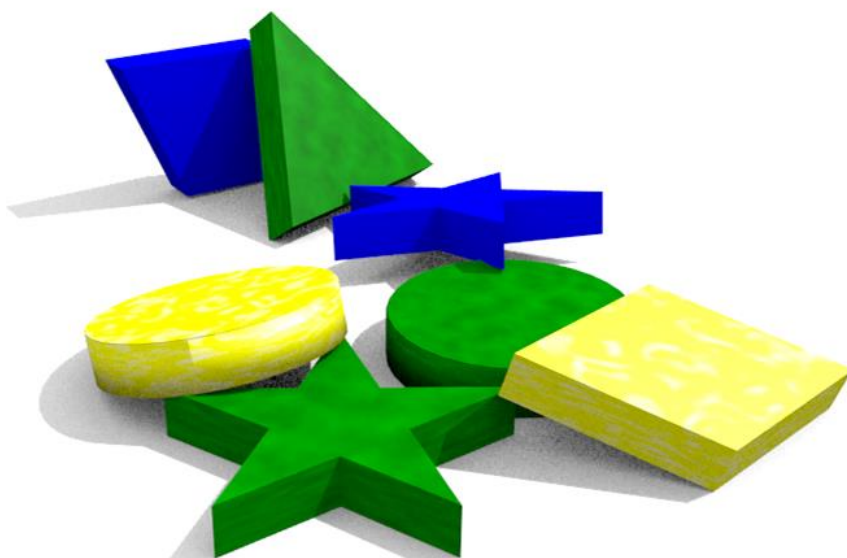


Lumpers and Splitters

In the studies of literature, biology, and software design there is a tension between those that focus on underlying similarities and merge items into larger and larger groups and those that emphasise obvious differences and partition them. For example, in linguistics there are periods when every new instance is seen as a distinct new tongue and other phases what were once thought to be separate languages are reclassified as dialects within a single one. In biology there are those that catalogue each specimen as a newly discovered genera and those that claim even the most widely dissimilar individuals just illustrate diversity within a single species. This phenomenon has long been recognised, more than 150 years ago Charles Darwin wrote "those who make many species are the 'splitters,' and those who make few are the 'lumpers.'"¹.

Take a look at this picture (on the right), is your first thought "shouldn't the triangle push the blue diamond over" or "why is the yellow star missing"? I suspect that each one of us has an innate bias, either towards grouping things together or towards distinguishing between them. The obvious temptation is to believe that this is a trait you are



born with. I know that my personal bias has always been towards consolidating things together (indeed on more than one occasion I have been told that I take this tendency to extremes). In my case this predilection has driven my choice of profession and even the topics I find interesting.

It might be overly fanciful but I suspect that extreme lumpers (like me) are naturally drawn to topics where complex behaviours arise from the interactions between a small group of fundamental principles. Topics like abstract mathematics, physics and the theory of computing. In contrast splitters are naturally interested in subjects that cannot be boiled down to the blind application of simple rules, topics where knowledge of a large body of details is essential. Things like history, biology and geology. That would certainly go some way towards explaining the "concept gap" I have encountered when attempting to agree on the principles of good data handling with certain clients and colleagues. However attractive such a characterisation is it cannot hope to explain everything, but maybe it can help illuminate one aspect of how we should communicate. Alternately it's all the deluded view of someone that sees simple principles underlying even the most complicated interactions.

So are you a lumper or splitter? Has your personal bias changed the way you approach the handling of technical data? Could applying a different style improve the eventual results?

¹ in a 1857 letter to Joseph Hooker