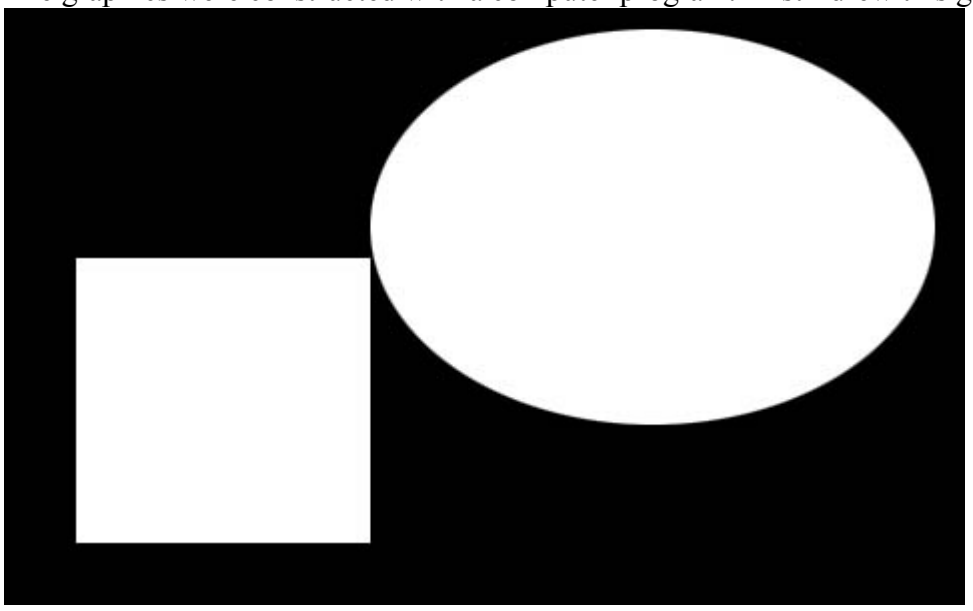
Words and Distinction
By Jody Hey
1999
(ICAPB, Univ. Edinburgh & Dept. Genetics, Rutgers University).

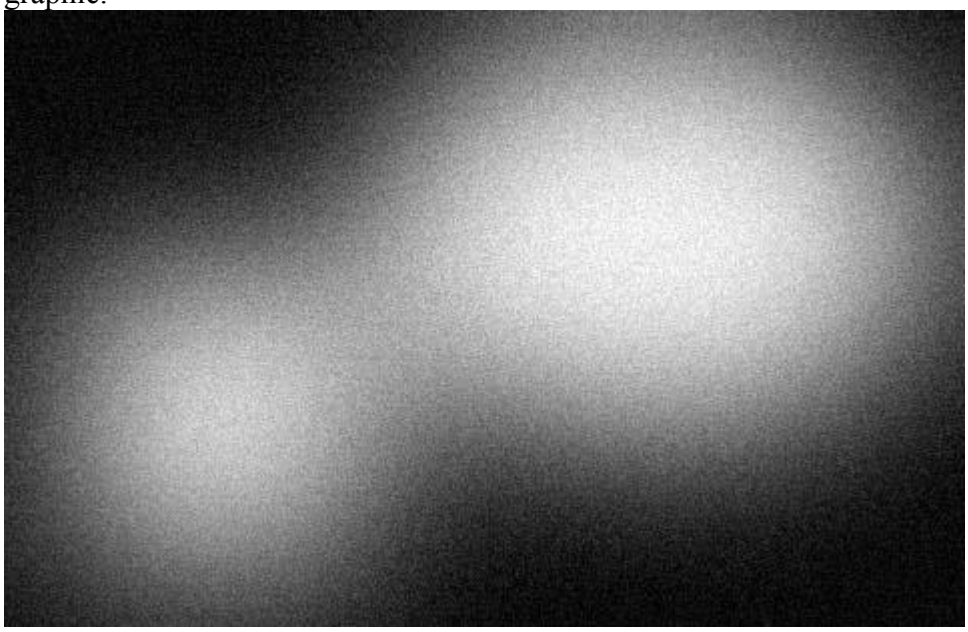
Introduction and Methods

This is a summary of the results of a small experiment conducted by email. The purpose of the experiment was to explore the relationship between apparent clarity of distinction among portions of a graphic and the words used to describe the graphic. The research was motivated by a wish to understand modes of references within scientific contexts (i.e. the nature of the real world and discourse about it) that do not harbor strong distinctions. While on sabbatical - clearly having way too much time on my hands, some would say - I have been thinking about vague words and vague objects that come up in biology. However the literature on the relationship among such things is tricky. On the one hand, there is a body of literature on vagueness as an attribute of language (e.g. Williamson 1994). But on the other hand, there does not seem to have been much done on how our language reacts to vague objects - on the ways that we describe nature when it does not come as seemingly distinct entities. Quine does have a bit of discussion on this, but not much (Quine, 1960). Nor have I found a literature in psychology on the topic. Certainly perception has been studied (edge recognition etc.), but my questions are more concerned with our descriptions of cases when edges are not to be found.

I figured one way to get a hook into the issues would be to conduct an experiment in which people provided descriptions of pictures that varied in the discreteness of their parts. With a sample of descriptions from different pictures, I could compare the words people had use and see what came out of it. The experiment was not strongly hypothesis driven (though I was surprised by the outcome) but was rather undertaken as an exploration.
The graphics were constructed with a computer program. First I drew this graphic.



Then I used the computer program to blur the picture, which resulted in the second graphic.



I emailed these pictures to a bunch of folks that happened to be on my list of addresses for one reason or another (nearly all are scientists). Each person that received the email instructions also received one attached file, a jpeg of one graphic or the other (never both). A total of 126 mailings went out for each picture (on Fri Jan 15, 1999), with random assignments of graphics to persons.

The primary instructions were as follows:

Please write down in words a short description of the graphic I have sent you. Your description should be sufficient to give a reader who has not seen the picture a fairly good mental image of it. Be terse, but use complete sentences.

These instructions were designed with the aim of placing the respondent in the particular, but familiar, context of explaining something to someone who cannot see what is being explained.
A total of 86 responses were received, including 41 for the non-blurry picture (sometimes denoted hereafter as N) and 45 for the blurry picture (sometimes denoted hereafter as B).

Pretty much without exception each response included a discrete block of text that was clearly intended as a response to the request. Thus it was a simple matter to remove unrelated text that was sometimes included, and to count the words in the response. Within the responses, only a little bit of editing was needed: beginnings like 'here goes' or references to my sending the picture were removed; '&' was changed to 'and'; '%' to

percent; and spellcheck was run. A few changes were made for easier word counting. Numbers like '1.5' where changed to 1 point 5. Similarly, expressions like '5x7' where changed to '5 by 7'. Apostrophes were removed.

Some responses included more than one description and these were taken all together as one response. Often we endeavor multiple descriptions when explaining something, so I saw no reason to tinker with responses that had multiple parts to them.

Results

The histograms of response length are shown in Fig 1. The mean for the blurry picture was 58.7 words, while that for the non-blurry picture was 74.2 words. The difference in means (or 'locations') is significant by a Kruskal-Wallis test of ranks ($H = 5.78$ $p < 0.025$). The variance for the N group (1601.2) was considerably larger than for the B group (599.7). This difference is also noteworthy ($p \approx 0.001$, $F_{obs} = 2.67$ $F_{0.001,45,41} \approx 2.7$).

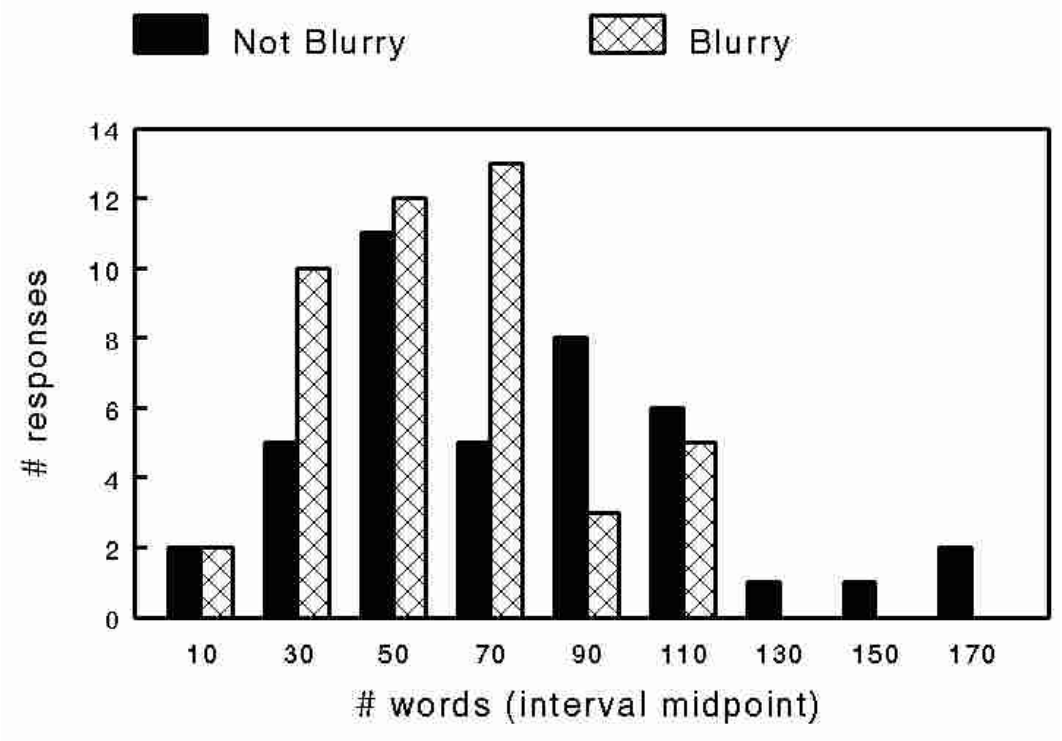


Fig 1.

I computer program was written to count up distinct words and generate histograms of word frequency (i.e. # of words used just once, twice etc., throughout each complete sample group).

Despite lower total word counts (2669 for group B compared to 3063 for group N), the number of distinct words (used at least once) was considerably higher in the B group (624 for the B group compared to 383 for the N group). These numbers are totals, and

not adjusted for the difference in the number of responses in each group (45 for B, 41 for N).

The difference between the two frequency distributions shows a clear pattern, with the low frequency classes having markedly higher counts in the B group than in the N group, while the N group was consistently higher in the highest frequency classes. The difference in the singleton class (words that appeared only once) is especially striking, see Fig 2.

The Shannon-Wiener diversity index for group N was 4.57, while it was 5.27 for group B.

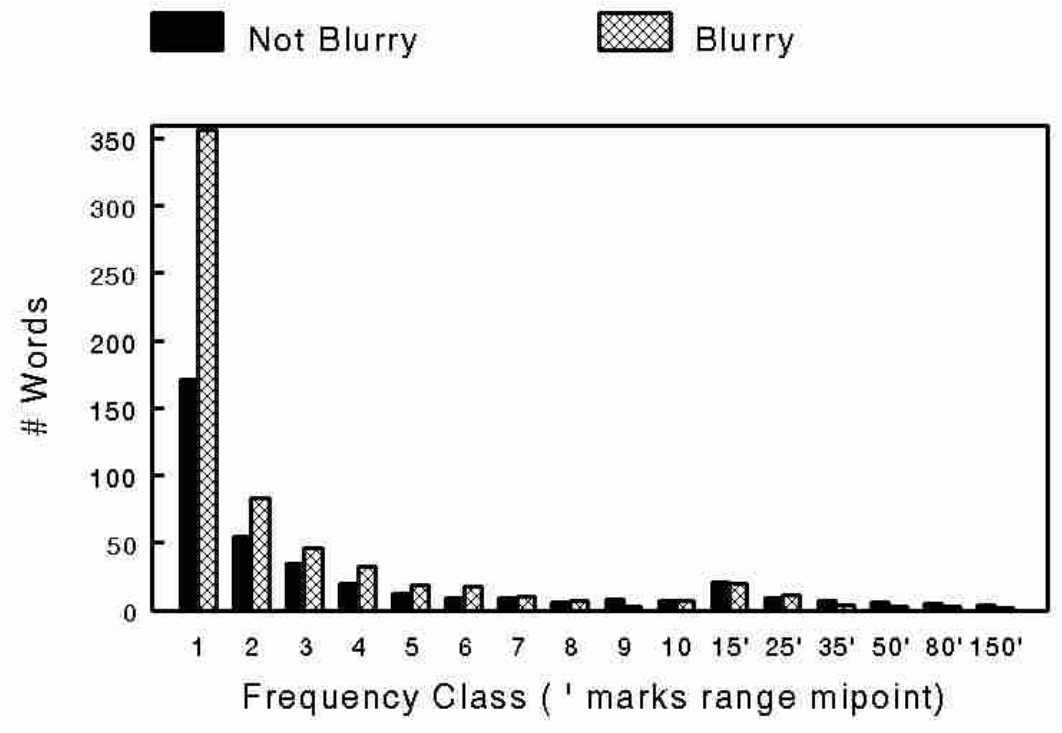


Fig 2.

Discussion

The purpose of this little study was to generate some kind of foothold for thinking about the ways that scientists give descriptions to things, and how these descriptions vary with the blurriness of their subject. Language is an inherently digital medium of communication. Whether written or spoken, words are made up of members drawn from a distinct finite set of subunits (letters or characters, and phonemes respectively). Also it is easy to think of noun-like words has having some kind of one-to-one correspondence with their referents. Thus I began this study with an *a priori* guess - a wrong one I see now. I supposed that the N picture was more ordered and simpler - certainly it was easier to draw - and that respondents would more easily find words to describe it. Thus I expected shorter responses in the N group. Instead the responses for the B group had a

lower mean length.

The main effect I had not foreseen was the efficient use of metaphor by the B group respondents. Almost every description appealed to another image that the respondent supposed was likely to be familiar to a reader. Here are some examples:
...two fuzzy whitish clouds against a black background...
... Looking down on two hills, or two piles of snow, with a valley in between them and plains around them ...
...someone's nude bottom placed on a Xerox machine...
... A car coming towards you. Two headlights on a dark, foggy night...
... the molecular force field of a molecule of sodium chloride in solution ...
Sometimes multiple metaphors were employed:
...They fuse where they touch resembling a fat bowling pin. It could describe a supernova explosion or a multidimensional plot...
...somewhat dumbbell shaped . . . like a gourd. The only other thing I can recall from memory that this looks like is that of an early stage of a budding event for some fungus . . .
Was metaphor the only route to description? Apparently not, as shown by this sample:

> *. . On a black field there are thousands of small, light dots. The dots are centered around two points; one in the upper right third of the field, and the other in the lower left. I'll call these two points "centers". As the dots superimpose upon one-another (increase in density), they increase in light intensity, so that there is a gradient from dark grey at the diffuse edge of the two centers to light grey at each center. . .*

I think the use of metaphor largely explains the frequency distribution difference (Fig 2), as there were many different metaphors and even among similar ones, the choice of adjectives varied considerably.

The two groups of respondents also differed in the extent to which they referred explicitly to their perceptions of the graphic. In particular, several in the B group began with something like "I see ...". Below is a table listing counts for a few words that were used in reference to the respondents perception

Word	B group	N group
see	8	2
impression	2	0
looks	12	2
seems	2	1

The reason for the wide variance among the N group responses is not clear. Reading through them the impression I received was that these respondents approached their descriptions as if they perceived it to be an orderly task that one could pursue with any level of precision. Thus in response to the instructions (.. give a reader who has not seen the picture a fairly good mental image ..) the variance in the descriptions seems to reflect variance in the detail that respondents saw as necessary to meet the requested criteria.

Here are fairly typical short and longer descriptions from this group:

> *This picture shows two white shapes on a black background. In the lower left is a square. In the upper right is an oval, with the longer axis along the horizontal. The upper right corner of the square is nearly touching the left edge of the oval.*
>
> *This is an image of two white figures on a black rectangular background, the long axis of which is oriented horizontally. The length of the rectangle is about 50 percent greater than the width. A white ellipse, with its long axis parallel to that of the rectangular background, lies near the upper right corner; its length and width (along the major and minor axes) are about 60 to 70 percent as long as the length and width of the rectangle. A white square lies near the lower left corner of the rectangle. Its upper right corner abuts the ellipse at a point slightly below the left terminus of the ellipse's long axis.*

But this explanation of variance among the N group does not by itself explain the difference between the two groups. Why did not the blurry picture respondents also react with similar variance? My best guess extends from the difference in the use of metaphors. Once a B group respondent had committed them self to finding a metaphor, they seemed not to pursue precision with the same vigor - nor variance in vigor - as did those in the N group.

This small study probably does not merit too much interpretation. I will add a couple comments about my sense of the results and what it seems the respondents have done to generate them. This sense is that the individuals in the two groups of respondents felt differently about their goal. The N group seemed to be addressing a straightforward task in an orderly manner, though with lots of variation on how in depth that task should be. The B group responses seemed to suggest that the writers had struggled a bit for a description, and then, having found something that would serve, avoided dwelling on more detail.

An important issue that this study does not address is how successful have the two groups been in generating "a fairly good mental image" of the graphics. For most of the N group descriptions I don't doubt that a reader could draw a figure that held the basics of the original. But this may be true of the B group as well. As any poet knows, metaphor is efficient and powerful stuff. An interesting followup study would be to ask some people to draw diagrams on the basis of some of the texts of the respondents (and ask a different group to judge the correspondence of the new diagrams with the originals).
Another question that a psychologist might ask is 'what did the respondents think they were doing?'. In my original email to people I said very little about the purpose. Academics think they know a test when they see one, and so the responses may have been different if the instructions had been, say, more positively misleading about the real goal. Many respondents did convey, in their description of the graphic or in ancillary text, that they wondered what kind of test they were taking. But I am not too bothered by this, as most scientists also value clever, pithy explanations in the course of regular scientific discourse.

Here are a couple quick thoughts on the design of the experiment. The 'treatment' can be seen as no-blurring/blurring. With one caveat, I think this aspect of the study was simple and reasonable. My major concern is that not everybody's' viewers worked the same way or generated diagrams of the same size. Probably this is not a cause of the patterns that I have described. For example, the non-blurry diagram, which was probably more easily drawn by most viewers, is the image that generated the higher variance. Also, there were very few responses that referred to the absolute size of the figure, or to colors that were not black white or gray. The sampling scheme is a bit more questionable. The population from which the respondents came is peculiar (my email list), and the responses were self selected. Only a minority of folks responded, of course. Quite a number of email addresses had expired or were out of date in some way, and many people did not bother to respond. However it is encouraging that nearly equal numbers responded in the two groups, suggesting that there is not a strong interaction between treatment (B or N) and the likelihood of responding.

Thanks very much to all those people who responded.
Jody Hey

References
Williamson T. (1994) *Vagueness* . London : Routledge
Quine, W.V.O. (1960) *Word & Object.* John Wiley & Sons, NY.