

# Cognitive-Linguistic Mapping

*How apparently severe disagreements can just  
be illusions, stemming from one-sided  
translation errors in two-way communication*

November 2011

## Part 0 – Preamble and introduction

*Wherein I apologise for writing this, excuse its vast length and admit that this is the only section written in the first person...*

*In the run up to November 2011, a few friends of mine decided to partake in National Novel Writing Month (NaNoWriMo for short) and I thought “Aha! There's something I could try”. The aim was simply to write 50,000 words in a month, November 1<sup>st</sup> to November 30<sup>th</sup> inclusive. Not a difficult task, but one that can occupy your mind quite a bit. Naturally, I wasn't interested in writing fiction. The idea of forcing out some prose without the ability to go and backtrack and review, or plan ahead and develop things slowly and naturally simply doesn't appeal to me. So I wanted to do some pop-science – the issue remained, however, what the fuck would I do?*

*The initial plan started to grow slowly, a semi-biographical tale about skepticism and skeptical methodology. Despite the grand vision, which may yet still be realised at a later date, it ended up growing too slowly and never properly took off in the time I had. This is why I don't think it's possible to write something good when in a rush. By the time the end of November approached I didn't have too much in the way of words. Oh well, perhaps call it a moral victory. But at the same time, I found myself reading about and thinking about and even developing my own ideas about communication, words, literature and how they relate to*

*skeptical and rational discourse. These ideas were undoubtedly not original – as almost every time I've looked on Less Wrong recently I found another post that actually agreed with me and exemplified what I was thinking. One or two of those may have even slipped through into parts of this document. So I had spent November not writing my pop-skepticism novel, but actually formulating another idea entirely. By around the 28<sup>th</sup> I found myself in the position to start writing it down. A few hours later I hit 8000 words of tripe and had more to add and parts to review and alter. And so, in less than a week I'd approached half the word count of a NaNoWriMo entry – it's just a shame that some of this fell inside December; otherwise I may have “officially” won the game by a nice and comfortable margin.*

*Why so long? Especially when what I'm writing here is pretty straightforward? I suppose the answer to this lies in a parable set forth by Robert Pirsig in Zen and the Art of Motorcycle Maintenance. In the parable, a student wishes to write about America, but can't quite think about what to write. Her teacher advisers her to narrow it down to just one town. She still can't think what to write. Then she's advised to narrow it to one street. Lacking any words to say she's told to narrow it down again to just a building. Finally she finds herself writing about the individual bricks in that building and, forced to think at the very basic level as if she's just seeing something for the first time, the words flow easily. A similar thing kind of happened here and it ran away from me quite readily. I saw an idea, and decided to write it as if I was looking at it for the first time, explaining it and writing things down. For the length, I am truly sorry. No one wants to read this sort of crap.*

*It's also difficult considering that this is continually expanding and developing. I've altered parts as new ideas came to mind or were read elsewhere, and I've added new sections as I came to realise that I needed*

*to clarify myself more. Perhaps it will turn out to be for nothing, and someone will turn around and say “actually, here's the book that explains it all much better”. I'd actually be happy with that, I know I've got something from this little exercise so far. Just now I feel it's at the point where it might be able to be read by someone else and they can point out what parts are bullshit or not. Otherwise, it risks running out of steam.*

## Part 1 – No psychic powers

*How we can prove that someone knows what the hell we're on about and how simply asking for a beer is far more complicated than you might think...*

How do we know what someone else is talking about? How do we understand *when* they know what we're talking about and when they don't. In one sense, we can't and it's impossible; we simply don't possess the psychic powers required to know this. So, like the scientific approach where evidence is gathered to support a proposition without needing to outright *prove* it (and we keep our minds open and are aware of what evidence *disproves* it) we need to amass a body of evidence rather than outright prove. The evidence must support our hypothesis – that “This person knows what I'm talking about” – and also must be consistent – that all indicators point to this person knowing what I'm talking about throughout the conversation. This seems intuitive enough: Ace asks for a bottle to be passed over to him, Gabe says a bottle is a bottle is a bottle and so can't get confused. They agree on what it is by definition so know what one wants to pass to the other. The end result is that an idea has been transferred from one mind to another, and then the success reflected in the fact that a matching action is performed that was expected by the first person. Ace expects Gabe to pass a bottle (though not the bottle opener, as he can use his teeth) and Gabe then passes a bottle over to him

to drink. The clink glasses knowing fine well that what amounts to a minor miracle has just occurred.

This leads us to a clear indication of how to tell if someone has understood; you can form effective two-way communication where responses fall within a range of expected replies. This is not to say that a person *knows the answer* if they are asking a question, but simply that they have in their head a set of ideas that they expect to hear in response. They have the form of an answer already in their head, as if they didn't they would have struggled to form the question in the first place. We form a reasonable expectation of what the result will be from what we say, and anything lying outside that expectation indicates that something has gone wrong somewhere in the great chain of communication. Less specifically, if it is not a question being asked, then we might be making a statement and expect it to elucidate some reaction within a certain range.

After several more beers go down the hatch, Gabe eventually refuses to pass over yet another bottle and tells Ace he's had enough (the freeloading swine!<sup>1</sup>). While that doesn't *obey* the request, it's certainly not outside the realm of expected responses to the original statement. In this case, it is an response coherent with and suited to a request, it just happens to be in the negative rather than affirmative. Ace is aware of a set of potential responses that match up with his statement; and should the given response match up with part of this set, then he *still* has positive evidence that communication has occurred and ideas have been transferred correctly. Should Gabe hear “dude, pass me a bottle, I'm thirsty” and instead interpret this as “please go do nasty things with my sister” (if it's possible to get any further from the actual meaning of the request) then clearly something has gone horribly, horribly wrong. The intended idea simply wasn't transferred, and when Ace later catches Gabe

---

1 Or perhaps he's simply sick of the sight of Ace using his teeth to open the bottle.

with his sister he will have a remarkably *salient* piece of evidence for that miscommunication. That action lies entirely, at least for a given value of “entirely”, outside the expected set of responses coherent with Ace's initial request/statement. It lies so far outside that realm of expected responses that the initial request may as well have been anything and Ace could have kept quiet instead. This is a more general form of evidence that communication has occurred correctly; someone has meaningfully influenced another person. If they could have said *anything at all*, then the influence wasn't meaningful. Good interpretation during communication is almost *defined* as meaningful influence.

In another way to ensure a consistent hypothesis (that the person really does, and always will, understand), simply asking a question in a slightly rephrased manner can indicate proper comprehension. If someone is trained only to parrot off one answer without understanding all the implications of it, asking a question in a modified way will not produce the same result. This is akin to using common sense; solving already solved problems quickly. The mental effort is low, but the applicability is limited and new situations risk causing confusion as we frantically search, in a machine like way, for a new answer. We will see later that the same idea can be represented by different words or different combinations of words, and so we can confirm the transfer of a correct thought, given certain caveats, this way. More generally, we aim to use different methods to probe the thought processes of another person by using words (generated by our own thought processes) and matching up responses with a set of *expected responses*. The more ways we have of doing this, the more evidence we can acquire for the hypothesis “this person understands me”.

Mutual understanding then becomes a testable hypothesis because we have expectation of sensory input; namely a coherent conversation, or

a suitable reply, or someone passing a beer bottle when asked. If we get an inconsistent sensory input, we know a miscommunication has occurred. But the big question is this; how far can we get before we *notice* this miscommunication? We know in science that one data point is never enough, and that experiments either need repeated or have their conditions changed to thoroughly test the variables and get a good feel for how reality operates. In a flippant application of logical positivism, we could say we can never actually “prove” that we're communicating properly, we're simply waiting for that one piece of evidence to disprove it. In simple model situations doing this is easy enough, and usually one data point is enough because the potential reality is artificially limited by the simplicity of the model. In these model systems, mere word games, miscommunication is very easy to spot and forms a useful exercise for thinking about the kind of mismatches and the kind of evidence we would look for in probing the success of real communication.

If someone constructs a simple expression composed of “pass” “to me” “Object X” and the receiver passes Object Y instead, clearly the problem lies in what they understand the object to be. If they were to start juggling Object X instead, then they're either showing off needlessly what they can do and wasting time, or because it's a model thought experiment and the possibilities are artificially limited, they've misunderstood “pass” and “to me”. The action, the sensory experience, lies outside the expected, and the question may as well not have been asked in the first place. It's important we know what responses match up with our requests and how, and what we can do to ensure these all match up in the future. Simply define Object X and Object Y as a brick and cement respectively, and introduce a few more objects into the model and you can start to build a house. Effective communication will produce the sensory experience of a house being built from that material – ineffective



communication leads to a pile of rubble. This holds true in this simple model system that just involves the rapid delivery of these objects, and where the rest of the terms in the sentence possess little to no ambiguity.

Of course, languages aren't really model systems like that! We have scope for ambiguity and complexity, and there are many, many points where a potential miscommunication can occur. We may not notice that the house is in fact a pile of rubble until it is nearly complete, or perhaps we might look at the pile of rubble and fail to notice that there is a house in there somewhere. In the real world, the nasty and distasteful thing that it is, we might not be able to easily spot where the miscommunication has happened, and tracking down the error becomes an arduous task. A conversation may progress quite happily (or not) between two people until they realise that they are, in fact, talking about two different things entirely.

A plethora of comedy sketches have been built on this failure to notice a miscommunication, not least the famous “Fork Handles“ sketch, based around homonyms of common everyday items: fork handles ' four candles; hoes / hose; and so on. Few, however, realise how many seriously heated debates actually occur based on the same failure to notice such simple miscommunications.

Given that the real world isn't full of simple word games where one-to-one, ambiguity free correspondence between words and actions is a real and clear-cut thing, there has to be some method to deal with the problems that *do* occur. Words have a particular salience to us and so need to be used properly to facilitate communication – to avoid unproductive arguments or circular discussions based on easily avoidable errors. And so the ultimate goal of rational discussion is reached; we cease repetitive argumentation and discuss things in a constructive manner to generate and share ideas. When recognising how words work

(and how they don't) and then finally putting measures in place to circumvent their inherent problems, statements that seem controversial can become *testable* – and at least so far that these can be practically tested, agreements should be able to be reached. By no means can colourful disagreements and conversation be avoided and replaced with machine like churning of data, but with a realisation of when someone actually isn't talking about the same thing *you* are talking about, debate turns into a more constructive discussion of issues, instead of people staying at loggerheads over silly things like definitions.

Is arguing over definitions silly? Simply contrast it with the “Fork handles” sketch; the sketch centres on two people effectively arguing over the definition of the three syllables, *for-kan-dals*, and the humour derives from this word play and the performers increasing agitation over their confusion. Do these three syllables mean “four candles” or “fork handles” – and could this not be solved by merely writing them down and checking the spelling? After all, words have definitions in dictionaries and so they can be easily cross-checked. But this would be to miss the point – indeed, the sketch ends with them doing just that and finding that the spelling/handwriting is equally ambiguous. Imagine a world consisting *only* of the verbal, say before we had written language. There would be no running to dictionaries for solutions there, they just wouldn't exist. Similarly, there is no need for us to expect that help is available even when we do write things down as that too is subject to the same possible ambiguities of the *for-kan-dals* set of syllables. That the sketch mostly takes place in the verbal, rather than written, world is very much irrelevant, the nature of the ambiguity is equally the same and relies just as much on a miscommunication; the failure to fully transfer an idea from one mind to another.

The outcome of good communication should be that when you use

words *properly*, with all the caveats about ambiguity and definitions in place, and then agree on empirical facts (and tests for empirical facts) there is no way that you can disagree on a definition of a word, and it's also unlikely that you would disagree on anything else.

Like most noble goals it sounds easy when put like this – and almost too obvious to need stated – but can be quite a challenge principally because people seem to treat words *backwards*. They might disagree on definitions, and think this *matters*, as if how we like to define arbitrary combinations of vowels and consonants dictate our thoughts, rather than the other way around (think again to whether defining *for-kan-dals* would define if someone wants “fork handles” or “four candles”). Years of indoctrination to appreciate the beauty and use of language not as a means to an end (effective communication) but an end in itself (poetry and prose), must have taken their toll on the world to produce such an illusion! To poets, authors, and literary students, a suggestion that language is nothing but a tool is vulgar and despicable – and so it seemingly goes onto a pedestal to be worshipped and never lowered to such a pathetic level as “tool”. Perhaps they would be right to think this in certain circumstances; but to *outright deny* that language needs to be manipulated properly and with care and precision, like any work tool, is often where this thought eventually concludes. Simply put (not to distract from the skill of cleverly arranged wordplay of poets, writers and comedians) we do *not* communicate rationally if we neglect proper manipulation of word-terms. A world where *for-kan-dals* can mean two things is highly damaging to communication of rational thought, and we can quickly gather evidence that we aren't transferring ideas effectively when we ask for fork handles and get four candles back in return. The wit, humour and supposed “beauty” of this word-play trick is of no *practical* use to us. We need to communicate rationally by

treating language as a tool as a means to an end, and so we must design it or work around its flaws to meet that end.

Take the concept of the “Grammar Nazi”, which most people will have encountered at least once in their life. Of course, it's Godwin's Law<sup>2</sup> firmly in action to compare someone who is picky with their language to Nazi Germany, but it's still a term with a wide recognition – the entire point of the comparison is to drag along all the negativity associated with Nazism into finickiness about grammar, even if such a comparison isn't warranted. Now certainly, not knowing the difference between “there”, “their” and “they're” can be an infuriating fault for a reader, not capitalising “i” is so frowned upon that most word processors instantly correct it and you need to take great pains to keep it in lower case, and double negatives are simply confusing. Ending sentences with prepositions, starting them with prepositions, verb agreement, and so on, fragmented or run on sentence structure, colon and semi-colon use – all these rules are constantly enforced in both spirit and also in literal letter in Grammar Nazi-land. Even mispronunciation is a symptom of this pre-occupation with the correct *form* that language takes.

These rules are, of course, there for a reason; a garbled sentence structure doesn't facilitate good communication, not agreeing your subjects and verbs leads to ambiguity with who you're referring to, similarly ambiguity of double-negatives<sup>3</sup> and at least knowing the difference between the homonyms of “there” is a good indicator that you're not a *complete fool*. At the same time, you must ask yourself if meaning is fundamentally altered by an incorrect use of most of these

---

2 Godwin's Law, named after Mike Godwin who coined it, states that “as an [internet discussion] grows longer, the probability of comparison to Nazis approaches unity”. It's often evoked as a sign that people have simply lost all perspective of the discussion. And *how*, this term itself is comparing someone who is picky with the structure of words to people who were picky about who should be allowed to live!

3 “Please do not fail to not uncheck this box if you do not wish to opt-out of our organ donor programme.”

rules? In 95% of cases (assuming 95% of statistics are made up on the spot) someone can easily correct you on your spelling, punctuation, pronunciation or grammar when you've made a mistake – although beware the ones that are a little too eager or quick to do this. It seems trivial at first, but consider for a moment; if someone can correct you, and quickly, then this must mean that the *intent* came across loud and clear! If someone corrects your pronunciation of a word, they must have understood the word you “meant” to use and how it “should” be pronounced. Intent made it across from mind to mind despite a problem within the formality of language itself; it's like magic. Despite the flaw in form, the correction actually is positive evidence in support of the hypothesis “this person knows what I am talking about”. If the meaning had been lost by an error in formal language, a correction would not be so easy or obvious. Where intent hasn't come across because of a garbled sentence structure or improper use of a colon a correction isn't readily available. Someone can point out the error, but they can't *solve* the ambiguity, at least not with certainty. In English it mostly happens when you try to compound extra detail into a sentence:

*“The woman hit the man with the walking stick.”*

*“Did you spot the girl with the telescope?”*

The ambiguity should be obvious in these two simple examples. Does the man have the walking stick or is he being hit by it? Does the girl own a telescope or was someone peeping through it to find her? On their own, these are pretty ambiguous sentences. If we had more information we could make a reasonable guess at solving the ambiguity. The sentence before might clarify who was in possession of the walking stick or the

telescope, for example, or the objects in question might not be so ambiguous in their use, as if you were to ask “*did you see the girl with the telephone*” it's slightly more obvious that you're not seeing through a device designed to transfer sound only.<sup>4</sup> On their own, however, we have multiple options and no available evidence to say either way. An example where evidence is available could be this:

*“She heard a knock at the door, came down in her night-dress  
and opened it for the visitor.”*

We can make a reasonable inference in meaning here. It's already been established that there's someone at the door, and you are aware from this that you can open a door and that visitors can be let in through it. But if it was a *Carry On* film we might expect somewhat differently – many apologies to those who don't quite get that one. While extra detail is needed to shore up a particular inference of what it is exactly that this woman is opening for the visitor, it is still biased by our existing preconceptions. Preconceptions such as our willingness to jump the conclusion that the visitor is going to experience an “oo err, missus” moment, or is just going to walk through a door into a fairly boring situation. Next, look at this example, spotted in a set of instructions for undergraduate chemistry students to follow:

*“You can now calculate B. You already know  $\Delta v_3 : \Delta v_4 = 2.27$   
 $\text{cm}^{-1}$  as calculated from an infra-red band at  $881 \text{ cm}^{-1}$ .”*

This is infra-red spectroscopy and the calculation of rotational and

---

<sup>4</sup> Even given modern technology where mobile phones have video calling available, the inference in this one can be considered *very* clear.

vibrational constants of molecules, but the actual science isn't important to know. Note the humble colon here, which flummoxed a number of students into thinking that  $\Delta\nu_3 : \Delta\nu_4 = 2.27 \text{ cm}^{-1}$  was a ratio of some kind to be calculated. Indeed, this is a reasonable inference when looking at mathematical symbols like that; the consistent string of non-alphabetical symbols seems to imply that the entire thing should be taken together as one unit. If your experience is seeing equations laid out like that, then you would probably jump to that conclusion based on your expectations. However, the reality was that it is a *grammatical* colon rather than a *mathematical* one. It is, in fact, simply stating that the undergraduates already knew  $\Delta\nu_3$  from earlier and the colon was there to indicate that this was entirely new, but related, information about another value,  $\Delta\nu_4$ . Such is the trouble when combining the hard world of formal mathematics and the soft and fuzzy world of language!

These are just simple examples of ambiguity, and for the most part the actual intent of the phrase *does* come across or can be inferred with reasonable accuracy from other contextual information. Yet despite this ultimate success in transferring meaning, the formality of language can and still does take precedent over the formality of thought during casual conversation. We still have dual-citizenship with Grammar Nazi-land, opening us up to corrections about simple spelling and grammar issues, mere formalisms, regardless of whether intent is transferred. It's an intriguing possibility that this attitude could be the root cause of communication problems as it causes people to suddenly think that the form of words have a *real effect* on thought. In fact, it's really the other way around. The form of words don't alter our thought patterns, our thought patterns are what influence the form and use of words. Mispronunciations, for instance, merely stem from what our prior experience of those words are – mispronouncing a word is really no

different to speaking it in an accent, it's just what our existing knowledge tells us about what sounds to make. Thinking about this in the wrong, backwards, way is known as “mistaking the map for the territory” in logical fallacy speak. The fallacy isn't always the most obvious one to see at first, but its effects are wide-reaching. It stems from the idea that a map might represent an area, but it is not *the* area itself. This confusion has lead to many literary satires on the concept, where 1:1 maps are created to ensure the most detail possible is included, and so instead of a map of a country someone uses the country itself. After all, it does work just as well, if not better. Maps like this, though, don't exist in reality. No one can build a 1:1 map at all, locating with precision every atom of the territory, hence why it's used as satire and fun in literature and used to represent a logical fallacy. In terms of converting thoughts into words this is the same as saying that we don't have psychic powers; we can't generate a 1:1 map because that would mean our map of a thought practically *is* a thought. So we rely on maps, not just in the obvious sense where they're stored in the back of a car, waiting for that one time when you actually travel somewhere you don't know,<sup>5</sup> but also in terms of communication.

No one would consider that a round black dot with a cross on it *was* a church with a spire, yet that is what one appears as on an Ordnance Survey map. Contour lines represent hills, sometimes to a great degree of accuracy, but they are *not* actual factual hills, with elevation in the third dimension and grass and rocky outcrops and the occasional goat. Yet despite the obviousness of this particular fallacy when described, people still insist on making it when it comes to words. All other things being equal and assuming no ambiguity is present, the spelling, punctuation and

---

5 GPS is slowly killing this exact analogy. But even so, the GPS device still contains a map which represents what you're driving through, and most work by trying to snap an *approximate* position given by the Global Positioning System onto the *exact* position that most makes sense on its internal map.



grammar of a sentence are only a *map* of what someone means; they *represent* what someone means, but they are not *what* someone means. Their choice of map or legend can be poor, hence why we do use spelling punctuation and grammar *correctly* to *avoid ambiguity*, but that doesn't necessarily invalidate their underlying thought processes – it simply means someone needs to clarify themselves properly and this will be a topic that comes up later.

That words only form a map and not a territory isn't the only problem. It's the simple fact that translating between the two isn't necessarily straightforward – indeed the proposal here is that this is the main route of the problem. Maps are fine if someone agrees on the legend, the scale and a few other conventions, but we may not have the luxury of that in language – or more specifically, we don't have the luxury of the psychic powers required to ensure that our legends are the same. We can't look at each others territories to check our respective maps against them. Communication is like handing someone a map that *you* know is correct for the landscape of hills and valleys that you want them to explore, but once that map has been handed to them you can't confirm that they're taking that map to the same, identical, set of hills and valleys. Could a dot and cross representing a church with a spire on your own map be replaced by something like a church with a tower instead? Or could the error be more blatant and in fact you set someone off walking down a valley with a completely wrong map that leads them in entirely in the wrong direction. You can't really confirm this until they return their own map to you and you compare. Like the symbols on a map that can have a few different meanings (church with a tower, church with a spire for example) words and phrases too are really laden with hidden inferences that you make inside your brain, and when you use these words the inferences are used just as much as the supposedly agreed upon

definition. These hidden attributes may be part of your own map and legend, but not as part as someone else's – and of course, telling someone to go to a particular grid reference on a map and find the object represented by the black circle and cross is far easier if you're both using the same map and the same conventions to lead them to the church with a spire, not a church with a tower, or something else entirely. In language this problem is (almost, at least) independent of perfect spelling, punctuation and grammar – as people who like to use proper terms for language will know, semantics (meaning) and syntax (structure) are two different things. Good syntax can be a tool to help clarify meaning, but it's not so essential that meaning can't be read at all without it. This is why different languages can have different arrangements in their sentence structure and why we easily parse something like “pass me now the beer” or “pass now beer the me” as meaning “pass me the beer now”, despite some of those sentences not having good structure from the point of view of the English language and Grammar Nazi-land. Good semantics can be considered completely different from good syntax, and even with perfect syntax your meaning can be completely garbled precisely because your map doesn't work. If your view of language is rooted firmly in *only* adjusting syntax (as most people tend to be taught) then you will have more trouble when miscommunication arises. You will be looking for errors in form when the error can be far deeper and embedded in how you've drawn up your own map itself. An in-depth look at this problem and how to solve it is the focus of this extended essay.

## Part 2 – Maps and territory

*Establishing the cognitive-linguistic map, and most importantly, making sure it's plugged in the right way round...*

In the inter-war years of the 1930s, the race was on to decipher Enigma, the German military code that encrypted their orders and communications. Even though war hadn't broken out and that, at the time and without hindsight bias, looked very improbable, diplomatic tensions were still high enough to warrant constant surveillance of the military machine being built in Germany. The Enigma code was a complex and involved scrambling system, where electrical signals corresponding to a letter were passed through a number of wheels and back again to convert them to a different letter before being transmitted. Such security was essential, as you can't be sure who would ever receive your message when it was beamed through radio waves in every direction. Enigma was more than a simple Caesar cipher, where letters are just swapped around in a fairly arbitrary yet straightforward fashion. Such a thing is very easy to crack – for example, the most common letter in the English language is e, and so whatever the cipher routine, you could presume the most common letter would still correspond to e. In these simple ciphers, every instance would also correspond to that letter, and you can deduce the code simply by looking at its encoded form very easily. Sherlock Holmes

deduced *The Adventure of the Dancing Men* in much the same way, spotting the patterns in different stick figure drawings to crack the code. In Enigma, however, the three or four wheels that were used to scramble the individual letters would rotate after each entry, altering the sequence that the letter would be scrambled to. You might type A and get G, but the next time round you could type A and get J. Eventually the wheels might come around to the same configuration again, but that amounted to thousands of letters – for all practical purposes this would never be seen, removing one possible tool for decryption. The amount of entropy introduced into the code because of the number of wheels and their constant realignment was enormous, and without either great mathematical computing power or exact knowledge of the construction of the wheels it would be difficult to crack. In addition to this, the wheels would initially be set at a different point for each message, and that too encoded by Enigma itself, in order to make the exact cipher different each day. But that wasn't all Enigma had going for it, while the wheels did most of the work, a “plug board” could randomly swap pairs of letters, adding another layer to the code, and these could be changed at will leading to even more possible combinations of encryption. Any command or piece of information sent by Enigma would then have to be translated with these exact settings of wheels and plugs, or the gibberish produced by Enigma would come out equally as gibberish. Every time this was reset, the code would have to be broken again from scratch.

Enigma was not without some cryptographic flaws and is probably considered pretty weak by today's standards and was weak compared to Lorentz, the high-command code for which the first truly programmable computer was built to decode – a full discussion of these things is unfortunately beyond the scope of this little side story. Aided by some procedural slip ups that allowed access to the tables of wheel arrangement

and plug board settings, it was properly “cracked” long before the war broke out, by a small Polish team in 1932 to be precise, and Allied cryptographers were deciphering the code quite readily throughout the war.

However, all this talk of the complexity of Enigma and code-breaking is, in fact, a red herring of sorts. Figuring out how the wheels scrambled the information and gaining information as to the correct settings for all the variables would be for absolutely nothing if it wasn't for one, remarkably simple, problem: how did the keyboard actually translate to the first wheel?

This seems so simple, and you would be forgiven for nothing thinking it to be even relevant, yet was a key problem. The entire project of deciphering Enigma rested on this. Perhaps this is a slight exaggeration, but few would doubt the importance of this first step. Without knowing this first, highly crucial, link up within the encoding machine there could be no decoding. The wheels could be known perfectly and the daily settings translated in the clear, but if the translation to the keyboard didn't match up there was no way any amount of cryptographic analysis could break the code in reasonable time. Could the letter A on the keyboard translate to the position J on the first wheel? What if it mapped to L? Was it random or was there a sequence? Were the positions kept the same as the first commercial version of the scrambler where they hooked up in the order found on the keyboard, Q, W, E, R... and so on?

This was a problem faced by Marian Rejewski in the pre-war years as he worked on deducing the operations of the wheels himself from mathematically analysing the code. It wasn't a security problem, as such, as this mapping was (from the point of view of cryptanalysis) a simple Caesar cipher as described earlier. But it was a significant obstacle in

exploiting some of the features of Enigma itself to decipher it, such how certain sequences repeated among other fascinating techniques. Without this one simple starting block, the message would stay garbled and the techniques required to decrypt Engima's wheels wouldn't work – unless you simultaneously guessed and compared your results with every permutation of keys to spot the patterns that elucidated the wheel combinations.

In the end it was simple alphabetical order. The A key mapped to the first position, B to the second, C to the third and so on. But this says nothing in particular, it was no barrier to security if cracked but it would be a barrier to the methodology of cracking the code in the first place. The key point is that this sequence didn't *have* to be like that, it could have been any arbitrary combination. It was the same in each machine and once cracked that would be enough – but that first point of translating the signal from the letters the operator pressed into the machinery of Engima was crucial to understanding the entire code and breaking it from just its output. The Enigma code could spit out an E, but if the operator had wired that up to the light bulb under the letter O, things would be problematic and there would be no way to guess that you had the code correct. In fact it's sometimes considered a great and inspired guess that Rejewski figured out that the keys mapped in such a straightforward alphabetical manner like that.

This is where the analogy to communication between people comes in. Engima's scrambling was complex, but ultimately built on rules, much like language, which is *very* complex but still based on a few rules. Enigma could be deciphered by exploiting its rules, again much like language. And like language you could look at it to deduce some of these rules. For instance, you don't have to understand the language to spot

certain features like punctuation, or that certain words get used a lot like “a” and “the” that seem to signify something, or that they only appear before some words and not others. You could have a good go at deducing some features of a language having never seen it before or being taught it. The Enigma code, and most ciphers, are like that, although such features are often mathematical and not linguistic in nature. But figuring this out is useless without figuring out how those rules mapped to the actual input of the encoding machine! Otherwise you're stuck trying to deduce an extra layer of complexity that acts as a significant barrier to you cracking the *real* complexity of the code itself. And so, understanding the syntax of language is useless to our ability to create good communication unless we also know how the input into the “machine” of language matches in the first place – the semantics, to use the fancy word. So where Rejewski had problems cracking Enigma until he correctly guessed that first initial wiring state, we will have problems in understanding and decoding language unless we know that very first state that transfers thought into words. Without this, we can struggle crack the code of language and communicate well, but while we might be able to correctly guess some of the form of a language the true meaning will still elude us. The previous section established this; in order to connect our words to an expectation of sensory response (to confirm the hypothesis that someone understand us) we need to know that semantic bridge between thought and words.

Ostensibly, one could say that running to a dictionary provides a solution, but in fact it doesn't. Words are not there to represent their meaning (something that you *can* get from any dictionary); their meaning is there to represent a thought, an idea, a raw feeling within our own minds. This is something you certainly *can't* get from a dictionary, but only from another person by comparing their response to a particular symbol and matching it with your own expectation. Words and language

form nothing more than a convenient, if sometimes convoluted, way of transferring these thoughts and feelings to others.

First person's thought → Words → Second person's thought

It seems obvious when phrased like this, but has many implications. The above scheme is how we'd like things to work if we were to design the system from scratch. In many ways, computer communication (something we *have* designed from scratch) works like this because a computer “thinks” in exactly the same code that it uses to communicate. You can consider two computers hooked up together to be effectively the same computer – with a few caveats to account for competing protocols, operating systems, file hierarchies and so on. Even so, these are still artificial systems and can be coded by us, due to knowing their specifications and design, to interact on a 1:1 basis. A computer doesn't need to actually provide an interpretation of what it merely *assumes* another computer is saying, they match up perfectly by design. Humans, on the other hand, don't have the benefit of this; we don't have psychic powers and cannot “mind meld” as a computer can. We first have to encode our thoughts, feelings and ideas into words and sentences before firing them across the air, or in squiggles on paper. This is the single step that determines the success or failure of communication, by which it's meant that the thought conjured in the first person's head closely matches the thought conjured in the second person's head.

The analogy to Enigma isn't that this process consists of convoluted cryptography. In fact you could say the analogy is there to present the convoluted encryption process deliberately as a red herring, a complex distraction to draw attention from the far, *far* simpler mapping problem. Enigma's wheels ran on rules like language (they are the



syntax), but nothing within the cipher itself stated which way the keyboard would match up to the first wheel (the semantics)! You can make some inferences about semantic, just as in the analogy you could decrypt the simple Casear cipher that results from wiring the keyboard up. However, in the interplay of actual language and thought the situation is a lot more complicated. The inferences that you make about what words mean aren't as simple as the ones Sherlock Holmes makes to solve the dancing men.

We can attempt another analogy to this. Most people today use keyboards in the Qwerty configuration, that is, the top row of keys reads Q, W, E, R, T, and Y. Unless you're already a Certain Type of Person, this will be the default you use and your keyboard will look something like this:

Q W E R T Y U I O P  
A S D F G H J K L  
Z X C V B N M

Not many people are aware that alternatives do exist. Such as the Dvorak Simplified Keyboard, which arranges the keys in a different way. Now, if you're a Certain Type of Person, you might actually be using a Dvorak keyboard layout right now, and might be looking forward to a good rant about how much faster and efficient it is. Well, unfortunately this is not one of those. This is just going to be a simple illustration of the concept of mapping and the barriers it can form. If you know how, and if you don't the internet can surely tell you, you can change the keys on your computer to the Dvorak layout. In this layout, the letters are arranged like this:

P Y F G C R L  
A O E U I D H T N S  
Q J K X B M W V Z

Which is considerably different if you're used to the usual Qwerty layout. But you don't need to pick up the letters off your keyboard and pop them back in a different combination for this to work, the computer can do it for you. You could then type a message in your Qwerty keyboard and have the computer map it to the Dvorak layout for you. If you were to type a message, it would look like this:

D>NNR MF BAM> CO OAM

It looks scrambled. Simply because you were typing in Qwerty but the computer interpreted it as Dvorak, the message is gone. Though you can get it back if you know the layout of the two keyboards and can translate between them or fancy performing a frequency analysis of the Sherlock Holmes variety. The point is that you could take the message above and pump it through an Enigma machine, or any code from a simple cipher to a complex 128 bit modern encryption process. When you put it through the rules-based encryption process and back again, you will always get this nonsense message “D>NNR MF BAM> CO OAM” back out again. Knowing the complicated encryption and decryption algorithms stands for little if you don't know the simple act of how the letters match up with the input on the encoding machine, in this case simply whether the keyboard layout is Qwerty or Dvorak. This is the exact same problem that faced Marian Rejewski when trying to decipher Enigma, and it is the exact same problems humans have when communicating. The only difference is that Rejewski was faced with far

more than two layouts, indeed it was potentially any permutation of 26 letter keys, and human communication is faced with even more layouts still, that are fuzzy, and change, and aren't available for easy checking.

And so, the map we all possess between a raw thought and a word or phrase to represent can be similarly arbitrary. It needs “cracked” on its own and no amount of study of language in complete isolation will convey how this map works; understanding all rests on assumptions about the map, and a lonely prayer that the assumption is right. This is why book translators simply can't rely on a dictionary to make their translation and require extensive notes from an author to convey their meaning properly. Similarly, translators of ancient texts need to know about what an author meant and not just what their simple dictionary says a word means, often involving a lot of historical analysis to get it right and a lot of footnoting and annotation to the final product to note what alternative meanings could actually exist given this context. There are many problems that have arisen from people reading translations of ancient texts and failing to take account of this fact. We accept this sort of thing when translating into another language, but we rarely accept this for reading in our own even though the principle is the same. A writer writing in English is writing in *her* interpretation of English, a reader is reading in *his* interpretation of that English. The fact that these are both “English” is only a matter of cultural convenience and a recognition that, for the most part, the overlap in syntax semantics matches up pretty well. One can claim, obviously that they're the “same language” and so are built on the same rules “by definition” but again, without access to psychic mild-melding powers we are *not* capable of confirming this. The more subtle implications, inferences, connotations and feelings conjured by words and collections of words go far beyond the simple dictionary definitions that are traditionally used to say that “this language is the same as another”. It

is this mistake, the illusion that “*my English is identical to your English*” that informs most of our miscommunication, which lies in the world of the broad connotation, sweeping generalisation and the lazy categorisation that words end up representing.

The mapping process connects the realms of cognitive thoughts, and the linguistic words. Though it would be a mistake to platonify this too much and take it literally; this linguistic domain can actually be any signifier, a word, a sign, a squiggle, an gesture and so on. They are things in the real world that signify something in our mind. Again like most concepts, the study of this has a posh word; “semiotics”. These shouldn't be taken in complete isolation whether they're words, or generic signs, but should be thought of in the terms of how they map onto actual thought patterns inside someone's mind. One particularly interesting extension would be to ask if this linguistic domain extends to the world of “art”. Consider that the origins of writing lie with ancient glyphs, derived from pictures of real objects. Later in the section regarding Real World categorisation we will revisit this in a little more detail, but for now consider how a picture represents an object, a pictograph is a simplification of this, and then eventually by several evolutionary steps we form a letter, and then words and sentences. The difference is fairly arbitrary so we could apply the concept of mapping a thought pattern onto a symbol in as generic a way as we like.

A simple and idealised map would look like this:

$$Q \Leftrightarrow q$$

Where  $Q$  represents the thought and  $q$  represents the word.  $\Leftrightarrow$  is just another symbol, stolen from symbolic logic, used to signify how they map onto each other in reasonable correspondence. In this scheme,

thoughts are all represented by upper case letters, and their respective words, or more general representations such as a sign or symbol or picture, shown by the lower case. They don't need to be written, but they do have to be somewhat physical; an entity that can be used to transfer, *sans psychic powers*, a thought to someone else. Audible compression waves in the air or squiggles on a page are treated equally here, both are manifestations of  $q$ -space. These aren't just the smallest possible units;  $q$  could mean a single word, or it can mean an entire book of words.  $Q$  could be the feeling associated with one particular word, like “hot” or “cold” (and some other information such as *how* cold or *where exactly it is* cold.), or it could be the experience of that entire book of words. As we will see later, this realm can be quite vast and produce unintended side effects when we don't know, with suitable precision, what the answer to these additional questions are between person to person.

The lower case  $q$  can even represent a *definition* of a word; because  $Q$  itself is not a definition, it is a raw thought – it has no words except those that the map associates with it. It's convenient to use “ $Q$ ” as the representation of the most general case because the Latin word for a raw thought or feeling associated with an event is “qualia”. A little used word in many cases, but a useful one for this purpose. Our qualia map on to some way of representing them, hopefully with the eventual aim of causing the same, or similar enough for the purpose, qualia to be generated in another person.

In other words,  $Q$  is the realm of the cognitive and  $q$  is the realm of the linguistic, or at least semiotic if you want to step outside language and generalise it to *anything* that can represent a thought.  $\Leftrightarrow$  then represents the cogitative-linguistic map between the two domains, acting as a legend to tell you what part of the map corresponds to what part of the territory. There is a problem though: this model isn't entirely complete as it stands.

The symbol  $\Leftrightarrow$  implies some degree of symmetry between the two, as the arrows move in both directions and the symbol itself is perfectly symmetrical. We make an inference from this sign that it means “goes both ways”, but that's not necessarily true of the cognitive-linguistic map.

In the realm of  $Q$ , we have a system of thought processes, cognitive ideas, implications, connotations and preconceptions. In the realm of  $q$  we have words, which are symbols and physical things in the real world. It should be immediately obvious that altering a word, or altering the definition of the word, shouldn't change anyone's mind. Defining *for-kan-dals* as “four candles” won't stop the customer wanting “fork handles”. Surprisingly, this is something people seem to forget when attempting to argue from definitions. What if you want to argue the definition of “religion” to make an atheist religious? Anyone debating religion on the internet will have come across this one before, it's quite a common, but amateurish, accusation. Would changing this make the atheist religious? Even if yes, and by definition the atheist becomes religious, the follow up question is simply: so what? We would have to only make a judgement based on what that definition of “religious” has become. Making this change in definition hasn't fundamentally altered what the atheist believes or their attitude (in fact, this actually creates greater problems for any religious believer putting it forward as a serious argument, as it dilutes the meaning of “religious” to the point where it ceases to have any *useful* meaning). This no more holds for more complicated terms and subtle redefinitions than it does for simple agreed upon terms and major redefinitions of them. We can take *every* definition in *every* dictionary, and *every* exemplification in *every* encyclopaedia of the word “book” and change it to “small slimy creature, known to inhabit the rainforest in Brazil, with an unusual mating cycle that is triggered by the presence of chocolate” and you would certainly *not* see every book in

your personal library (even this one if you choose to “define” it as a book) turn into a small slimy creature, known to inhabit the rainforest in Brazil, with an unusual mating cycle that is triggered by the presence of chocolate. The obviousness of the fallacy in this sort of process should be apparent with or without the intentional absurdity of the example. Even if we reached inside of language itself and made it so that “book” actually represented that said creature, rather than appearing as an odd printing error in every dictionary, we still haven't turned a “bound set of pages with writing on” into a “small slimy creature, known to inhabit the rainforest in Brazil, with an unusual mating cycle that is triggered by the presence of chocolate”. It almost resembles the omnipotence paradox, where we ask God to make a triangle with four sides, and He responds by altering the definition of “triangle” to “object with four sides”. We haven't altered the physical realm by changing words, so we cannot really hope to alter the cognitive realm (which contains ideas based upon the physical realm) by changing words.

Of course, like all things when phrased in an extreme way the fallacy is obvious. But how is this absurd example qualitatively different from any of the more subtle examples that might be seen? Evolution is a religion, atheism is a religion; this is morally right, this is morally wrong; this is good music, this is bad music; and so on. Does this matter at all? Consider a situation where one were to define “art” as they liked. That “art” is only a visual medium, or that “art” is *any* creative medium, or that “art” must have some pretentious purpose and is independent of anything that is simply aesthetically pleasing. This wouldn't change the nature of paintings, or sculptures, or music, or performances, or prose in any way. Nor would it change the nature of another individual's appreciation of them. It would merely alter what someone, and at that just one person, considered as “art”. One person dismissing a landscape painting of a

bridge as “is *not* art” is not going to change the opinion of someone else who might consider it “is art”. Here it looks like there is a disagreement of what art is, or that there is some “correct” definition and one of these people is plainly wrong. Is the landscape painting art or not? We can continue the thought experiment a little more and say that the second person eventually agrees with the definition of art provided by the first, and so must then switch to thinking that the landscape of the bridge “is *not* art”. But their *actual opinion* will not have changed at all – they have simply shifted their linguistic categories around to fall in line with someone else. One could easily rejoinder to this with the phrase “*It's okay, have the word “art”, I'll pick my own arbitrary combination of vowels and consonants and use that to describe what I'm talking about.*” If we can use that phrase to solve an argument, there really isn't an argument at all, only the illusion of one caused by mistaking the map for the territory.

Clearly the relationship between concepts and that which represents them is only a one-way street in reality. Words, therefore, are given meaning *based on* our thoughts. Not the other way around. Our linguistic categories may be manipulated by arguing from definitions, but the cognitive categories themselves are not. A clearer way of representing the cognitive-linguistic map, therefore, would be as so:

$$Q : \Leftrightarrow q$$

Here we've introduced an symmetry to the symbol mapping  $q$  to  $Q$ . We still have the arrow indicating a back-and-forth relationship, but the *defining characteristics* only move in one way. That is, words and symbols,  $q$ , are defined in terms out our thoughts and concepts,  $Q$ , that we wish to represent by them. The converse is not true and so we can also



say the following in faux-symbolism speak:

$$q \neg \Leftrightarrow Q$$

That is, it only works one way around. We define our words based on the experiences we have of them. We do not define our experiences based on words – or their definitions. Again, this should be obvious. We know the sky is *blue* because we've seen (if we have functioning sight, at least) the sky and know what *blue* is because of it. We do not know what *blue is* based on just being *told* that the sky is blue and what we think the sky looks like is isn't *changed* by being told it's of a different colour.

What of words influencing other people? Consider this outline of a communication scheme as we want information to travel:

$$Q_a \Rightarrow q \Rightarrow Q_b$$

We clearly want the process to evolve forward from Person A to Person B. Person A thinks thought  $Q_a$  and wants Person B to think thought  $Q_b$ . And, despite the labelling to keep who's thought is who's clear and separate, we want  $Q_a$  to precisely match up to  $Q_b$ . This means encoding thought  $Q$  into a linguistic construct  $q$  and then, very critically, for the second person to interpret it the exact same way. Except we can't have how words and thoughts define each other work in reverse. The communication scheme is actually better represented like so:

$$Q_a : \Leftrightarrow q \Leftrightarrow : Q_b$$

Now the asymmetry has been re-introduced we can see that both people must agree on their individual cognitive-linguistic map to ensure

good communication. Whether  $Q_a$  and  $Q_b$  match is dependent entirely on the set up of the map for that one individual. If they differ, we get a miscommunication. So if we assume two different thoughts,  $Q$  and  $R$ , that map onto the same word depending on the person we get a map that looks like this:

$$Q_a : \Leftrightarrow q$$

$$R_b : \Leftrightarrow q$$

Which produces the communication scheme:

$$Q_a : \Leftrightarrow q \Leftrightarrow R_b$$

And so Person B maps the wrong thought onto the intended word. But there is one problem with that last sentence there: who is *actually* wrong? Person A, or Person B? How can we tell who is wrong, or are both in fact correct? In this model of the cognitive-linguistic map the fact that  $Q$  and  $q$  are the same letters is only for convenience in the general case, to keep track of thoughts that are intended to map onto what word. We could phrase it with different letters for both the thoughts of Person A and Person B and a different letter again for the word or phrase in use – you can't tell from this formulation who is wrong:

$$X_a : \Leftrightarrow z \Leftrightarrow Y_b$$

This is where people will go running to dictionaries to save them, they hope that by defining a term they can get their thoughts to match. Except there is a problem with this:  $q$  doesn't represent just a word. The linguistic space here can represent anything, from a single word to an

entire novel, and the cognitive space represents everything associated with it. Reaching for the dictionary only defines what  $q$  is more precisely in terms of the linguistic realm; rarely do they help solve the miscommunication precisely because:

$$q \neg \Leftrightarrow Q$$

The thought is not influenced by the word itself. Redefine atheism, theism, liberalism, conservatism, libertarianism, Marxism however you like; the thought that someone is thinking will not change. Perhaps they have got the word mixed up, and an agreed upon term of art is being misused, but that still doesn't change what they're thinking. The next section will deal with how to solve this miscommunication.

## Part 3 – Screw your words

*Definitions make answers arbitrary, so simply make sure you say what you mean instead and avoid the problem entirely...*

Definitions can make the answers to a question fairly arbitrary. A closed, yes/no question like “is this a book?” can be either yes or no depending on your definition of “book”, which is merely arbitrary combination of letters that cognitively maps to an idea in your own head (representing an object in the real world, which will be put into the mix later). Trying to strictly, coherently, and properly generate correct definitions and, more importantly, sticking to them, has been the first line of defence in philosophically and politically troublesome debate for years. How can we have a discussion on morality unless all people understand what “morality” means? More pressingly, what is “good” and what is “bad”? What properties do they have? Is one the lack of the other, and if so which way around and *why*? All of these issues aren't alone; there are countless discussions that can be derailed simply because definitions aren't agreed upon. Simply look back at the cognitive linguistic map to see this:

$$Q_a : \Leftrightarrow q$$

$$Q_b : \Leftrightarrow q$$

Each individual has their own map to exemplify the meaning of  $q$  in their own head. If both parties fail to realise this, then they can be derailed when suddenly their cognitive systems stop agreeing with each other and lead to something else. Consider the following, more representative map that interplays a set of thoughts,  $R$ , and two similar but significantly different thoughts  $Q$  and  $S$ , as understood by two people, A and B.

Person A:

$$Q : \Leftrightarrow q$$

$$R : \Leftrightarrow r$$

Person B:

$$S : \Leftrightarrow q$$

$$R : \Leftrightarrow r$$

So imagine a system where the line of reasoning represented by  $Q$  leads to a conclusion represented by  $R$ , but the reasoning labelled  $S$  leads to “not”  $R$ , or  $\neg R$ ,  $/R$ ,  $!R$ , or however you wish to define it quickly and symbolically. Because the map of cognitive processes  $Q$  and  $R$  lead to the same linguistic process  $q$ , confusion arises. The statement that  $q \Rightarrow r$  and  $q \Rightarrow \neg r$  applies equally validly for both parties in the conversation, Person A and Person B respectively. The two will now spend considerable time arguing their position because they fail to realise that what they're talking about isn't the same thing – in fact they may well agree, and their cognitive processes obey the scheme  $Q \Rightarrow R$ ,  $S \Rightarrow \neg R$ . The point is that should you agree on your definitions, standards of evidence and basic procedures to come to rational conclusion, your cognitive processes should, *in principle*, be the same. This sounds esoteric and convoluted,

but it can be applied to real world examples easily.

Often, the argument arises regarding things with strong emotional connotations to words, of which more will be mentioned about later. Imagine a debate between two modern academics of the day discussing the world of feminism and what it means in modern society. One argues that it is a force for good and progress in the world, the other argues that it is simply a redundant entity, ultimately unnecessary and the cause of more sexism than it solves. Surely if two viewpoints were so diametrically opposed, and based on reality as claimed, they could be settled by test. One claims that feminism creates equality and the other says it removes it and these both constitute reality; as they're expectations of sensory input of some kind. We expect to see a correlation between sexist attitudes and the local prevalence of feminism, and the direction of this correlation will settle the argument once and for all.

Yet debate and argument persist regardless. Indeed, any suggestion that it could be settled by a test might lead to greater problems. When you request to settle such an argument you imply one side is wrong; and to accept you might be wrong is not a pleasant position for those attracted by the prospect of “debate”.

To see where the problem arises, think about how you would go about testing this hypothesis: “feminism is a cause of more sexism”. You very quickly realise, almost instinctively when you consider it as a question of testing an hypothesis, that the very hypothesis rests on definitions. This is just like how the closed question of “is this a book?” rests on the definition of “book” to make the answer non-arbitrary. A test for a hypothesis, real-world fuzziness aside, is almost invariably phrased as a closed yes/no question. It is, in principle, a binary of a “real” or “not real” effect, or at least in more reality-orientated terms it is either “consistent with evidence” or “inconsistent with evidence”. To

understand this and answer the question we need to know specifically what to look for in terms of evidence and simply can't go into answering such a riddle blind, charging around demanding evidence for terms we haven't fully understood ourselves. The trouble is, people actually will do this because they think the terms *are* settled and understood; they just don't realise that they're only settled in terms of their own cognitive-linguistic map. The answer to the hypothesis “feminism is a cause of sexism” relies entirely on how you define each term in the hypothesis statement, making it somewhat arbitrary unless everyone is agreed on what it *actually* means. Defining sexism is one thing, defining feminism is another, and this makes things even more complicated and you mix-and-match these individual definitions to produce a statement that could mean an increasing number of different things. So the answer is not just dependent on how you define one term, but on both terms and how they interact with each other to form a wider system of actual meaning.

So let's get back to the hypothetical debate by our Ageless-Faceless-Gender-Neutral-Culturally-Ambiguous-Debating-Persons on feminism causing or not causing sexism. It's likely that you've probably thought of the one arguing that feminism is good is a woman, relatively young, liberal leaning and politely spoken while the one arguing that it's bad is actually a grey haired and wizened academic with a horrid demeanour and no stake in reality. This is the sort of connotation that comes across when using a term with a strong emotion response, regardless of what definition we try to give something, even hypothetically for the purpose of conversation, or thought experiment. We could define “feminism” as something closer to Nazism (or whatever) and it would probably still conjure similar images we already accept as belonging to “feminism”, whether it be protesting suffragettes, bra burning (yes, it's a myth, but highly pervasive, which is the point) or sex-

positive columns in progressive women's magazines. These connotations and images we associate with a term persist even when arguing *by definition*. The problem with arguing *by definition* is that we place ourselves under the illusion that these connotations and implications aren't there any more – but they are. We could, hypothetically, say that for our hypothetical discussion with our Ageless-Faceless-Gender-Neutral-Culturally-Ambiguous-Debating-Persons is taking place in some alternative universe where “feminism” is practically synonymous with something else suitably opposed like Nazism, or perhaps the philosophy of noted women's rights campaigner Phyllis Schlafly. Yet we'd still find it difficult to swallow this “definition” in our own world, such is the strong emotional attachment that we have to words and phrases, despite their contextual definitions we're given to work with. We could argue, *by definition*, that feminism is simply synonymous with “gender equality” and nothing else with no strings attached and yet despite this those strings would still be attached, based on our prior experiences and pre-conceptions. Saying that something is this way *by definition* doesn't detach the strings, it simply denies that they're there.

This issue with definitions going awry between different people goes into deeper and more subtle territory than, for example, Ali G subtly mocking a politician by using the term feminism in a way synonymous with lesbianism – “do you agree that all girls should try feminism at least once? Yes? Really, man? Wicked!” – but the effect is, in fact, the same. Two people talking about something and not quite getting that they're not referring to the *same* thing despite care to use the same language and make a paltry attempt at “defining it for the purpose of our conversation”.

Our own historical context is constantly defining and redefining our terms for us. We have a history of using words to represent concepts and we're arguing continually about them, exposing ourselves to what



they mean. The two concepts of “Nazism” and “feminism” are two terms that provoke so much emotional response that invariably complaints will be made that they were even put in the same sentence, or implied to the same thing, or compared at all. The choice of using them together just above was very deliberate for this reason, it would illustrate that emotional attachment almost immediately without having to even point it out. Mentioning them is likely to cause masses of offence – even if pains are taken to ensure that you're not *really* conflating them, but putting forward a hypothetical exercise in the nature of definitions. This is a very strong attachment not just to a word, but to the cognitive-linguistic map *itself* –  $Q : \Leftrightarrow q$ . It's not just the fact that we love a word, we are in love with what it represents, albeit only in our own heads, and how that word succinctly symbolises it for us. An attack on the word is therefore an attack on the thought we choose to have represented by the word – even if the attacker doesn't map the same thought onto it.

This is no different to how a flag can represent patriotic feeling, and seeing a flag burn is an attack on that feeling. Patriots, angry at the misuse of a national symbol, aren't just angry at the misuse of the symbol, they're literally seeing it as an attack on the associations they have with that symbol – usually the good aspects of the country and nothing else. The flag burners, however, see the flag as a symbol of their enemy and their bad aspects (and little, if nothing else). Bad aspects, that is, from *their* perspective. It's mistaking the map for the territory, sure, but it's a pretty salient map that looks, to the user, an *awful lot* like the territory itself. It's a powerful draw, and built into us because we need such quick and intuitive maps in order to communicate with speed and efficiency. There's no good in communication if we *can't* do this as otherwise we would take a long time to say even the simplest things. Do we really need to clarify that “going out to eat” means going to a restaurant, rather than

merely wandering outside to eat? Well, if a miscommunication occurs then yes, otherwise not at all. We all tend to make similar inferences and so things proceed quite smoothly. A definition that is agreed on by all, and so causes no miscommunication issues, is an already-solved problem and so we will leap back to it again and again. Eventually this one particular use becomes so ingrained by this process of use and reaffirmation that to think the word could have another would be, almost literally, unthinkable. Now, this in itself isn't a problem. Reusing the same answer for the same question again and again is a fine enough tactic. It's all about efficiency and the look-up table of answers in our heads is far more efficient than doing complicated thoughts and calculations from scratch each time. The problem occurs in a failure to recognise when a question is different and the same answer given yet again doesn't actually help. This causes us to stumble, and because the use and reuse of the answer is so deeply embedded, we would rather question the validity of the question than of our response.

We could try a thought experiment of intentional remapping with other terms; imagine defining “Good” as consistent with murder, spreading disease or stealing, and then imagine defining “Bad” in more positive terms, helping, giving to charity and so on. It would be a difficult thing to communicate to someone. It would take a lot of cognitive effort to keep mentally switching your language around for every instance. Even if we were to take the courtesy of defining our terms in advance, we'd still have this niggling doubt that they were the right terms to use. Emotional attachment and prior experience drives this.

This is all due to the connotations attached to words that give them extra meaning that the dictionary can't convey. A dictionary can tell you what “marriage” means, but it can't tell you what it's like to experience one. The definition of “death” would have a different level of salience to

you if you had recently lost a loved-one and the dictionary won't convey that to you. These are the extra thoughts that are difficult to shake off even if you wish to lay out your definition in advance. Our experiences and interactions with terms feed back into the definitions in our own heads. We might develop our idea of  $Q$  to form something slightly different,  $Q'$ , but this will happen so slowly that we fail to realise that the change in our cognitive idea hasn't been matched by a change in our linguistic representation. Thus at one point in our life we might map  $Q : \Leftrightarrow q$ , and at a later point in our life our experience will have morphed this to  $Q' : \Leftrightarrow q$ . And moreover we fail to realise this subtle change, and presume everyone still agrees with us when they too use  $q$ .

For many people in their 20s or older, the phrase “I believe in God” will hold a much different meaning now as it did when they were 6 years old, whether they were brought up religious or not. Assume, like many people who *are* brought up religious that, aged 6, they simply believed by default. Their idea of God would not have been fully formed, and it might not be theologically “correct”, simply because they had not yet been taught everything in a more adult sense – aged 6 they were likely to have been patronised by the awe and wonder of a loving God, and may have been spared the more complicated absurdities or theological implications or other more involved areas of belief. Their 6-year-old idea of how the Bible sticks together and its historical background would be limited. Ultimately, their own personal use of the phrase “I believe in God” will be fundamentally different compared to an adult because of the context in which they understand the terms “believe” and “God”. In fact, consider what it might actually mean to a 6 year old; perhaps a simple parroting of what they've been taught, simple acceptance that it's not just *true as a belief* but just how it *always was and is and that there are no other belief system*, and add to this no real understanding that God is a transcendent

concept. They may well think of God as a real person, complete with arms, legs, a head of hair and some preferences for what to watch on television, and would stand on a cloud making animals out of parts tucked away into a desk draw. Contrast that idea with what adult Christians believe God is and suddenly the phrase “I believe in God” is not really mentioning the same thing *at all*. Believers might want to wiggle out of this by saying that the child's idea is incomplete, but “at least it is there”, but that doesn't actually respond to the accusation that the child isn't believing the same thing; if it's not complete then it plain and simply is not the same concept as formed in the adult brain! That's the *point*. Do we therefore conclude that our 6 year old wasn't a Christian? Certainly their view isn't in line with what the adults think when they say “I'm a Christian”, indeed the description above is more remarkably different to the adult Christian view of God than this adult Christian view is to Hindu or Islamic views. Could we even go as far as to say that, given the definition of God agreed upon by adult Christians, the 6 year old is, for all practical purposes, an atheist?

Belief in God is one thing, and one might want to worm out of realising that a younger person's view of God is totally different to an adults. Apologetics and theology is all about trying to desperately hand wave away these kind of implications and wouldn't accept them. There are other examples where we *would* accept this line of reasoning, and as a result even encode it into law. For example, what is a 6 year old child's view on sex? We can almost feel the fury building as such a touchy subject is raised, and can almost hear the frantic bashing of keyboards as angry emails are written at the mere mention of this! But really, what does a 6 year old view it as? It's important to understand this because it's the founding idea behind one of the more important moral caveats we have to the near-sacred idea of “informed consent”.

Ultimately, the punch line is that the law judges minors to not know what sex means in the context that an adult understands it. In fact, this isn't just a legal accident, children really don't develop the necessary brain functions to make certain judgements until around puberty. It's physiologically embedded in them that their knowledge and interpretations of the world will not match up with those of adults. We recognise this, and so don't let them drive, vote, fight wars or have sex with adult members of the population – even if they do actually say yes. Informed consent implies that someone gives their consent, that they are okay to proceed, for an activity whether it be sex or medical treatment. More importantly, it implies that they are *informed* and know all the associated risks, consequences (good and bad) and are prepared to accept them should something go wrong (or right!). A minor can easily give “consent” to any activity they choose, but we deem children incapable of being informed enough to make a choice of informed consent precisely because we don't believe that the cognitive processes that they form will match those of an adult. This fact is realised enough in this case to encode it into law.

These are all issues that arise from the fundamental fallacy of only holding that  $Q : \Leftrightarrow q$  holds for all people equally, but also because some people think that the inverse is true, and that  $q : \Leftrightarrow Q$ . E.g., that the phrase “I believe in God” is logically synonymous with the thought process associated with belief in a higher deity, specifically of the Christian tradition. This is emphatically not the case as it is reflected in reality.

Despite these numerous asides and explorations, we are beginning to build up a picture of what is wrong with the discussion about feminism and sexism being held by our Ageless-Faceless-Gender-Neutral-Culturally-Ambiguous-Debating-Persons. Their problem goes somewhat beyond how they simply *define* their terms, as the definitions themselves

may have some other hidden connotations associated with them. These connotations are constantly informing and biasing what images are conjured up in their head while they're discussing topics such as “feminism”, “equality”, “sexism” and so on. These connotations do not match up in each others heads. Imagine one of them, it doesn't matter which but again we can presume your prior ideas and imaginings of this hypothetical pair informs you of which one, realises this. They, completely out of character for a professional debate enthusiast, come to the sudden and correct conclusion that they're talking about completely different concepts. Their disagreement over the hypothesis “feminism leads to more sexism” stems entirely from how they're defining their terms. Good, it's an epiphany of a kind and some progress at least. This realisation is the difficult first step on the road to true clarity in communication.

So, the discussion by our two protagonists (remember that you still have full reign on how you're visualising them, if indeed you are) takes a short, sharp, but very significant, step... backwards. One asks; how do we *define* feminism?

“Well...” says the one arguing against the hypothesis that feminism causes sexism, feminism is just ideas about gender equality, “It's this notion, hidden from us by history, that men and women are all equal in the eyes of reality. In some respects it's the bizarre and strange notion that women are people too!” Audience applauds. Okay, so that is what they are talking about, and they've suddenly realised that this is different to the other, and is the reason that they have clashed so much. Presumably then, the other agrees with the definition and allows them to move on?

“I disagree!” screams the one arguing *for* the hypothesis, and that feminism causes sexism. So it transpires that in their definition, feminism is synonymous with the idea that women are superior, and need to adjust

their level of rights by restricting those of men. In this definition of the idea at hand, equality isn't the goal or the method, and so the hypothesis fundamentally follows from it. The audience probably boos at this. After all, this is accusing people who identify as feminists as complete and utter berks. Our debaters have realised they're not discussing the same thing, and that their views on what feminism *is* differs. But then what happens?

Anyone who has ever reached this stage in a discussion will know exactly what comes next: *arguing over definitions*. The emotional attachment to the word takes over, because even though people have realised their definitions clash, their attachment to the cognitive-linguistic map remains.  $Q : \Leftrightarrow q$  remains firmly intact for both parties and the cognitive clash remains, translated through the medium of linguistic clashes that are being incorrectly translated back into the wrong qualia for both parties. This is what leads to arguing by definition: or more precisely, arguing that *my* definition is the One True Definition. Even reaching for the dictionary doesn't help because often the terms that cause this sort of problem are the terms that are beyond a mere dictionary entry to describe. These are the terms that have entire histories, sub-movements, people and alternative interpretations. More importantly, they also have the hidden set of ideas that no one really admits to when they give their “definition”. This is problematic and yet still going back and defining your terms is often thought to be one of the most basic tenants of how to host a proper and meaningful discussion with someone. Just as you can't discuss morality until you've put on the table what your ideas of “Good” and what your ideas of “Bad” entail, you can't have a meaningful debate on feminism and sexism without agreeing on a definition of both, and then how the two concepts interact. You're still asking a closed, rather than open question, and so the definitions are still contributing to the arbitrary nature of it.

But where to go from there? If two parties are claiming ownership of  $q$  for themselves, asserting that their map is the One True Definition, how do you break the tension? Left to fester, such a conversation can be at best unproductive and at worst actually *reinforce* the bias that someone has that their own definition is true. Our Ageless-Faceless-Gender-Neutral-Culturally-Ambiguous-Debating-Persons will actually end up learning nothing from each other, run out of patience with their stubborn opponent and go away assured that they're still correct because any *right thinking* person would never have disagreed with them. “I can't believe that they refused to accept that feminism is just gender equality!” one cries out when recounting the failure of the debate, “what a bigot!”. The other is equally vexed by the experience “That person still thinks feminism is just gender equality, it's obviously more than that, otherwise why would we have that word!” Both are, to a degree, very correct. Yet they would probably never realise this.

The simple fact is that despite attempting to hammer out definitions in advance both parties are biased by their prior experience with the terms they are using. They have approached it laden with these ideas already formed, found that they “respectfully” disagree, but failed to lose their attachment to the simple terminology and the fact that these words don't represent what they are saying *when viewed from the perspective of their opponent*. There's nothing particularly special about singling out “feminism” as an example here, except to say that it's a remarkably easy term to play with when discussing this concept. There are a massive range of views of what it means; it's not a case of just what dictionary definition someone wishes to use, or the fact that this One True Definition has been twisted. It simply the case that people are bringing their own hidden meanings and connotations along to the table and treating them as intrinsic properties.



Remember that in the cognitive-linguistic map,  $Q$  stands for all raw thoughts associated with a concept (indeed, it's quite fuzzy even to categorise it like that) and no more represents a mere one or two sentence long dictionary definition than  $q$  represents a single letter of the alphabet!  $Q$  contains all these connotations, hidden meanings and subtle differences between synonyms. Oh, and they only can ever be sure to apply to yourself and you can't really prove what someone *else* has associated with a word – no psychic mind meld available, remember. All of these factors are informed by prior experience of a term, and are difficult to shake off, and so are very real to the users. They pollute what the two-way process of communication uses to make translations between the system of thoughts and the interpretation of words.

The communication scheme  $Q_a : \Leftrightarrow q \Leftrightarrow Q_b$  represents not just the obvious differences between  $Q_a$  and  $Q_b$  but all the subtle connotations that are along for the ride. A person with experience as a gender equality activist that treats male and female issues identically and with equal importance will associate the term “feminism” with gender equality of that style. A person with experience of those protesting specifically women's rights as exclusively more important issues will associate “feminism” with something more gender specific, rather than gender equal. There is no judgement being implied here, it's simply describing two different approaches to what people *think* is the same idea because it uses the same word. Someone who's experience of a *feminist* is Andrea Dworkin won't quite agree with someone who's experience of a feminist is Maggie Mayhem. Very importantly, in most cases when people experience new things from outside their prior experience, they don't intentionally mutate  $Q$  to match (such a thing tends to be a slow process and biased by a few other factors, not least confirmation bias), but perhaps shunt it all into a different set of thought patterns entirely,  $R$ . This

is easier than completely realigning their thinking, and confirmation bias and cogitative dissonance are two effects almost certainly at work determining this. So many fallacies and false arguments also arise because of this principle; the old canard of No True Scotsman, where someone doesn't *really* belong to a group because they display abhorrent behaviour, or the phrase “I'm not homophobic, I have a gay friend and he's alright, but...” where someone's prejudice stands in face of evidence.

We can formalise this concept, because equations and diagrams are fun, a little like this:

$$\begin{aligned} Q_a &: \Leftrightarrow q \Leftrightarrow: Q_b \\ R_a &: \Leftrightarrow r \Leftrightarrow: R_b \\ (R_a + Q_a) &\Leftrightarrow (R_b + Q_b) \end{aligned}$$

That is, even though persons A and B might disagree on how they categorise thoughts that relate to  $q$  and  $r$ , they still actually agree in the grand scheme of things, and if they were to redistribute their thought processes they would be able to cognitively categorise in the same way, and possibly agree with each other. They still possess the information required to communicate effectively, combining  $R$  and  $Q$  for both of them leads to the same result. No True Scotsman, therefore, only exists in their mind.

With respect to our featureless protagonists in the feminism debate (remember that you're still allowed to age, gender and give them ethnicity them yourself. What are you visualising? And why?) they both clearly know about groups of people who are for sexual equality and groups that are for female specific rights. These cognitive categories exist for both, if you push them for enough specifics, but they've both generated sweeping

generalisations and mapped them into a broad linguistic category. There is nothing, *per se*, wrong with this apart from the laziness of using too broad a category, as categorisation is a key part of how our minds decide to communicate with each other. The trouble starts when they use the *same* linguistic category ( $q$ ) for their *different* cognitive categories ( $Q_a$  and  $Q_b$ ) and so have bundled their own subtle connotations and implications in with the word that they choose. Hence why simply arguing about definitions gets you nowhere. Just to contrast this, imagine if it was a far simpler and more blatant mapping error like this:

$$\begin{aligned}Q_a &: \Leftrightarrow q \Leftrightarrow R_b \\ R_a &: \Leftrightarrow r \Leftrightarrow Q_b\end{aligned}$$

In this case, it would probably be easier to argue from definitions, simply because the word is transferred around, but the connotations deliberately stay within their own thought groups. This is what people *think* the error is when they argue from definitions, but the reality is almost invariably a different thing altogether. The illustrative and hypothetical feminism debate is not about one party misunderstanding the definition of feminism, it is about how both parties have disagreed on how they are categorising sets of attitudes, how they attach their hidden positive and negative connotations to this, and then how they have dragged these across into the words they have used. These words then generate a different set of qualia in the opponent, and they judge it as an attack on their own values. Simply saying “for the purpose of this discussion we will define it as...” doesn't help because our debaters cannot let go of their emotional attachment and will be constantly biased by their own use prior to the debate and the “for the purpose of this discussion we will define it as...” phrase.

## Part 4 – Implying connotations

*Avoiding the hidden pitfalls of careless use of language – in other words, say what you bloody well actually mean, not what you've been taught is supposed to mean it...*

So how can a discussion proceed? By this it's implied that the discussion proceeds in a constructive, useful way, where both parties actually form an effective communication scheme and, because the hypothesis like “feminism causes sexism” is clearly a matter of empirical fact once the definitions problem is solved, come to an agreed conclusion. Running to a dictionary as the first line of defence clearly doesn't work; it doesn't solve emotional attachment to a word and it doesn't eradicate hidden connotations and biases. Similarly, inventing any new word probably won't work because, regardless of what the word *is*, a cognitive-linguistic map must be formed that will start to include all biases, connotations and subtle distinctions. At best these will be new inferences, at worst you just bring across the old ones and have moved nowhere. It's a tempting idea to try, but it doesn't necessarily work. You simply can't just swap *q* around for something else and think that solves the problem as the problem is in the mapping itself. Existing words only approximate thought (at least with respect to two-way communication), and so adding new ones muddy the waters further. It doesn't stop the generation of new

words being a useful illustrative exercise, whether it be an entirely made up word like “seleshquob” or Richard Dawkins' coining of the typo-like “theorem” to try and get around the connotation that the word “theory” is synonymous with “guess”. But the limits of this must be understood – where the Dawkins example fails is that the connotations associated with the first term may well be drawn into its illustrative replacement regardless of how well and rigorously it is defined first. His readers would *know* that he intended to replace the word “theory” with what looked like a spelling error of a mathematical term as he mentions this explicitly. While he described the actual term “theory” with brilliance, mapping it onto a new word and merely *suggesting* that it was there to replace “theory” automatically conjures up the connotations and implications of that word; namely, that no matter who many times you insist it means “well tested body of predictive statements founded on empirical observation”, people will still think it means “wild ass guess”. A true synonym isn't just a synonym of definition, but also of connotation.

Running to the dictionary is out, running to the thesaurus is also out, and finally running to the name-generator is also out. This leaves us, once these options are eliminated, with the simple case of evasion. This seems like running away from the problem, but it is the route to solving it. Because our connotations are drawn across the cognitive-linguistic map, avoiding the problem terms that create this draw in the first should be the first port of call, not simply redefining terms. In the first instance, this removes the lazier mental shortcuts people use when communicating that eventually lead to making vast sweeping generalisations. So in the hypothetical debate, it's best to not just redefine a term, but simply don't use the term “feminism” at all, use the term “people for gender equality”. This isn't to say to use it in the sense that “feminism = people for gender

equality” as some propose – the suggestion is to literally take the left hand side of that expression and scratch it out, ignore it, forget about it. Even further, it would be good to scratch out “gender equality” and replace it with what it actually means to you, as if you were to ask 10 different people you'd probably get 7-8 different responses. By actually replacing a term with what it is supposed to mean, we can start on the quest to remove the excess mental shrubbery that people carry with them when they use their own pet terminology on their own terms. Similarly, do Christians annoy you? Well, scratch out the word Christian and replace it with what you mean. “People who enforce their arbitrary religious beliefs upon others through legal protest annoy me” is far more productive an expression, and might help you identify why you'd also say “Muslims annoy me” or “those New Atheists annoy me”. This act does not simply redefine or replace  $q$ , it forces an exemplification of  $Q$  from the very start by removing the general map of  $Q : \Leftrightarrow q$  and replacing it with perhaps a larger number of much smaller and more precise maps – how and why this works will be covered a little later. The advantages of precision and specifics outweigh any loss in conciseness.

But this is not the only application of the idea of removing problematic terminology. Removing the lazy mental shortcuts is also of benefit to an individual themselves, as they can crystallise their ideas more clearly and without reliance on jargon. If someone responds to a problem with the phrase “we need more common sense”, it isn't necessarily obvious what they mean. Indeed, even to themselves they may not know what they mean and could just be parroting a particular politician or newspaper column they happen to enjoy and mostly think that they agree with. “Common sense”, as seen elsewhere, is pretty unreliable for the most part. But again, that depends on the definition of common sense and what connotations are associated with it. When

someone says “we need more common sense”, what are they *actually* thinking and implying by their use of those 4 vowels and 7 consonants split into two words? Obviously there could be some air of being an “everyman” in there somewhere, that they're part of the common people with common sense and they don't need any complicated procedures to live by. “Oh, we don't need all those rules, we have common sense!” they might rejoinder to someone when faced with a massive rulebook dictating how to act in situation A, situation B, situation C and so on, in the belief that their one simple rule of “common sense” will save them *because* it's simple. Or there could be other negative associations with it.

But we can expand the idea of hidden connotations and hidden meaning to other terms in a particular statement, not just the obvious like the nouns that beg for someone to put forward a definition and a specific category. When we say “common sense is unreliable”, we should be able to rewrite it without using the term “common sense” *and* “unreliable” at all. So if we just avoid the terms, and by extension the positive or negative connotations associated with it, we arrive at a more meaningful statement that is not only empirically testable but also clear. We can instead say: “Our basic rules-of-thumb developed for us to respond rapidly to previously solved problems do not necessarily apply, or can produce misleading results, when applied to novel problems for which pre-existing correct solutions are not available.” Wordy, but precise, empirical and far more useful. And it's all down to first realising that how we use the term “common sense” defines our answer.

On a roll, let's look at “marriage” for a moment, particularly in the context of same-sex marriage. Why should gay people be allowed to marry? Well, that's probably a no-brainer, so more specifically why should it be allowed given that there is the option of a “civil partnership” instead? Examining this question we can look at the connotations

associated with the words and phrasing used. Principally in the word “marriage”, which has (erroneously, considering history) been associated with loving, respectful relationships that last for years, and all the pomp and circumstances that are associated with a wedding, from the dresses to the cake to the drunk uncles you kind of *had* to invite. Civil partnerships, on the other hand, seem to conjure up this idea of people getting married only for legal convenience, the associations with such spectacle simply aren't there. It might be *desirable* for such connotations of spectacle to be avoided, particularly for those growing cynical of the concept of marriage as described above (as cases of heterosexual civil partnerships can illustrate) but this is irrelevant to the question of why just one demographic should be denied access to one and not the other. Civil partnerships and marriage may have the same legal definition, and even dictionary definition, and even the same definition if someone wants to say “for the purposes of this conversation will we define it like so” – and so on the face of it someone might appear correct to say “why the fuss over marriage when civil partnerships are available?” but those additional connotations are *not* brought along for the ride automatically when conjuring up qualia with a new term. This seems a contradiction to the discussion of Dawkins' use of “theorem” above where the accusation is that the connotations *do* come across. But the wider point is that you cannot truly hold control, unless you work up to your terminology very carefully, what connotations are brought across when another person interprets your term. In the case of “theorem”, the mere hint about it being used as a synonym of theory has firmly planted the connotations of “guess” and “unfounded” in the eyes of detractors, for whom such connotations will be second nature. With “civil partnerships” and “marriage”, those campaigning for exact equal treatment will actively resist connotations being brought across, because that is what their



experience and their desires inform them. Indeed, they might generate new connotations to go with the new term or bring over ones from similar words. Certainly the term “civil” implies some form of soulless law making, while “partnership” seems similarly devoid of emotional meaning and somewhat tainted by its own generic character. You can try, but simply influencing the words and definitions doesn't easily influence the cognitive grouping itself;  $q : \Leftrightarrow Q$ , isn't the way around that it works. So even with precise legal and dictionary parity between marriage and civil partnerships ( $m \Leftrightarrow cp$ ) you don't have precise parity between peoples' *concepts* of such things ( $M \Leftrightarrow !CP$ ). This parity, that is functional equivalence, might exist in the realm of the real world (the influence of which is discussed a little later), but not in the interpretation of it *as used by real people*. Does this mean that they're wrong, and should get over it? Certainly *not*. Words should reflect our use of terms, not the other way around. Definition-level parity between two terms ( $m \Leftrightarrow cp$ ) means that in the cognitive-linguistic map they can be effectively summarised by a single linguistic term (this is the apparent aim, is it not?), in this case we substitute them for a generic  $q$ . So this actually leads to this situation:

$$\begin{aligned} M &: \Leftrightarrow m \\ CP &: \Leftrightarrow cp \\ (m \Leftrightarrow cp) &\Leftrightarrow q \end{aligned}$$

And from this we can generate the ambiguous communication scheme below:

$$M : \Leftrightarrow (q) \Leftrightarrow CP$$

Which is exactly the usual mapping error described above that is present in normal miscommunication! The fact that we are using different terms has caused biases, preconceptions and connotations to be built around the terms in the cognitive side of the equation. So the phrase “but gays can marry, they have civil partnerships” is, unfortunately, *wrong*, because  $CP \Leftrightarrow !M$  no matter how much we try to work the words around to mean something different. It is correct only as a legal fiction, not as a reality inside someone's mind, where it truly matters.

So back, finally, to the hypothetical feminism debate by our featureless protagonists – still imagining one as a gray haired man, tenured professor of humanities with a dismal and dismissive attitude, faced off against some liberal young woman who moans constantly and thinks all sex is rape? Oh well. In this situation we find that the hypothesis “feminism causes more sexism” is useless as stated because it doesn't tell us what to look for, as both cause or effect. We need to completely remove the problematic terms and produce something meaningful. It doesn't have to be too wordy, although this tends to be an inevitability, but it does have to remove all troublesome statements with strong emotional attachments, ambiguity or potential for mistranslation across the cognitive-linguistic map. So let's consider this instead:

*Campaigners, authors, journalists and activists focusing on the promotion of equal treatment of both sexes, and those not associated with the gender binary, under the law and under the prevailing attitudes of others, lead to less unfounded and unnecessary favouritism for or against either men or women on grounds of their sex or gender.*

“Well...” they say, scratching their chin and deep in thought as if

something profound had just been said “...when you put it like that I suppose we're in complete agreement.”

## Part 5 – Cognitive architecture

*Building a model of thought so that it can relate to what we say with precision. If we have the time to spare, that is...*

The thesis as it stands is thus:

- 1) Words form only a representation and interpretation of actual thoughts.
- 2) This representation is formed through a cognitive-linguistic map, relating thoughts to words (or symbols, if you wish to extend to the most generic case).
- 3) The main route for misunderstanding, and unconstructive dialogue, stems from how this map differs (sometimes minutely) between individuals.
- 4) Failure to recognise this probably stems from the importance put on words and language as an end in itself, rather than a tool.

So far, the cognitive-linguistic map has been dealt with in a fairly abstract sense.  $Q$  represents the generic thought while  $q$  represents the generic word or phrase to represent  $Q$ . A few other letters were brought in to try and get across greater differences; such that we can be sure that  $R$  and  $S$  represent significantly different thoughts, but that  $Q_a$  and  $Q_b$  are similar enough but still ultimately different in a way we can't tell until we hit a stumbling block in a discussion. Obviously, it is possible to expand

upon this considerably and explore not only what they mean in the abstract, but what these letters are actually composed of. Much, in fact, the same way that complex mathematical functions can be represented in a short-hand manner with a single letter, masking their complexity when a higher level manipulation is required and expanded again to their full form when we need to see what their effect truly is. Indeed, saying that miscommunications occur because  $Q_a$  does not equal  $Q_b$  in a conversation is almost tautological in its obviousness; of course people will disagree if their thoughts are, in fact, different. It's merely an assertion in its present form, if hopefully a convincing one. One could argue against the thesis above by asking where people get their definitions of words from in the first place. It obviously must come from somewhere, and people do – if vaguely – agree on terms. So maps between  $Q$  and  $q$  must not be as arbitrary as implied, and surely people must know how to pick the right word or phrase to represent themselves. Without responding to this and stating how thoughts can match up, and differ, in more explicit terms there is nothing *but* tautology.

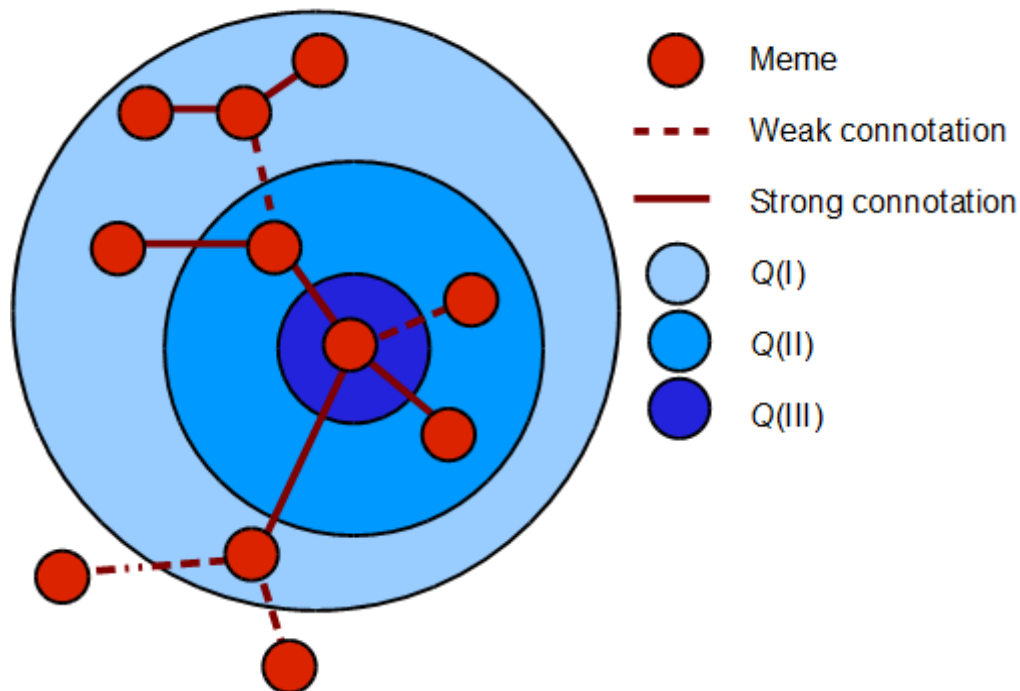
Establishing what  $Q$  actually is the important next step in more thoroughly describing how the map changes between people and *why* removing problematic terms entirely is an excellent tool for dealing with miscommunication, if one that is a little expensive in terms of raw brain power. Dealing with  $q$  and the  $q$ -space is easy enough – we're dealing with it right now. Because  $q$  consists of language itself it hardly needs exemplification as everything is practically an example of the linguistic space, we communicate in it so are likely to get the idea, especially if we're predisposed to mistake the map for the territory anyway. Dealing with the much vaster and more difficult to truly represent  $Q$ -space is something else entirely. One can attempt to describe a thought  $Q$  in terms of words, but that would still simply be a form of  $q$  again, and risks

mistaking the map for the territory as people assume a thought,  $Q$ , actually is a combination of words, even if they do “accurately” describe it. But no matter how precise, even if you can represent  $Q$  perfectly with 100% precision and no room for error and ambiguity when dealing with two-way communication,  $q$  still isn't  $Q$ .

In an ideal world, we would communicate our explorations of  $Q$ -space with a psychic mind meld. As it stands, we don't have such a tool at our disposal – otherwise we'd also have no need for the  $q$ -space to represent it in the first place. Yet a way to denote  $Q$  is still required to explore it, and for this a non-linguistic system, or at least one that avoids the stereotypical and obvious trappings of a language, is required. In this case a visual representation of how unique and individual thoughts might pack together to form a qualia in  $Q$ -space, and why a single word might lead to different interpretations when mapped onto another person. From this visual representation we can also see commonalities between definitions and how the unfortunate connotations can be inadvertently dragged along with a single word when translated by someone else's cognitive-linguistic map.

For this, we can consult a visual representation of something like a memeplex, a collection of individual units that join up to form a wider idea. This is the  $Q$ -space architecture that forms a thought pattern, which can then be mapped to a word or linguistic construct:

## Memplex within Q-space



A meme in this is the smallest unit of cultural knowledge, or perhaps the smallest unit of a thought. They've often been thought of as analogous to genes and that they can even undergo natural selection. This is an idea that is not without its detractors and controversy, but the idea of splitting thoughts into individual “units of cultural knowledge” and examining how they can attach and interact is a very attractive one for when trying to represent a wider thought pattern. It makes a nice starting point for how to assemble a visualisation of the interactions of things in *Q*-space, but a few caveats (mostly the reiteration that this is simply a useful model) need to be made clear.

Defining these individual units becomes slightly problematic if taken too literally; is “do not believe in God” a simple and fundamental meme? Perhaps, as it is a single thought when expressed like that, and is literally a yes/no answer to a relatively (well, deceptively) simple question. But attached to this particularly statement are concepts of what

“belief” means and also what “God” means. Even then, these terms may also have units within them to explain what they are. This is perhaps where the concept that language can't *perfectly* represent thought stems from; the paradox that you can't easily define a “fundamental” thought and yet an infinite regress of definitions seems completely and utterly absurd. Otherwise, how would you know what *anything* meant if there was an infinite regress? But that simple terms are, in fact, *deceptively* simple is a well established point. Indeed, look at the fact we need to remove problematic terms to proceed when a discussion becomes derailed by wars of definition; invariably they're replaced by longer, often *much longer*, sentences. As  $Q$  grows more precise and, in one particular sense gets smaller,  $q$  seems to increase in size. It seems paradoxical, doesn't it, that a smaller unit of cognitive thought requires a larger unit of linguistic thought? This is simply due to the fact that greater precision means trying to *exclude* connotations and hidden implications. A sweeping generalisation automatically includes them, at the expense of the fact that not everyone agrees *which ones* are included, while a precise description must purposefully avoid them and often specify which ones are excluded as well as which ones are included. And so we get an inverse relationship where, ironically, we need more words to describe a smaller portion of the cognitive  $Q$ -space. A more thorough treatment of this with information theory will come later.

This is a rough model, and a mere *representation* of how thoughts are organised that tries not to rely on language too much (in case it is mistaken for a description, like  $q$ ) it's not necessarily perfect and has its limitations, and so what constitutes an individual “meme” or point in the graph is in fact just whatever is necessary to form a coherent and formal model at the time. This is a hand wave, and something of a cop-out of a very important point, but until we develop psychic powers this is the sort



of modelling that we will be forever lumbered with. Computers don't have this problem, except where they need to start translating their own language into human-readable output.

In the quick outline of a potential generic memeplex above, specific labels of those red, meme-like, points have been deliberately omitted. Partially to keep the schematic as generic as possible, but also to enforce the idea that these individual memes are *not* individual words or, even more misleadingly, simple statements. As we've seen, simple statements may well be *deceptively* so, and highly misleading to have as a single point. Once this has been established, *then* we can start hazarding a guess of what the individual red circles on that graph actually mean in *linguistic* terms.

The next feature of the model is that you can have several versions of  $Q$ . In this case the more specific ones, encompassing fewer thoughts, are grouped in higher order.  $Q(I)$  is the broadest collection of associated memes in the web,  $Q(III)$  is the narrowest and most precise. These don't necessarily map onto different words, however. The cognitive-linguistic map may well come out like so:

$$(Q(I) \Leftrightarrow Q(II) \Leftrightarrow Q(III)) : \Leftrightarrow q$$

This may raise the question of why we should have multiple cognitive domains in this model  $Q$ -space if they map to the same word; but would that not be mistaking the map for the territory again? Three different domains of this  $Q$ -space and the cognitive realm no more need to map to three different  $q$ -space terms in the literary domain than the inverse is true. We might use the same word to represent different thoughts and collections of thoughts at different times, as our  $Q$ -space and the constructs within it being highly context dependent. Or, what if

we use  $q$  when we mean  $Q(\text{III})$  when we've previously been using  $q$  quite rough and haphazardly to refer to  $Q(\text{I})$ . People failing to notice this little switch is also a source of many communication problems. This is why replacing troublesome terms with greater precision is desirable. Take a look at the more involved map for various terms:

$$\begin{aligned} (Q(\text{I}) \Leftrightarrow Q(\text{II}) \Leftrightarrow Q(\text{III})) &:\Leftrightarrow q \\ Q(\text{I}) &:\Leftrightarrow a \\ Q(\text{II}) &:\Leftrightarrow b \\ Q(\text{III}) &:\Leftrightarrow c \end{aligned}$$

In this case,  $q$  is our sweeping generalisation. We have access to  $a$ ,  $b$ , and  $c$  so should use those to specify our thoughts more explicitly. In this case we need three linguistic terms, and because of the irony of excluding inferences these terms will be more complex than  $q$ . But we get what we pay for, we can now distinguish between three different subsets of thoughts, communicating them to someone else with greater precision than before because they won't get distracted by our sweeping generalisation. So instead of potentially ambiguous conversation:

$$Q(\text{I})_a :\Leftrightarrow q \Leftrightarrow: Q(\text{III})_b$$

We get precise conversation:

$$Q(\text{II})_a :\Leftrightarrow b \Leftrightarrow: Q(\text{II})_b$$

Of course, this now raises the issue of the fact that the second person must agree with the  $Q(\text{II}) :\Leftrightarrow b$  map, setting us back to square one. But the chances of this map matching up and the two thought

patterns overlapping become greater as the ambiguity is lessened and we focus in on more specific ideas. The ability for the hidden connotations to start polluting the actual meaning is reduced as they are purposefully excluded by such specificity. Categorisation in  $Q$ -space is unavoidable, and indeed necessary for brains and languages to work, but it is better to have good categorisation than bad categorisation; i.e., *relevant* and *precise* categories rather than *broad* and *general* ones. At least then we can begin to shut off the hidden connotations associated with a specific thought.

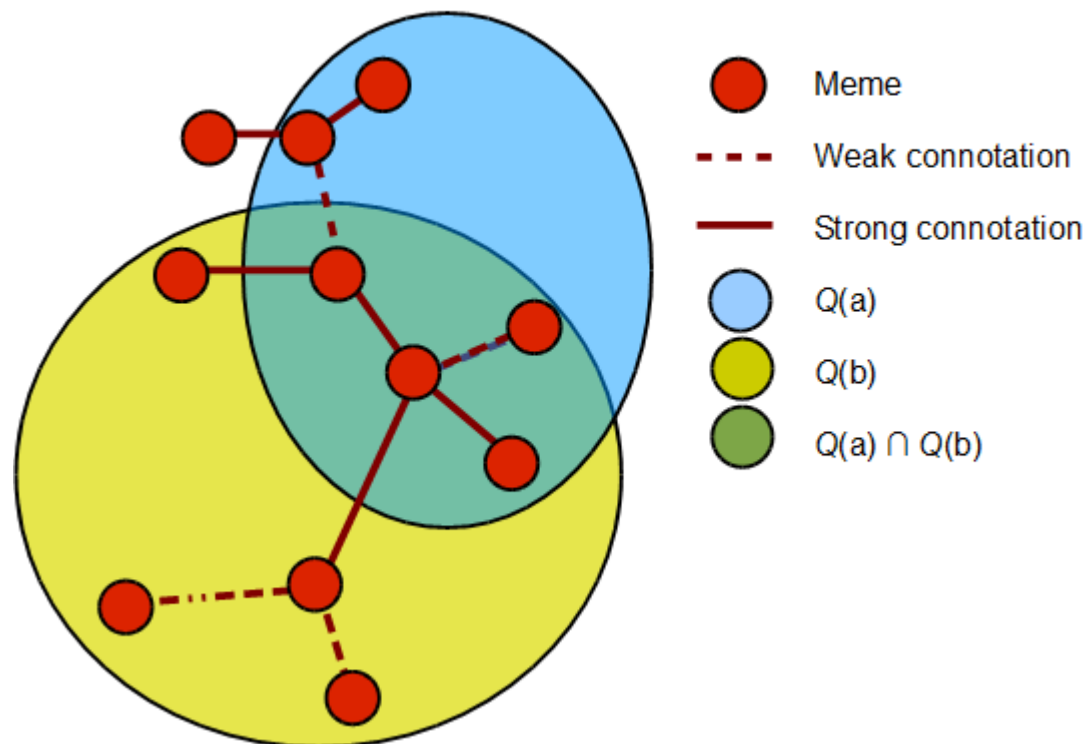
We have already come across connotations and associations when using words several times. In this case, the diagram uses connections to join ideas together, thus representing the connotations attached to ideas. A central one might be “self-defines as politically liberal” and a connotation to this is “pro-choice”. There's nothing special about the “weak” and “strong” lines, this is simply to suggest that some connections are indeed weaker than others – and to add more problems, some people the connections might be stronger than with others. A person might only consider “pro-choice” as only weakly attached to “self-defines as politically liberal”, but another may view the connotation so strongly that they presume them to be almost synonymous. It's these stronger connotations that are probably the most troublesome, not only being more difficult to exclude when switching to specific language but also leading to horrid things like unfair stereotyping (racism, homophobia, political ostracism and so on). Perhaps as a general rule, we can say that the stronger this connection, this connotation or inference, is the harder we have to work to exclude it when it comes to describing what we mean. If someone takes pains to describe only  $Q(II)$  in that diagram – only the red dots within the boundary marked  $Q(II)$  – then they will have to put more effort into ensuring they don't actually include the stronger connotations

that lie outside  $Q(II)$ . Importantly, the minimum effort required to do this is as much as the *strongest* connotation in all people involved in the conversation. Someone who *doesn't* view “pro-choice” as perfectly synonymous with “self-defines as liberal” will still have to make a good solid effort to distance them if talking with someone who does have that strong connection between the two. Failure to ensure this results in a mismatch of ideas.

If we imagine that an “atheism memplex” contains a central point marked “do not believe in God” and another for “do not believe in ghosts” then some may imagine a stronger connotation between the two, someone else might see it as weaker. Hence the debate about whether atheism, if you take it as a belief system must *also* include not believing in anything supernatural *at all* – ghosts, psychics, astrology, homeopathy and so on. Indeed, the connotation is there that God is bundled up with all things supernatural, but it depends on the user how *strong* that connotation is. In some people the bond between the ideas may be so strong that it is nearly *impossible* to separate the two concepts. This isn't a problem with reality, although it's worth asking the question “how is God special next to any other supernatural proposals”, but is simply a case of how someone chooses to categorise things and how strong the connotations are between them.

Here is the more important diagram, representing why communication errors can happen and what  $Q_a$  being unequal to  $Q_b$ , which has so far been dealt with in a more abstract sense, actually looks like:

## Q-space architecture of persons A and B



Now we're starting to get somewhere, although the reality is still far fuzzier. To get a more accurate picture add at least a thousand more red dots, hundreds of thousands of connections between them (all of different intensities that are again person dependent), then blur the big circles representing a particular version of  $Q$ . Oh, and have it move and evolve while people develop their ideas over time. Given this, you should be able to forgive the intense level of simplification in the schematic here. The fundamental miscommunication comes because the additional connotations associated with a central concept are mismatched between  $Q_a$  and  $Q_b$ . Yet there is still a significant enough overlap for A and B to perhaps not notice that they've ran into an error in translation involving those connotations. That intersection is enough to give the illusion that they're both talking about the same thing, when they might not be. A clash over the differences is possibly inevitable if these differences become highly salient. The Ageless-Faceless-Gender-Neutral-Culturally-

Ambiguous-Debating-Persons from the previous section know they are discussing feminism, and sexism, and so their  $Q_a$  and  $Q_b$  intersection will probably include things like “women's rights”, “gender equality” already (and subsets of memes that explain what these mean, although perhaps these don't entirely match) but then differ in the connotations attached to these that they're dragging in with it in order to define them. One has a strong connotation that gender equality means treating men equally to women, and that to do otherwise in any form is bad, the other has a strong connotation that gender equality means standing up for women's rights in the face of historical precedent to favour men and that simple “equality” is not practical given this. Again, it's worth reiterating that neither party here is actually wrong, they simply view things differently in terms of subtle implications – their *collective* mistake is to continue to group these things under the same word and continue the illusion that they're talking about the same thing. Both group these within a *similar* cognitive category, and map them to the *identical* linguistic category. Arguments arise because of the memes contained within  $Q_a$  and  $Q_b$ , but not their intersection,  $\{Q_a \Delta Q_b\}$  for short.

Here we're already getting into the mystic world of Set Theory, and in particular the realm of “fuzzy” sets, which is a fairer representation of the concept. A set is, almost obviously, a set or collection of items. In mathematics it could be a set of numbers like  $\{3, 6, 18, 19\}$  and so could be compared with another set like  $\{3, 12, 16, 19\}$ . Or it can be more descriptive and less arbitrary, like the set of positive integers or the set of real numbers and so on. In this basic flavour something is either part of the set or it isn't; it's distinguished by a membership test that we use to put something into the set (or not) given anything. The set of objects in the real world with two legs, for example, would have a membership test that looked to see if the object had two legs. A human has two legs, a

chair (at least most chairs) has four. This seems almost self explanatory but the set isn't always so obvious next to the membership test. For example, the set of objects called “cars” might have a membership test of “four wheels and an engine”, but that would also include buses – meaning the test might need to be refined. In keeping with the theme of replacing lazy mental short cuts with more substantial descriptions, reporting the *membership test* is usually more precise than reporting just the name of the set. Doing so either highlights where our mental biases are or if the set is even meaningful. As an exercise to the reader one could ask what the membership test for the set of “black people” is, and you might quickly find that concepts like “racism” cease to have any real sense.

Indeed, doing this for sexuality is what has lead to more refined ideas about gender and sex. Questioning the membership test for “homosexuals” for example, leads us to question if someone is a member if they sleep with someone of the same sex, but don't tend to form romantic attachments. Or perhaps the situation where a man only forms romantic attachments to other men, but still sleeps with women. It is out of questioning *why* we group people like so we begin to realise that many of the categories that seem to come naturally aren't actually very useful. This leads us to an unfortunate conclusion: the real world simply isn't built to be categorised as a hard set. Thus we need to be more flexible in our usage of categorisation as well as trying to apply more relevant and unambiguous criteria for the categories we produce. “Fuzziness” has often been the word used to describe this, and has a strict mathematical definition for being applied to numbers. This allows for more flexibility than the simply yes/no of whether something passes a membership test or not. It is the equivalent of taking the diagrams above and seeing them through squinted eyes.

So we can refine the cogitative-linguistic map slightly more, by

stating that the  $q$ -space and  $Q$ -space are both sets of some kind; one linguistic, one cognitive, both slightly fuzzy. Information is transferred between the different realms on the assumption that they have a degree of correspondence:

$$\{Q_a\} : \Leftrightarrow q \Leftrightarrow : \{Q_b\}$$

Using, or perhaps abusing, a bit more set notation we can attempt, in a hand waved way at least, to quantify the level of disagreement between Person A and Person B. We already know that the cognitive-linguistic map means that whatever thought  $Q$  is conjured up by symbol  $q$  cannot easily be altered because of the one-way mapping.

Unless we take measures to improve precision then  $Q_a$  and  $Q_b$  will be fixed and will be slightly different. So we need to directly compare  $\{Q_a\}$  and  $\{Q_b\}$  and ask under what conditions can “good communication occur:

$$\{Q_a\} \Delta \{Q_b\} < \{Q_a\} \cap \{Q_b\}$$

In words, this means that the difference between the cognitive categories should be less than what is common to them. But this isn't just in terms of sheer *quantity* of memes; there could be thousands of individual differences between two people's cognitive categories and yet people would happily get along. It is a matter of both quantity *and* salience, how relevant these individual ideas are to someone. It's likely that a religious believer shares more memetic components of their concept of “religion” with any other believer on the planet, and these thoughts might well be quite powerful ones too. For example, that faith trumps reason, that a supreme deity exists, the nature of this God to be



good, and so on. Yet the most minute of differences in doctrine are likely to be highly *salient*; where you worship, what you call *your* God, what you wear on Sunday. Hence blasphemy tends to be just merely the smallest deviation from established religious dogma – a mere speck of fly turd on a page to turn one letter into another, one word into another and a whole meaning on its head. It's this that possibly generates the greatest amount of confusion in non-believers: “Catholics and Protestants both worship the same God... so *where's the problem exactly?*” If the meaning is salient enough, then this difference can overpower so many similarities.

However, when we realise that we're actually talking about different sets, then the differences, in fact, become meaningless – as the existence of differences have been taken into account already by the fact we recognise the limits of the set. Just like how an open question, as opposed to a closed one, already takes into account our uncertainties in defining our terms, which would render the answer arbitrary in the case of a closed, yes/no, question. We can then start talking in more precise terms to accurately describe our thought *Q* and communicate it to the next person, avoiding any unnecessary warring over words.

## Part 6 – Beyond abstraction

### *Un-waving the hands and justifying cognitive architecture...*

Previously it was attempted to make a non-language based model for thoughts, the cognitive side of the cognitive linguistic map:

$$\{Q\} : \Leftrightarrow q$$

This was to avoid attempting to represent  $Q$ , a thought process, in terms of words and as such the diagrams were left fairly abstract with a lack of labelling. Attempting to describe a thought,  $Q$ , in words would merely produce  $q$ , a linguistic representation of that thought. The entire point here is not to mistake the map for the territory, which could result from thinking that  $Q$ , a thought, could be so accurately and easily represented by a string of words,  $q$ , that it actually was that string of words. This is of paramount importance precisely because the interpretation of the cognitive-linguistic map is person dependent, and cannot be confirmed to agree:

$$\{Q\}_a : \Leftrightarrow q \Leftrightarrow \{Q\}_b$$

And so with a mismatched map between two people, good

communication is not possible. So far this has been dealt with fairly abstractly, attempting to use terms and symbols to present the idea of a map in the most general sense possible. And again, when trying to represent  $Q$  in a non-linguistic way, more abstraction was required.

But what do we actually mean by abstraction in this sense? Abstract implies that something is not particularly real, as if it is just a model representation of what is really being discussed. Some abstractions might be difficult to describe, other abstractions may be difficult to think about. This is, of course, relative to the thing actually being talked about. We may think of thoughts as abstractions, because they're difficult to describe – but no one would deny that they actually exist. Alternatively, we may think of language as an abstraction because they attempt to describe thoughts, but only approximately. Across the cognitive-linguistic map,  $\Leftrightarrow$  represents a process of converting between two concepts that are abstractions of each other. Each is very much real in its own world. Neurons fire in particular patterns to produce a thought, words on a page are very much physical objects that encode information. If we work in one realm, the other is abstract. As such, it's all relative here.

This abstraction has lead to one of the key assertions made about language: that words cannot, with total accuracy, describe thoughts and feelings. Examples are abound to demonstrate this. Can one describe the concept of love? Or fear? Or effectively describe any emotional response? Because of this, we sometimes think of thoughts as somehow “abstract”, we simply can't write them down and clarify them. It is said, therefore, that no number of words can fully describe what someone is saying and so ideas can only be stated approximately. However, it's important to go a bit deeper into *why* this should be the case, and the appearance of relative mapping errors between two individuals more concisely explains this apparent assertion much better than the idea that

thoughts simply *cannot* be written as words.

As mentioned before, Ace wants a beer. Along with this desire are several connotations associated with it. The feeling of thirst, the desire to drink more booze, the favourite brand. How this feels mentally and physically, memories of the last time he had one, and imaginings of future circumstances where he'll want one again, are all wrapped up with this thought. This thought goes beyond what is “strictly” implied by the phrase “pass me a bottle”. If we broke this down into a simple word game, it describes only three things with concise semantics. Firstly, the object (a bottle) then an request/action (to pass it) and a clarification to the request/action that states a destination (to me). But the connotations mentioned above are still attached; a request for something to drink *implies* a desire to drink, for instance, even though the phrase under examination isn't something like “pass me a bottle *because I'm thirsty*”, which attempts to communicate a reason. We would rightly conclude from this drunken thought-experiment that mere words cannot convey all thoughts and feelings succinctly, if at all. Indeed, the irony that a more precise, and arguably smaller, thought requires *more* words, possibly tending to infinity.

Suppose, however, that Ace recognises all the feelings he has at that exact moment and notes that what he feels can be described as a memplex-like combination of qualia. To move away from the general case, we'll call this set of thoughts  $\{D\}$ . Ace has now identified a set of thoughts, the cognitive side of the map and so, in principle, can precisely put this down in words. Like so:

$$\{D\} : \Leftrightarrow d$$

Perhaps earlier in the evening his thoughts were slightly different.

Certainly he would be more sober, but perhaps instead of a bottle he was more interested in a cocktail. Perhaps his taste buds were still functioning or he hadn't eaten so wanted something more substantial. All these connotations were there and Ace would be aware of the exact mental state he was in, and chooses to label this state  $\{E\}$ . Similarly, at this point in the evening he could represent this in the non-cognitive realm in exactly the same way.

$$\{E\} : \Leftrightarrow e$$

It matters little what  $d$  and  $e$  are. Ace may represent them as “pass me a bottle” and “pass me a cocktail” respectively. The fact that we wouldn't readily recognise these words as associated directly and specifically with all the thoughts in Ace's head (a complete representation of his mental state at the time) is not the point, at least not right now. He could make up two new terms, like “Iranu” or “Uvavu” and we couldn't argue that they *didn't* represent those two sets of thoughts  $\{D\}$  and  $\{E\}$ . Ace has, from his perspective, perfectly and precisely represented his mental state in words. He could do it in a dozen words, he could do it in another. He could write these words down in a diary and, at a later date, read and rediscover them, recalling exactly his mental state at the time. The words have encoded the information he needs to recall it because he's aware of his own map. Simply put, the assertion that “words cannot truly *represent* a thought or emotion, such as love or fear” is wrong. They can do this, because we choose to give them that meaning through a cognitive-linguistic map of our own.

With that assertion readily disproved it's time to, in fact, prove it again. This seems almost pointless, but is actually a key part in not necessarily proving the assertion wrong, but at least clarifying what it

really means considerably. How to go about this is probably obvious now; from *Ace's perspective* the map holds, and represents thoughts perfectly. He has mapped his mental state onto something in the physical world with perfect accuracy, from *his perspective*. The problem arises because that map may not necessarily hold true in someone else's head. They may guess, they may infer, they may use their own connotations, but they do not have the precise map that Ace has; they have their own. Particularly as he has made new words to represent his precise thoughts from scratch, it would be difficult for someone else to know what he was talking about without some way in:

$$\{E\}_a : \Leftrightarrow e \Leftrightarrow : ?$$

So while Ace has correctly encoded his mental state as far as he's concerned, Gabe can't decode it. They both understand the language perfectly fine. They can understand the syntax and basic semantics that converts the word game “pass me a cocktail” into the action and object (we can presume the overlap  $\{E_a\} \cap \{E_b\}$  is great enough for this to hold true), but the connotations of Ace's desires and feelings about *why* he wants it have remained hidden and effectively lost in translation. They both have working Enigma machines, probably with the right wheel settings, but their keyboards are wired up a bit wrong so that last step in getting the full *intended* meaning is lost. Gabe's own interpretation of  $\{E_b\}$  is biased by his own prior experiences instead, and may think that Ace's motivations are different to the ones he actually has, or at least thinks he has.

This seems like an awful lot of beating about to simply come back to the original point that “words cannot, with perfect accuracy, reflect thoughts”. True, but the point was to say that it is slightly more

complicated than that. To simply state that words cannot accurately reflect meaning is an assertion, to say they can't *communicate* because two (or more) individuals do not necessarily agree on the same mapping scheme is far more closely representative of what is going on with this assertion.

This is, incidentally, why simply replacing philosophically or politically troublesome terms with a brand new term you've just invented can present its own problems. You first need to communicate your map to another person, but then ensure consistently that the map agrees throughout. The temptation is to create this new word with an existing one as its base; creating nothing more than a synonym and risking dragging across all those extra connotations and inferences with it. This especially applies when using existing mental short-cuts to aid in a description of a new term, and so it needs simply to have these short-cuts peeled back so that the words used most closely represent the actual substance. Once thoroughly enough described, a new “term of art” can be created, but unless that definition is set in stone and all parties ruthlessly tested on their agreement of it (this is effectively what qualifications are) it is still open for new interpretations to be bolted onto it. In order to help describe something in a way that most closely resembles the actual substance, we need to leave the cognitive and linguistic realms behind and take a detour into some place that is entirely different; the Real World.

.