Bros would be the tightest example of a subordinate category (since it still includes different members, such as Super Mario Bros on the NES and Super Mario Bros on an emulator). Somewhere along this spectrum, we could find Main Franchise Super Mario Games, which I would argue still remains within a subordinate category and Platformer Games which would perhaps be somewhere within basic level categories. As you can see, categories as a model is still nominal benchmarks, but it allows us to explain concepts better. In this presentation, I will not make arguments for games or indeed soup belonging to any word category, but I will be making claims of higher level category, which means that the object in question is more like a superordinate category, and lower level category, which means that the object in question is more like a subordinate category.

Text-Based Experiment

Based especially on Rosch and Mervis' 1976 experiments, we ran an experiment to determine what a soup is. I adapted Rosch and Mervis' earlier experiments a little bit to suit our circumstances. While Rosch and Mervis had constant access to new groups (allowing her to subdivide her experiments into multiple parts), we were not as blessed. Additionally, our experiments were also being run as a basis for game design, which I will explain in more detail later on.

We ran one experiment, on 4 different focus groups of 10 or 12 people each, to a total of 44 respondents. While we did not explicitely employ quota sampling, we had a very healthy mix of respondents, with an almost equally split male / female divide, and an age range spanning from 18 to 54 years. One of the above groups was also a control group that was explicitely told what we were testing – they were all linguistics students who had studied Rosch and Mervis' theories. This was done in order to see whether there was any difference in results when respondents knew their cognitive frameworks were being tested, which from our control, did not seem to be the case, although I will make a few references to the control throughout the presentation.

Our respondents were put into pairs, each being shown a word – soup or bird – which they were to keep hidden from their partner. They were then instructed to list down all the components that they felt defined that object within a minute and a half. This means only half our respondents (22) were involved in this part of the soup data collection (since the other half were writing about birds). In this part, we were primarily looking at formal properties of soup. We wanted to see if there was a pervasive formal property in all these soup definitions, which gave us the following results

Commonly listed properties
Liquid: 41% (13 type incidence)
Edible: 32% (24 type incidence)
Spoon Usage: 18% (4 type incidence)

Bowl Usage: 18% (4 type incidence)
Hot: 9% (4 type incidence)

The difference between property and property type is that the former lists the formal property exactly as they listed it, while in the latter, I made judgement calls to include certain properties together. This is why "commestible" has a large incidence rate than the number of respondents – there were 6 properties that were included in this property typing: edible, eaten with a spoon, eaten, can be digested, food, eaten when it is cold. Some respondents listed more than one of these property typings. Some type of commestibility was mentioned all but one time, so the closest we can get to a property that spans across would be commestibility with a 95% incidence rate. Rosch and Mervis noted a similar phenomenon in their experiments; the only items with a 100% property incidence were fruit and vegetables, with the formal property "commestible" being given.

From this experiment, if we were to exclude commestibility, soup seemed to have a lower generability rate to birds, with the latter holding 5 different formal properties with a 50% incidence rate: feathered, pedal, beaked, winged, and lays eggs. I would speculate that this is either because soup is in a higher word category than birds (meaning that it holds many more differing members than birds would), or because real definitions for birds have been more rigidly developed thanks to biological typologies.

In the second part of the experiment, we made the respondents switch papers with each other and, based on the formal properties that the other person listed, they were to guess which item their partner was describing. Again, this means that we had 22 results, half of our respondents were trying to guess bird as opposed to soup. In this part of the experiment, we were trying to establish the closest non-category members, or more clearly, things that are not soup but are very close to being soup.

Over 50% of the respondents managed to guess that their partner was describing soup. The rest of the results were scattered, with 2 respondents guess either coffee or water, with single respondents guessing pudding, bottle, drinks, rice, and lemonade. This experiment which tried to establish the closest non-category members seemed to focus on the liquid element of soup, with the exception of rice (and perhaps bottle). This shows that while being liquid is something that soup often is, it is not a core feature only within soup, which would suggest that it is perhaps not the formal property we use to distinguish soup from not-soup.

This experiment would also suggest that soup is a higher category member than birds – the mirroring experiment showed that all 22 respondents guessed either bird, or a lower category member of the bird family, such as eagle or chicken. Meanwhile, no respondent guessed a type of soup as their answer. The formal properties of birds are much more descriptive of birds, and the members of birds are more clearly defined than the members of soup.

This would be confirmed in the third, and final part, of our experiment. In the last part, users were told to swap the paper one final time, while folding the paper in a way that hid their guess. The original formal property writer would then have half a minute to list three different types of soup or bird, and another half minute to list three different types of their partner's guess. If their partner guessed correctly, then they would null the second half of this experiment.

The soups listed were various and often unnamed. The soup mentioned most, at 7 times, was chicken soup, which might not necessarily immediately evoke an image of a particular type of soup – it is a soup with chicken in it. A creamy French style chicken soup is very different than a Cantonese chicken bone soup. The named soups most often mentioned were Minestrone and Gazpacho, both appearing at 5 times each. The rest of the named soup appeared only twice, such as Aljotta, or once, such as Broth, Consommé, Clam Chowder, Goulash, and Veloute.

As I explained above, this seems to indicate that the members of the soup category are ill-defined, especially compared to bird. Respondents were very happy to name lots of different types of birds, including Eagle at 6 times, Sparrow at 6 times, Chicken at 5 times, and Duck at 4 times. The only non-named results we received for bird were Birds of Prey, which appeared twice, while Hunting Birds, Flying Birds, Migrant Birds, Tropical, and Birds in a Soup, appeared once. This shows a complete opposite trend from soup.

Ludic Experiment

We ran this experiment for a few reasons. Firstly, we were trying to understand better what a soup is, and whether it was a valid contender for definitional fuzziness. If respondents were very clear as to what a soup is, then running a secondary experiment which would take considerably more time to execute, would not be particularly useful.

However, as Reviewer No. 2 correctly pointed out, there is a distinction between pository definitions and negatory definitions. The text-based experiment was a mixture of both. The first part was pository – we gave people a word and asked them to describe what it is. The second part of the experiment was negatory. We gave people formal properties and asked them to tell us what they thought the object being described was.

The game we designed was only negatory – we gave users procedurally generated soup, and they had to decide whether it was or was not a soup. The soup was procedurally generated by making arrays of the most common formal properties, which we listed above. These contained at least one confirmatory, one negatory item, with some containing middle ground sections. For example, within the commestible section, we had edible food (carrots), edible

but harmful (poisonous mushrooms), inedible but harmless (a cocktail umbrella), and inedible and harmful (batteries). Edible food confirms the formal property, inedible and harmful negates the formal property, while the other two sections do neither, but shed a bit more light about what constitutes commestibility – whether it is capability of eating or advisability of eating. Each section of each formal property meets each section of each other formal property at least once. So if there are three sections for Bowl Usage (bowl, bowl-like, not bowl-like), each of these three sections would be paired up at least once with each of Commestibles four sections. Likewise with the other formal properties.

Since the game was only negatory, it did not replicate the entire text-based experiment, but rather, it replicated the second part of the experiment. More importantly, it replicated this part of the experiment much better than the text-based version for various reasons.

Advantages and Disadvantages

Firstly, in our text-based experiment, the respondents had to decide whether the formal properties constituted a soup or not based on their partner's idea of what a soup is. Needless to say, some of the formal property lists received from their partners were less than ideal. This could have been avoided if we simply gave the respondents a list of definitions ourselves. However, this would stil remain as an individual definition, as opposed to the social definition that the ludic experiment provides. In the game, our procedural generation was based off of a pository definitional exercise of 22 respondents, but this could be scaled up. Rosch and Mervis' original experiments had a much larger sample size (>100).

The second advantage that the ludic experiment gives us, over the text-based one is that cross-combining all of the formal properties listed is much easier when it is procedurally generated by the game's code, over the haphazard method employed when a resondent is given another respondents list of formal properties. As mentioned in an earlier paragraph, each possible variant of each formal property is matched at least once with each other variant of each formal property at least once. This allows us to not only record which formal properties are highly evaluated in terms of soupness, but also record whether certain formal property matches are more highly evaluated than others. So while liquidness might have a certain incidence, liquidness tied with commestibility might have a higher incidence. While this is also recordable in the text-based experiment, the ludic experiment allows us to replicate it consistently every time.

A third advantage is the project's adaptability and scalability. If we find that a certain formal property intersection is consistently dismissed (for example, a <10% incidence), then we can exclude it from the procedural generation easily. Similarly, if we find that a certain formal property intersection has very varying results between one rendition of the intersection and another, then we can make introduce even more renditions to this intersection to determine which of these intersections is an outlier. With the text-based version, we would have close to

no insight to either of these, especially the latter. For example, to make the latter example clearer, if + Liquid and + Meat were given as a combination, in a text version, one person might conceive of a stew, while the other person might conceive of a ramen, because of the mild ambiguity of both liquid and meat. They might both agree that stew is not soup, but ramen is. However, they might answer differently because of the first projection that comes to them. In the ludic experiment, this would be less of an issue, as they would answer to a visual referrent with clearer liquid and meat referrents. And if we see a large discrepancy between stew results and meat results, we could instert a beef congee in there or a chunky chicken soup, and see whether stew or ramen is the outlier.

While the advantages of the second experiment are clearer, these advantages are moreso related to the digital computational aspects of the medium, rather than the ludic ones. While the advantages are not as recordable as the digital computational ones, we can still hypothesise on certain advantages, at least to create space for discussion questions.

One advantage that a ludic setting might provide is the removal of the experiment setting from the experiment. From my observation, presenting the experiment as a game that collects data, as opposed to a data collection exercise made respondents less worried about giving us the correct type of results. If anything, correctness is shifted towards proper gameplay (what will "win" me the game), rather than proper results (what will produce the right data). Once the game ends, respondents were often surprised to see how vague their definition of soup was, which is a good thing, because it shows that they were not actively thinking about what a soup is during the game (and realising that their idea of soupness is vague). What would need to be tested, relating to this shift, would be whether trying to "win" a game can similarly affect results as trying to provide proper data. More side by side experiments would need to be conducted, with their data analysed.

Another advantage that a ludic setting might provide is how easily it can slip into mainstream consideration. Earlier on in the presentation, I mentioned some gaming outlets that picked the game up, just because they found it interesting, which caught me (and Stefano) quite by surprise. Sadly, we did not expect it to spread so much, and so our data collection is only local, but it does set a future consideration. I do not feel that a visual digital computational experiment in cognitive linguistics would be so readily picked up if it did not have this ludic element within it. While we are still in the realm of the speculative, there are precedents, such as Bogost's *Cow Clicker* and McGonigall's body of work that give some credence to this claim. Apart from allowing us to collect research at a much faster rate, it also allows our research to reach a larger audience, in a valid and accessible manner.

Moving Forward

There is a final advantage of putting the experiment within a ludic setting, which is considering whether the questions we posited for soup can be posited for games as well. In our experiment for soup, as we have pointed out, we have found that shifting from a textual

pository perspective to a textual negatory perspective changes the way definitions are formed. We have also seen that shifting from a textual negatory perspective to a visual negatory perspective also alters the way definitions are formed.

It is clear that the final part of our experiment could not be replicated like-for-like with digital games for a variety of reasons, with the biggest one being games' resistance to being reified. Displaying a frame or a shot of a game is a very different phenomenon from displaying a game in process, while a shot of a soup and a soup in process are not necessarily different. As Reviewer 1 pointed out, digital games are also not necessarily visual in scope, while soup necessitates being visual, which is why the closest point of comparison for our game would be Pedercini's *The Definition of Game*, which reverts to a textual negatory approach.

However, there is still merit to the exploration of digital game definition outside of the academic real definition work, or the nominal reactions to these flawed formal-property oriented constructions. While the experiment does not carry over, the results do.

In previous work (Harrington, 2017), I argue for an application of Rosch and Mervis' theories to the study of digital games, explaining the necessity of formal definitions to establish what the prototype is, considering category membership of specific games and specific types of games, and positioning games within the superordinate category to explain why reactions to the real definitions are so pervasive within digital game academia. There is a space for real definitions within games, as long as we accept that properties are by degree and to a certain degree. Our next step then becomes to figure out this degree.

The progression that this paper, along with our soup oriented case study, provide to this line of thought,is that this can be achieved through a different process of definition formation. Firstly, there is value in applying a collective method to the creation of definition over the very strict academic approaches that we are approaching. Accumulating previous academic texts as a basis for our definition or using specific language case studies to disprove these academic collations of definitions is not a valid method of collective definition formation. Since Wittgenstein (or rejections of Wittgenstein) seems to be one of the more popular methods to construct definitions, then there must be due rigour used to seeing language as social, not individual.

However, apart from using more social-oriented formations of definitions, SSSS also shows that there is some merit to alternate approaches to acquiring the definition. Utilising Rosch and Mervis' textual negatory approaches over the current attempts of textual pository approaches could already render different results and deeper insight. Had Pedercini's game been more than just commentary and applied better methodology, valid results could have been generated. Similarly, finding a way to alternate the predominant sense used to form a negatory definition, as we did in SSSS, could also equally render valid alternate results for us to consider.

Games

SOMETHING SOMETHING SOUP SOMETHING. Gualeni, S., PC, 2017.
THE DEFINITION OF GAME, Pedercini, P., PC.

References

Aarseth, E. and Calleja, G (2014) 'The Word Game: The Ontology of an Indefinable Object',

Academia.edu [Web: https://www.academia.edu/15504517/The_Word_Game_The_Ontology_of_an_Indefinable_Object, Accessed 25th December, 2016]

Arjoranta, J. (2014). 'Game Definitions: A Wittgensteinian Approach', *Games Studies, vol. 14, issue 1.*

Rosch, E. (1973) 'Natural Categories', *Cognitive Psychology*, 4(3), pp.328-350.

Rosch, E. and Mervis, C. (1975) 'Family Resemblances: Studies in the Internal Structure of Categories', *Cognitive Psychology*, 7, pp. 573-605.

Wittgenstein, L. (1986) *Philosophical Investigations*, trans. by G. E. M. Anscombe. Oxford: Blackwell Publishing.