Group Comparison Formulations

© David Hewson 2002

Abstract

We often use group terms that label a set of members. And we often compare one group with another. This can lead to problems. Korzybski showed up one of the problems with group words, that of identification. I.e. treating all the members in the group as though they were identical. He proposed that they were all non identical. This gives us some information about how to cope with some of the problems. The aim of this paper is to take this analysis further and to propose several other solutions to the problems we get into when we compare groups. A recent example of problems with group terms has been the persecution of Muslims or Arabic people in the US after the terrorism of September 11. Other problems include sexism, racism, ageism, etc.

Introduction

I'm going to talk about the problems we have when we compare groups. Theses problems involve group words. I aim to show how Korzybski's formulations help but only go so far. Finally I will talk about what we can do about it and look in more detail at the mechanisms involved.

What do I mean by group comparisons?

When we ask questions or make statements that compare two or more groups or sets of members I see us as involved in group comparisons. E.g. "Chinese are smarter than Indians", "Women live longer than men.", "Jews are friendlier than Muslims.", "Old people are not as capable as young adults.", etc. As you can already recognise, a lot of these statements would be considered as racist, sexist or ageist, etc.

Is this still relevant today?

Well, in the wake of the September 11th I noticed that a significant portion of the population still has problems with groups. I.e. There was a significant backlash against the Muslim and Arabic communities even though the vast majority of these people had nothing to do with it and abhor what happened. We can over simplify things in our abstracting by the misuse of group words.

How important is this too us?

I think that the people on the "receiving end" of this sort of behaviour would regard this as important. I'd hope that most general semanticists find the reduction of this form of unsane behaviour as important also.

An example of these problems: If we believe that "Men are stronger than women" then we may only hire men for heavy lifting jobs. Korzybski's logical fate in action. This happened to one woman who couldn't get a job loading 50kg sacks onto trucks because she was a woman. She took the employer to court and showed that she could do the job.

Some issues that affect the discussion of this topic.

The multiordinality of words can affect our understanding of group comparisons. I.e. words can mean different things at different levels of abstraction to people. So, people can mix up the group with the member meaning. For example, here's some reasoning that at each step seems reasonable but overall resembles gibberish:

"No dog has six legs. My dog has four legs more than no dog. Hence my dog has ten legs."

The problem here of course is that "no dog"$_1$ is different to "no dog"$_2$. One talks about what is not in a class and the other about no member.

Other examples include: "No decision is a decision.", "These boys will be boys.", "Women should be proud of women like Curie and Kendig." I.e. the group name and members name are the same. To talk about things meaningfully we need to compare apples with apples and not apples with fruit. I.e. our meanings should be on the same abstraction level.

This leads us to the next problem: ambiguity. The statement "Men are stronger than women" can been seen as ambiguous. Does the person speaking it mean: some or all? While I see the statement "Some men are stronger than some women" as true, it doesn't say much apart from that there is some overlap between the groups. And the statement "All men are stronger than all women." I see as false. For the sake of clarity I'll use the convention that when a comparison is made of the form "Group$_1$ is better than group$_2$ on some attribute" I'll take it to mean that "All the members of group$_1$ are better than all the members of group$_2$ on that attribute."

Early solutions to the problem and their limitations.

One problem with group comparisons is identification. I.e. treating all the members of the group or class as though they were all identical. Korzybski improved on this by giving us non identification. His solution was to index. Smith$_1$ is not Smith$_2$. See *Science and Sanity*[2] chapter XIX. But this only tells us what the territory is not. It doesn't give much extra information about the structures we are interested in. I will use graphs extensively to show up identification and the problems we have with group comparison.

Figure 1 shows up the idea of identity and then non identity. But there can be many sorts of non identity. They give varying degrees of useful information. For example, when comparing the heights of men and women I can show up at least four possibilities. Starting from the left:

1. Identity, where all the females have just one height and all the men have another.
2. Non-identity, where there is a uniform distribution of heights that don't overlap.
3. Non-identity, where the distribution of heights is non uniform and no overlap.
4. Non-identity, with both a "normal" distribution of heights and a large overlap.
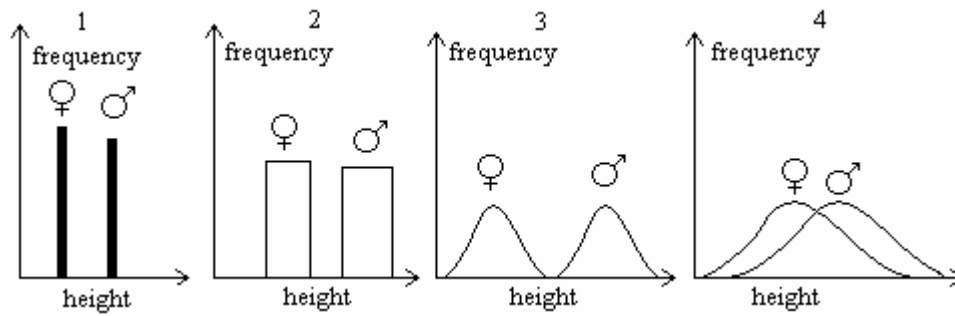
figure 1

As you can see there are many sorts of non identity but only the last graph is a reasonably accurate representation of reality. So non identity says there is a spread but does not give more structural information on how wide the spread goes or on the shape of the frequency distribution or whether there is an overlap between the two groups one is comparing.

Statistics has leapt into the breach and given us averages. However, comparison of averages between groups tends to fall to the abbreviation problem. For example: one could say "the average woman lives longer than the average man". One danger to watch out with this is our habit of shortening things or simplifying them by leaving off conditions. E.g. "Gasoline" becomes "gas". A "partial vacuum cleaner" becomes "a vacuum cleaner" which becomes "a vac", etc. So "the average man" becomes just "a man" or "men". And the saying "the average woman lives longer than the average man " (which is true currently for most nations) becomes " women live longer than men." which is not true as some die earlier than the average man. So identification strikes again when we treat the average as if it fits all. This again leads to stereotype thinking and prejudice.

What errors can "GS people" make due to the limitations of their use of GS formulations?

I have noticed in GS training that people tend to move from an Aristotelian to a non-Aristotelian orientation via the following steps.

Before doing a GS course some people have an Aristotelian attitude about Aristotelian formulations. E.g. it can only be either right or wrong. At this stage they identify the members in a group with the group. And to name a group gives them all the information they need (or can have) about any member.

After learning some GS they move on to having an Aristotelian attitude about non-Aristotelian formulations. E.g. it's **never** "only right or wrong". "Either/Or" thinking is totally wrong. Here they do not identify the members in a group with the group. And to some, to name a group gives them **no** information at all about any member in that group.

Finally some move onto having a non-Aristotelian attitude about non-Aristotelian formulations. They become more conditional. E.g. most of the time "only right or wrong" does not fit the territory. They do use "Either/Or" **when** that map fits the territory. I.e. they use the formulations in a more conditional manner. For instance in computing. The basic digital data is either 0 or 1. At this stage they also do not identify the members in a group with the group. And hopefully, to some, a group **may** give them **varying degrees** of information about any member in that group, depending on many conditions. This

information is never certain because of the "uncertainty principle". We are stuck with probabilities.

So the first two groups above believe that "Either a category tells you all about the members or it tells you nothing about them in comparison to another group." I want to move beyond this.

In this paper I try to show new ways of giving more structural information when comparing groups and using GS formulations in a more conditional manner.

Good communication tries to produce accurate maps for a topic in a situation to a person, etc. We also want economic maps. An error to watch out for here is that of trying to map to an inefficient level. I.e. I've seen some people trying to make their maps fit the structure of the territory exactly. Because "the map is not the territory" and "the map is not all the territory" we can never have a map that is exactly the same structure as the territory. I show up this idea of economy with the following graph.
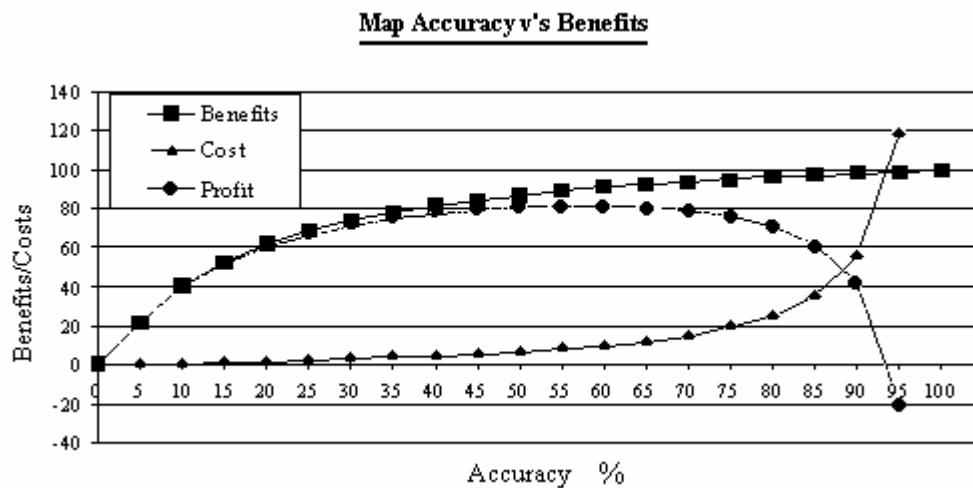

figure 2

Hence the most beneficial map is only about 60% accurate **in this illustration**. My solutions to the group comparison problem are like this. Not perfect but hopefully an improvement over what went before.

I have three solutions to offer. One graphical. One based on maths and operational philosophy. And the final one is based on English and ways to improve your every day language.

Graphics solution

Wendell Johnson saw graphs as: "the symbol of science". (*People in Quandaries*[1] page 125.) We can use graphics to show up how the members of a group are distributed on some attribute e.g. height, length of life, etc. A common frequency distribution is the normal distribution where the members are non identical and tend to cluster around the average.
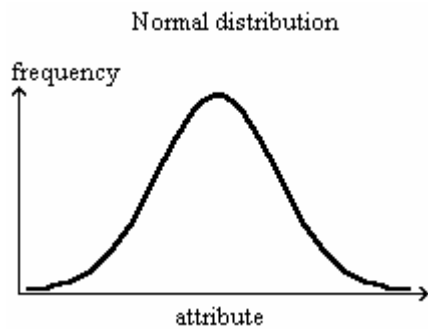
Normal distribution

figure 3

So when someone makes a statement that "Group X is better than group Y at some activity or on some attribute." you could think about the distribution of the skills or attributes of the members in these groups.

This automatically leads to non identification and possibly further inquiry like "Just how many members of one group do better than the best member of the other group?"

I see three main types of group comparison:

- Where you can get no information from the comparison.
- Where you can get some information from the comparison.
- Where you get full information from the comparison. (Uncertainty tacitly understood.)
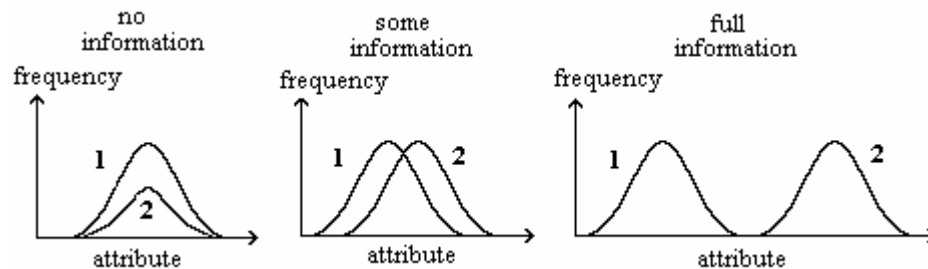
I will illustrate this graphically below.

figure 4

From the left, the first comparison might be between a whole group, say group$_1$ with another group, say group$_2$, which might be a good random sample of that whole group. Comparing these groups will give you no new information. For example finding a tall girl in a girl guides group might be just as likely as finding one in the general population as a whole.

The second graph shows the groups partially overlapping. Here you get some information but it is only partial and probabilistic in nature. For example finding a tall boy (over 6ft) at a school is more likely than finding a tall girl there.

Finally the groups 1 and 2 are separate. Here you get full[7] information about the members. E.g. To the question: "Are the women in the US basketball team taller than the professional male jockeys at the race track?" The answer is a "certain" yes.

The first and last comparisons we don't have too much trouble with. The middle comparison where there is a partial overlap creates lots of problems. Mathematics can help us here.

Math solution

Now that you have the graphs to visualise I will move onto the operational[3] solution which uses maths to help make the formulation more precise. When comparing groups, one operation you can do to compare them is to randomly pick a member from each group and compare just these two members on the attribute you are interested in. This is then repeated a significant number of times to overcome sampling bias. You can then look at the probability that a member randomly selected from one group is better than a member from the other group. Or to put it more mathematically, we look at the probability that A_i is greater than B_j for randomly selected members i and j. Once we have this probability we can then use information theory to tell us how much information the comparison will give us. This is measured in bits. 0 bits means no information and 1 bit means full information for the answer to the question "Is group A better than group B on some attribute?". If we compared a group with itself in this manner we would end up with a probability of 0.5 that any member randomly picked would be better than any other. This is the case where the two distributions totally overlap and have the same structure. Hence we get no information. So a probability of 0.5 is equivalent to 0 bits of information. If the two groups have no overlap at all then the probability is 1 (i.e. "certain") that the member from the higher group will score better on some attribute than the member of the lower group. This gives us full information so the information content is 1 bit.

I've left the mathematical details of this to the notes[6]. Just to clarify this I'll run through an example.

Say someone wants to employ some youths in a mathematical type job. One could ask "Do boys do better at maths than girls?". The following graph illustrates the results of a study[4].

Sex based differences in Arithmetic Reasoning
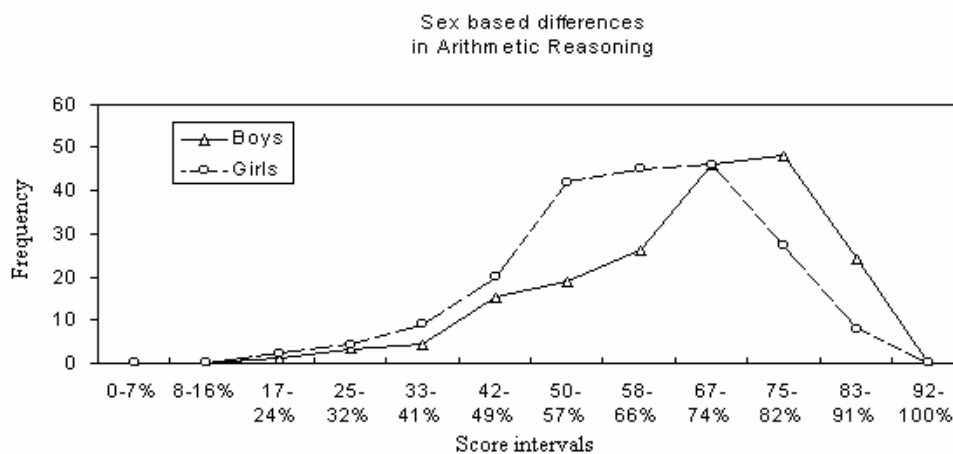
There is a large overlap between the groups.

figure 5

Do we get any information from this overlap. The answer is yes, in this case. If we apply the method of finding the probability that a randomly selected boy is better than a girl, we get a probability of 0.642. This is equivalent to 0.361 bits of information. Not certain, but we do get **some** information. The crucial question I see is: "Do we want to act just on this weak probabilistic information and hire only boys or can we do better?" I see acting on this information and choosing just boys for the maths job as sexist. If we want someone who scores at least 67% for the job, it disadvantages all those girls who scored this high. We can do better. Like Ed MacNeal[5] we can test the job applicants with applicable maths questions and choose only those who make the grade. The problem here is that people act on limited information when better information is available. Or they treat the probabilistic information as though it was certain.

So we can test out reality and find the factors that lead to the consequences we want, rather than just looking at groups and saying that one group has better consequences than another and choosing based that weak information alone. This is for the case where the groups overlap.

An example of this occurs with car insurance. Some firms charge based on groupings like age because these groups have different accident rates. E.g. Drivers aged 16-21 have more accidents than any other age group. But what factors lead to these accidents? Well I'd say that driver skill and attitude plays a large part. (Other factors like the amount of driving one does, when and where one drives will also have an impact, etc.) So the industry could test these and other relevant causal factors and charge according to them and not just base the charges on the easy to measure age group. This disadvantages all those safe 16-21 year old drivers.

English solutions

And finally one can change ones use of English to help solve this problem. Instead of asking "Is group A better than group B on some attribute" one uses a solution similar to what Korzybski recommended and one drops down an abstraction level. I.e. you talk about the members of the groups rather than the groups. In doing so you avoid identifying and you give more information about the group. E.g. "More women than men can have babies" replaces "Women can have more babies than men". There is actually an overlap in these distributions since some women can't have babies. So the general format is to move from: "$Group_1$ has more of attribute X than $group_2$" to "More $group_1$ members have attribute X than $group_2$ members." Another example: Instead of saying "In Australia during 2002, high school females were brighter than males", say "There were more females than male students in the top 100 Australian high school students in 2002." The first is false and the second correct. One can also talk about the members by using percentages. "Women live longer than men" can be replaced by "A larger percentage of females than males live till they are over 80 years old."

Another method of talking about the members is to compare them to some standard. E.g. "Year 7 & 8 students had 43% of girls as high achievers in reading and 25% of boys also made this grade." The grade of high achiever being the standard. So the inaccurate statement

of "Girls read better than boys" can be replaced by "More girls than boys can read very well." (Indexed and dated of course.) Again we compare members instead of groups.

To summarise. We can change our English to better compare groups by dropping down a level of abstraction and talking about the members. So use one of the following formats:

> More members of group A (*some comparison*) than group B members.
> A larger percentage of the members from one group (*some comparison*) than the members from the other group.
> Or compare both sets of members to some standard.

Conclusion
I hope I have described the problem more accurately and how we can improve our language so that it more accurately reflects reality. How we can cope when the groups do overlap, when we can't get full information but in some circumstances, we can get some information. We then need to realise the limits of this information and not over generalise it. We can then look for more useful information on which to base our decisions. E.g. test the territory. This will help us avoid acting in a racist, sexist or ageist manner, etc.

References/Notes:

People in Quandaries by Wendell Johnson

Science and Sanity by Alfred Korzybski

Operational Philosophy by Anatol Rapoport

Data on maths ability comparison quoted in "Psychology: A Dynamic Science" by Kurt Schlesinger and Phillip Groves p 295.

Score	Percentage	Boys	Girls
0-4	0-7%	0	0
5-9	8-16%	0	0
10-14	17-24%	1	2
15-19	25-32%	3	4
20-24	33-41%	4	9
25-29	42-49%	15	20
30-34	50-57%	19	42
35-39	58-66%	26	45
40-44	67-74%	46	46
45-49	75-82%	48	27
50-54	83-91%	24	8
55-59	92-100%	0	0
	Total	186	203

Mathsemantics: Making Numbers Talk Sense by Edward MacNeal
Method for group comparison and information theory. To compare two groups with a finite number of members do the following. For each member in one group derive the proportion of the members in the other group that are greater than it. E.g. if half the members are greater then the proportion is 0.5 These proportions are then all summed.
And finally the sum is normalised. I.e. divided by the number of members in group one. This gives a probability p(A>B) between 0 and 1. If using a continuous approximation, then you can use calculus to compute the probability.
You use a double integral that does the same as the above 3 steps.

Information theory.
To calculate the information content of a comparison I've developed a formula based on Shannon's theory which looks at the information content of a two option message as –
$p*\log_2(p) - q*\log_2(q)$
Where $p + q = 1$

7. I'm ignoring the uncertainty principle for the purposes of clarity and brevity. By "fully" I mean very close to certain. Or where the difference does not make a difference.