

# Week Two: Psycho-Analysis

## 1. Part One: McNamara and Sternberg on Models of Meaning

### *Competing Models of Meaning*

McNamara and Sternberg were aiming to test, experimentally, three models about the nature of concepts. The three models weren't really on a par, so they were comparing apples and oranges a bit from the start (though to be fair, they sort of realised this).

The first is our old friend the **classical model**. They attribute the classical theory to Russell and Frege, and in particular to "On Denoting" and "On Sense and Reference". Now while Russell and Frege would have held something like the classical theory, one will have to interpret their words quite freely to find it in these papers. Indeed, it isn't easy to find an explicit endorsement of the classical theory in classical works<sup>1</sup>. I thought it would be in the *Tractatus*, or in *Language, Truth and Logic*, but neither are particularly explicit in their endorsement of it. Interesting question for those who like historical research: who does explicitly endorse the classical theory?

The second is what they call a **characteristic-attribute model**. I suspect their presentation on pg 450 is a bit too classically orientated. The basic idea is that these are cluster concepts. For example, there are no necessary and sufficient conditions for being a game, as Wittgenstein pointed out, but we can list properties which most games have, and it seems sufficient to be a game to have a certain number of these properties. (It is important to note here that the classical theory should be restricted to saying the necessary and sufficient conditions are not too gruesome, or we will be able to restate Wittgenstein's theory as saying the conditions for being a game are just disjunctive. This is the colonising tendency of the classical theory which I mentioned last week, and think should be resisted.)

The reason I think the presentation is a bit too classical is that they think that for one of these models to work, we should be able to say precisely how the properties in the cluster interact. So they want to test whether an individual falls under the concept by having a certain number of the properties in the cluster, or having a weighted sum of the properties, or having the average weight of the properties they have be sufficiently high, or some such. Now I don't think any non-classical theorist should be committed to saying one or other of these is correct, they should (and usually do) say that (a) how the properties in the cluster interact will be different for different words and (b) there is simply no fact of the matter as to which is the right theory for some words. Remember that Wittgenstein's claim that 'game' refers to a cluster concept appears within a few paragraphs of his claim that the term is massively vague in the *Investigations*. In general there is a close connection between cluster concepts and vagueness, so the assumption here that there will be an answer as to which type of cluster concept theory is right seems to be a mistake.

---

<sup>1</sup> Actually, by 'classical' here I mean anything prior to, roughly, Jack Smart's early works. I suppose on this terminology *Word and Object* counts as the high point of early modern philosophy.

The third are what they call **mixed models**. Saying these are all different from classical models seems to commit what I called the Red Sox fallacy. If “defining attributes provide the necessary and sufficient conditions for category membership, whereas characteristic attributes determine the prototypicality of the members,” then the category is classical. But they say this is just their first kind of mixed model. The reference to Rosch and Mervis should not be read as a claim that Rosch and Mervis endorse this view, but rather as a reference for the concept of prototypicality.

The notion of mixed model is a bit broader than this. As the reference to Clark and Clark shows, they take an exemplar-based account, where the exemplars are picked out by defining attributes (presumably rather than by ostention) and the intension of the concept is the set of things sufficiently similar to these exemplars, to be mixed. Now I don't know how many people who base their theory of concepts on exemplars take them to be picked out ostensively, so I don't know how different this is to the standard characteristic-attribute model. (You might think the other option, having exemplars picked out by ostention, is implausible because we do share concepts but don't share exemplars. But possibly this problem can be avoided provided there are sufficiently many exemplars, and that they overlap sufficiently. See “Naming the Colours” in **Lewis** for one attempt to resolve such difficulties.)

And as far as I can see, even this mixed model may still turn out to be classical. It will depend a bit on what we take to be too gruesome. Assume, for the sake of the argument, that ‘water’ was a bit more classical than it is. So rather than being the stuff of our acquaintance which has most of the following properties: falls from the sky, fills the rivers and oceans, is potable, drinkable, transparent, and so on, ‘water’ was defined as the stuff of our acquaintance which fills the rivers, is drinkable and transparent. To be an exemplar of water, something has to be acquainted with us, fill some river, be drinkable and transparent. We could still, by Twin Earth kinds of considerations, say that water is really H<sub>2</sub>O, because the exemplars are all H<sub>2</sub>O. Now this is just Clark and Clark's model, but this (fictional) story about ‘water’ is entirely consistent with the classical theory. So I can't quite see how either the mixed models are non-classical.

The goals of the experiments are outlined on page 451. Note that they are not presuming at the start that the same model will be appropriate for different classes of words, though they do seem to place a great emphasis on proper names. Note also that they are interested both in the meanings of terms and in the criteria subjects use to judge whether an object falls under the term. I don't think the sentence about Kripke at the bottom of the left-hand column is even close to being right, but that doesn't seem too important in the overall scheme.

### *Experiment One*

The experimenters chose eight words from each of three categories: natural kind terms, artifact terms and proper names. The words used are listed on pages 471-473. One of the proper names was of a fictional character; another (Queen Elizabeth) seems very close to a description. The subjects (all 10 of them!) first listed attributes of exemplars of each of the words, being given one minute for each term. (Can you have an exemplar of a proper name?!) After

the lists of compiled attributes were compiled, the same subjects were asked which of these attributes were necessary to satisfy the term in question, which were sufficient, and which would be used to determine whether something was an exemplar of the term in question.

Given the way the experiment was set up, it doesn't seem too surprising that subjects came up with quite a few necessary conditions for satisfying these terms. But given the kind of necessary attributes listed, the semantic or philosophical importance of this project seems kind of dubious. As is pointed out on page 453, the necessary conditions for being Einstein include: being a scientist, inventing  $E=MC^2$ , and being dead. Well, I suppose that deals with possible conspiracy theories about Einstein still being alive and living in a hideout with Elvis and Nixon. The conspiracy theory becomes inconsistent, since Einstein is necessarily dead!

Now it might be possible that there are necessary and sufficient conditions for being Einstein. That is, some kind of descriptivism about proper names is correct. But given the arguments raised by Kripke and others for a direct reference theory, we know that the attributes will need to be phrased using quite technical language. In particular, they will need to refer to causal connections between the person named and users of the name, which generally hold in the actual world, but with qualifications for empty names<sup>2</sup>. In other words, these are not the kinds of attributes of named people one would usually think of when asked to list properties of a person within 60 seconds. So we can be pretty sure that any attribute generated using this experiment will not be a necessary attribute of objects satisfying some proper name.

Two interesting questions: first, why did the subjects go so badly wrong when discussing proper names, and secondly, do these mistakes mean that what they say about artifacts or natural kinds should be discounted? I suspect the right answer to question two depends on the right answer to question one, so we'll look at that first. Two answers suggest themselves. First, the subjects may have felt dumb answering "None of the above" too often, but since all the proposed answers were wrong, they put down wrong answers. (The answers were all wrong because the generating procedure made getting correct answers impossible.) Secondly, the subjects might just not have known how to think about modality correctly. Before Kripke it wasn't so unusual to hear things like "Invented general relativity" proposed as necessary attributes for people like Einstein. What was most valuable about *Naming and Necessity* was not that it showed a particular brand of descriptivism was wrong, but that it gave us new and improved ways of thinking about what is and isn't possible. Someone who hadn't read Kripke, nor any of the subsequent Twin Earth chronicles, might just not make the right kind of conceptualisations.

What does this mean for the other data? You might think that it means not a lot; inferring from an error on one task to errors on other tasks is just guilt by association. But that's not the argument I'm appealing to. Rather, I'm claiming that since the experiment goes so badly wrong somewhere, this shows by *modus tollens* that at least one of

---

<sup>2</sup> This kind of descriptivism is defended in the Kroon paper in the reading guide, and by Lewis and Jackson in various footnotes in the books I have recommended. (The Lewis footnote is in "Naming the Colours".)

the assumptions behind the experiment is mistaken. If subjects approached the discussion of proper names in a radically different way to how they approached the discussion of artifact terms, then perhaps the assumptions needed for different parts of the experiment are different, so this little *modus tollens* won't go through. On my first explanation of the errors subjects made, the false assumption would be that the necessary and sufficient conditions for the applicability of some terms will be included amongst the attributes listed in thinking for 60 seconds about exemplars of that term. On the second, it is that subjects have a working understanding of necessity and possibility. The first assumption is false for proper names, but is a starter for other categories. However if the second assumption is proven false by the experiment concerning proper names, the whole experiment is in trouble.

This kind of analysis of error is the main point I wanted to take from McNamara and Sternberg's paper. When thinking about the kind of attributes they discuss, remember the sampling method, and take into account what effects this might have on subsequent discussions. Also, remember that folk make philosophical errors. Don't take the word of an undergraduate on a tricky philosophical topic.

#### *Experiment Two*

This experiment was designed to test something quite different to the nature of concepts, although it doesn't seem clear that McNamara and Sternberg recognised this. It was designed to test how confident people were that an arbitrary object was an *F* given that it had some other features *G*, *H* and *I*. Now it isn't at all clear what this has to do with reference. I suppose you saw this immediately, but just to spell it out, let's change their example a little.

We all know roughly what an atheist is, we can all give a clear definition of an atheist. And that definition will make no mention of the Presidency of the United States, nor any other country for that matter. Now suppose I tell you that *x* is President of the United States, what is the appropriate doxastic attitude to take towards "x is an atheist"? Presumably, to be very confident that it is false. Does this show us that something was wrong with our definition? I can't see how. So it is hard to see how this kind of test should tell us anything about meaning or reference.

#### *Experiments Three and Four*

These were basically repeats of experiments one and two with some bells and whistles added. The whistle in experiment three was that they added negative properties as well as positive properties. So it was possible for the subjects to say that Paul Newman was necessarily not dead. (Good news if you're his life insurance agent.) The bell in experiment four was that the subjects were timed while working out which set of attributes made for a better exemplar of a target concept. This experiment seems rather useless semantically for the same reason that the experiment two was. And experiment three delivered the same mistaken responses as experiment one, so I don't want to talk about that too much.

In fact to close this discussion, I just want to note one astounding result of experiment three. As in experiment one, the subjects tended towards some kind of Searlean descriptivism. So having the most prominent properties  $x$  actually has tended to be necessary and sufficient to be  $x$ . But this tendency towards identifying defining attributes wasn't just a carry-over from thinking about artifacts and natural kinds. The subjects were more likely to identify defining attributes of proper names than of artifact terms or natural kind terms. For what it's worth (probably not a lot) this difference was statistically significant. So the question of what mistake these subjects were making takes on some degree of importance.

## 2. *Part Two: Rosch and Mervis on Family Resemblances*

### *Rosch and Mervis's Experiments*

It is not usual in this field to take the psychological data to assist the classical view. As far as I can tell from a quick sweep of the literature, McNamara and Sternberg are about as positive as anyone about the classical theory. The orthodox position is that various experiments have soundly refuted the classical theory. One of the most famous such experiments, and the one we will spend the rest of this seminar discussing, was performed by Rosch and Mervis. The paper reporting their results is reprinted as chapter 2 of **Intuition**.

We will be mainly looking at experiment one. There are several other experiments reported in the paper, but these are less interesting for our purposes. The first experiment was designed to test the classical theory, and to show that it fails. The other five are, I think, designed to discriminate between various non-classical theories. Maybe, the consistency of the results generated is good evidence for the particular non-classical theory they endorse, and this matters if the argument for the classical theory is inference to the best explanation. But the last five experiments certainly don't seem to tell directly on whether the classical view is right.

At the very start (p. 17 in **Intuition**) Rosch and Mervis say that the classical view is opposed to the view that semantic categories can have graded membership. As I said above, I don't think this is right. Of course 'classical theory' is a technical term, so they can redefine it if they like, but the view that categories are characterised by defining attributes is consistent with the existence of graded membership.

The first sketch of a non-classical theory they provide is the Wittgensteinian 'family resemblances' model of categories. So a category might consist of objects of the form AB, BC, CD, DE. It is worth noting that this is only a non-classical theory if we have, pre-theoretically, a privileged class of attributes. Why can't the attribute, being a AB or BC or CD or DE be the defining attribute for this class? Well, I think because the classical view is committed to defining attributes not being particularly gruesome. I don't think this is a mistake in Wittgenstein (though I used to think that for many years, and still have some sympathy for my earlier view.) I do think this assumption of privilege should be made explicit, because I think it is quite relevant to what comes later. Note that when we get to

part two of the paper, experiments three and four, Rosch and Mervis are quite explicit that they believe in 'basic semantic categories', so they wouldn't be too distressed by the need for this assumption.

The main focus for experiment one is on what they call "superordinate semantic categories." The claim is that subjects rate these, "as having few, if any, attributes common to all members." (p. 20) A superordinate semantic category is one, like 'fruit', which has other categories, like 'apple', 'pear' and 'banana', as sub-categories. Here's the experiment they ran to show this. For each of six superordinate categories ('furniture', 'fruit', 'weapon', 'vegetable', 'vehicle' and 'clothing') they selected twenty category members. So for 'fruit' the members ranged from 'orange' and 'apple' to 'tomato' and 'olive'. They were (relatively) careful for each category that the members covered the range of prototypicality. That is, that they had some highly prototypical members, some very aprototypical members, and a reasonable spread in between. The lists for each category are on page 22. We might wonder for some categories whether the 'bottom end' of each category was a bit high. A bar stool is an even less typical weapon than a screwdriver but is (at least some of the time) a weapon. A pair of rollerblades could, I think, be considered a vehicle, and an item of clothing, but in each case would be off the bottom of this scale.

Each experimental subject was given a set of six members, one from each category, reflecting a range of prototypicality. For each of these six, they were asked to list all the attributes they associated with that object. They were given a ninety seconds for each member. Here's their summary of the result.

[F]ew attributes were given that were true of all twenty members of the category – for four of the categories there was only one such item; for two of the categories, none. Furthermore, the single attribute that did apply to all members, in three cases was true of many items besides those within that superordinate (for example, "you eat it" for fruit). (pg 23)

Two other results are worth noting. First, the more prototypical a member, the more attributes it had in common with other members of the category. The measure of prototypicality, by the way, was one they had generated in previous experiments. Secondly, sort of as a corollary of this, there were many attributes in common to the highly prototypical members of each category. This ranged from three attributes shared by the five most prototypical vegetables, to thirty-six (!) attributes in common for the five most prototypical vehicles. On the other hand the five least prototypical members had very little in common. The precise results are given on page 25. This seems to be another instance of the *Anna Kariena* law: there is only one way to be an *F*, but everything which is not-*F* is not-*F* in its own unique way. The importance of the second result is that they think it explains the intuitive plausibility of the classical theory. Restrict one's diet to oranges, apples and bananas, and it is quite plausible to think that all fruit have something in common. A more rounded diet, one including pomegranates, tomatoes and olives, will lead one to the proper Wittgensteinian conclusions.

One occasionally hears people deride the assumption that there are necessary and sufficient conditions for the application of a term, as if we had somehow assumed that the Brooklyn Bridge was up for sale. Really, this assumption is no more than the assumption that dictionaries can be written, and without any reason to the contrary, seems perfectly harmless. Maybe, the Rosch and Mervis experiments provide a reason to the contrary, a reason for thinking that the conditions of applicability for terms like 'fruit', 'weapon', and perhaps philosophically significant terms like 'knowledge' and 'justice' are Wittgensteinian family resemblance conditions, rather than traditional necessary and sufficient conditions.

Well, the experiments don't show this at all. For one thing, the proposal looks incoherent. The claim is that there are no necessary and sufficient conditions for being a weapon, but something is a weapon iff it bears a suitable family resemblance to paradigmatic weapons. Well, bearing a suitable family resemblance to a paradigmatic weapon is a condition, so it looks like we just have a very short list of necessary and sufficient conditions, a list of length one. Jackson makes a similar point in response to Stich's invocation of Rosch's experiments (**Jackson**: 61). This feels like it's cheating, and I think I can say why it should feel like cheating. Just in case some of you don't have Jackson's book yet, I have included the Stich quote and Jackson's response to it here:

The prototype for *bird*, for example, might include such features as flying, having feathers, singing and a variety of others. In determining whether a particular instance falls within the category, subjects assess the *similarity* between the prototype and the instance being categorized. However, features specified in the prototype are not even close to being necessary and sufficient conditions for membership. So, an animal can lack one or many of the features of the prototypical bird, and still be classified as a bird. Emus are classified as birds although they neither fly nor sing (Stich 1992: 249).

The puzzle is that the Roschian view [Stich] describes as opposed to the search for necessary and sufficient conditions is itself a view about the necessary and sufficient conditions for being a bird: as he himself describes it, the view is that being sufficiently similar to the relevant prototype is necessary and sufficient for being a bird. Moreover, Stich supports the view by pointing out how it fares on the method of cases, for he notes that the view correctly classifies an emu as a bird (Jackson 1995: 61).

(Does everyone get the feeling that this is an odd thing for Jackson to say?)

In some contexts, we only consider properties that are above a certain level of naturalness. If I claim two things, say my carpet and the Battle of Avignon, have nothing in common, I will not feel threatened by an objector who points out that they share some gruesome, gerrymandered property, like being elements of {my carpet, the Battle of Avignon}, or being things which I have just mentioned. Say that the best analysis of *F*-hood requires us to use predicates which are below the contextually defined border between the natural enough and too gruesome to use properties. Then there will be a sense in which there is no analysis of *F* into necessary and sufficient conditions; just

the sense in which my carpet and the Battle of Avignon have nothing in common. This, fundamentally, is the reason I think classical theorists should be committed to a restriction on the kind of properties which can go into analyses, because we implicitly adopt such a restriction in our everyday talk about properties.

Jackson's argument feels like a cheat because he just shows that there will be necessary and sufficient conditions for any concept provided we are allowed to use gruesome properties, but he makes it sound like this proviso is unnecessary. And indeed three pages later, near the bottom of page 64, he explicitly writes that it is unnecessary. If Rosch and Mervis's experiments show anything at all, it is that this is true of some common terms in some everyday-ish contexts. They go wrong by projecting truths of a particular context to all contexts. I doubt they show even that much, for reasons we'll touch on soon.

*Polemical digression.* That some concepts expressed by English words are cluster concepts in this sense is a coherent hypothesis, and for all I know it might be true. As an analyst, should I care? No; when I collapse into my philosophical armchair, I relax *inter alia* the standards for naturalness of properties. And I know I can do this without changing the meaning of analysanda, so I have every reason to believe analysis is always possible. And even if the best analysis of a concept into its necessary and sufficient conditions requires using rather gruesome predicates, the analysis might still perform a useful function. *End of digression*

Philosophers aren't particularly interested in terms like 'weapon', so Rosch and Mervis's experiments only have *philosophical* interest if the results can be shown to generalise to terms philosophers care about. In other words, if can be shown that terms like 'property', 'justice', 'cause' and 'knowledge' are cluster concepts, or family resemblance terms. But there is a good reason to think this is false.

As William Ramsey notes (in his paper in **Intuition**), if  $F$  refers to a cluster concept, then for any proposed list of necessary and sufficient properties for  $F$ -hood, it should be easy to find an individual which is an  $F$  but which lacks some of these properties. To generate such an example, just find an individual which lacks one of the proposed properties, but which has several other properties from the cluster. So say that  $x$  is really an  $F$  if  $x$  has any two of the properties  $A, B, C, D$  and  $E$ . And say my analysis of  $F$  is that  $x$  is an  $F$  iff it is an  $A$  and  $B$ . Then by finding something which is an  $A$  and  $C$ , and hence an  $F$ , you will have found a counterexample to my analysis. And this schema looks perfectly arbitrary. So for any cluster concept, it should be easy to find counterexamples to the **necessity** of the proposed defining attributes. And this is just what Wittgenstein does for 'game'.

Conversely, it should be harder to find an individual which has the properties without being an  $F$ . If the proposed analysis is even close to being right, then having these conditions will entail having enough of the cluster of properties that are constitutive of  $F$ -hood to be an  $F$ . Let's go back to the example from the previous paragraph,  $x$  is really an  $F$  if  $x$  has any two of the properties  $A, B, C, D$  and  $E$ . Unless my analysis is really way off beam, anything which I say is an  $F$  really will be an  $F$ . A 'way off beam' analysis would be saying  $x$  is an  $F$  iff it is an  $A$ , or saying  $x$  is an  $F$  iff it is an  $A$  and  $G$ . But these will be corrected very quickly. The kind of competing analyses which



will be left in the literature after the first few iterations of the counterexample game will all list **sufficient** conditions for being an  $F$ , but they will all be wrong because none of them will propose the correct necessary conditions.

After that bit of *a priori* theorising, let's look at what kind of counterexamples are commonly provided to proposed analyses. Without presenting a scientific survey, it's my feeling that in most fields, analyses are failing because they do not provide sufficient conditions, not because they are failing to provide necessary conditions. The most comprehensive survey for any field is Shope's wonderful little book on the investigations into 'knowledge' in the 1970s. He lists about 200 counterexamples to different theories. And almost all of them are cases which (purport to) show that the analysis has not provided sufficient conditions for 'knowledge', in conflict with what we would have expected was 'knowledge' really a cluster concept. In sum, even if Rosch and Mervis are right that English is riddled with terms picking out cluster concepts, this is consistent with there being some terms which pick out classical concepts. And this idea of Ramsey's seems to give us an operational test to see whether a philosophically significant term is a classical or a cluster concept, and running the test seems to show that some interesting terms are classical concepts.

'Seems', because Ramsey doesn't think it shows this at all. He thinks the application of cluster concepts is going to be highly context-sensitive. Let's change our earlier example a bit. Say that  $x$  is really an  $F$  iff it has the two most contextually salient properties of the set  $\{A, B, C, D, E\}$ , and my analysis is that  $x$  is an  $F$  iff it is an  $A$  and  $B$ . Now you can produce a counterexample to the sufficiency of my analysis by changing the context to make  $C$  salient, which can usually be done by telling the right kind of story, and producing something which is an  $A$  and  $C$ . If this is right, my simple operational test for a concept being classical or cluster will be ineffective, because there will be counterexamples to the sufficiency of proposed analyses of cluster concepts.

So maybe this test for some concept being classical or cluster doesn't work. There is a quite powerful argument that most every concept has to be classical, or at the very least that every category which has surprising elements must be classical. Making such categories non-classical makes some rather mundane facts quite inexplicable. In this experiment the subjects weren't told which category each member was in, but for other categories they were. Imagine, as seems plausible, one of the subjects objected to putting the member in that category. Many people, particularly undergraduates, don't regard olives and tomatoes as fruit. ("Fruit on pasta? How absurd!") When the student asks why is this thing called a fruit, other speakers can provide a response. It is not a brute fact of language that tomatoes are fruit. It is not just by magic that we happened to come to a shared meaning for fruit that includes tomatoes, and that if faced with a new kind of object, we would generally agree about whether it is a fruit. It is because we know how to answer such questions. This answer to the *Why is it called 'fruit'?* question had better be a sufficient condition for fruitness. If not, the subject is entitled to ask why having that property makes it a fruit. And unless there are very many possible distinct answers to this question, which seems very improbable, there will be a short list of necessary and sufficient conditions for being a fruit.

Now in this argument, at least, ‘fruit’ was relatively arbitrary. Really the only assumption about ‘fruit’ that I made was that there are things which are fruit which some folk don’t realise or know are fruit. So at the very least for any category for which there are such surprising members, there will be a short list of necessary and sufficient conditions for being an *F*, for pretty much any *F*. I really don’t know how non-classical theorists would explain to these folk why a tomato or an olive is a fruit.

Third objection, this one specifically about ‘fruit’. Rosch and Mervis’s experiments could not possibly show that many superordinate predicates in English are cluster concepts. For they would, if successful, show that ‘fruit’ is a cluster concept, and it quite plainly is not. So by *modus tollens*, there is something wrong with their methodology. Some of the other categories they investigate, particularly ‘weapon’ and ‘furniture’ *might* be relatively cluster-ish but not ‘fruit’. As the OED says, a fruit is **“the edible product of a tree, shrub or other plant, consisting of the seed and its envelope.”** If nothing like this is right, then we couldn’t explain to the sceptical why we call tomatoes, olives and so on fruit.

Some people deny that the OED is right in its definition. They say that the seed-envelope cluster for any plant is its fruit, even if it is not edible. So roses have fruit on this definition, but not on the OED definition. My best guess is that the term is ambiguous (or vague) between the OED definition and the definition which does not imply edibility. Interesting and hard question as to whether this should be classed as an ambiguity or as vagueness. Even if the term is vague, if it is vague between two classical definitions, that seems classical enough for me.

Perhaps my little *modus tollens* argument is wrong. Rosch and Mervis don’t need to be right about each of their claims for some of the others to be right. That is, they don’t need the assumption that the experiment leads to correct conclusions at all times for it to be a reliable indicator. If this is right, then even if they are wrong about ‘fruit’ or whatever, they might be right about other categories. Still, given that it is possible that a term can be classical and produce the kind of data they display, we should know to explain their data.

The simple explanation is that the defining attribute of ‘fruit’ is not one that would occur to a subject when given ninety seconds to think about attributes of oranges, bananas, pomegranates and the like. Subjects would, in general, agree that all 20 of the fruit they list have this property, that anything which has these properties is a fruit. I said that classical theorists should be committed to defining attributes having a certain level of naturalness. Now one possible level is ‘property which would occur to an experimental subject within the first ninety seconds of thinking about the attributes of one of its sub-categories’. But is this really a fair constraint? Have we changed the subject when we bring properties like ‘being the edible product...’ under discussion? I think not, so the classical theorist is playing fair here.

And now it should be clear why Rosch and Mervis’s results have little bearing on the truth of the classical theory. Whatever explanation we give of the results they get concerning ‘fruit’ will be transportable to the other categories. That is, whether the result was caused by sampling bias (not enough time to think of all relevant properties) or by mistaken beliefs about fruit (or ‘fruit’), the possibility of such an error means that the other results

may be similarly contaminated. Indeed, any plausible explanation of this error makes it look like a quite repeatable mistake, so all of the data is compatible with the classical theory.

### *3. For next week*

The main focus will be on the Stich paper in **Intuition**. There is a reading assignment on this paper, as noted on the first handout. The whole paper is interesting, and will be discussed, but my focus will be on the very ambitious section at the end attacking analytic epistemology. The other two papers are there to give you a better feel for the kind of work that goes on in this area. The Tye paper is a nice link to week four, where we will discuss the prospects for a plausible definition of physicalism.

Remember the reading assignment is due on the Friday before our next class, that is Friday September 24.