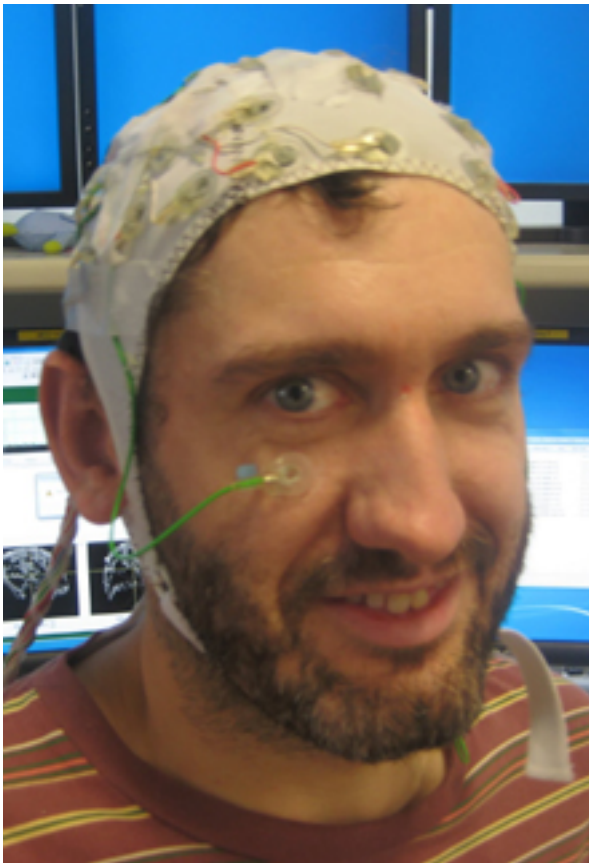


VIEWPOINT

Perfect match

BY JON BROCK

5 MARCH 2013





CONNECTIONS

People with autism have difficulty **recognizing facial expressions** of emotion. Their **eye movements are inaccurate**, and they have **difficulty miming**. They are, however, exceptionally **good at remembering melodies** and detecting pitch differences, although their brains are **slower to process sounds** compared with controls.

These are the results of just a few of the studies reviewed on SFARI.org in recent months.

All rely on comparisons between people who have autism and matched control groups of people without the disorder. The assumption is that statistically significant differences between the groups tell us something fundamental about the nature of autism.

Control groups are an essential part of autism research, providing a benchmark against which to assess those with autism. Finding, for instance, that participants with autism score an average of 68 percent on a test is meaningless if you don't know how people who don't have autism do on the same test.

A control group can also be used to try and rule out alternative and perhaps uninteresting explanations for group differences.

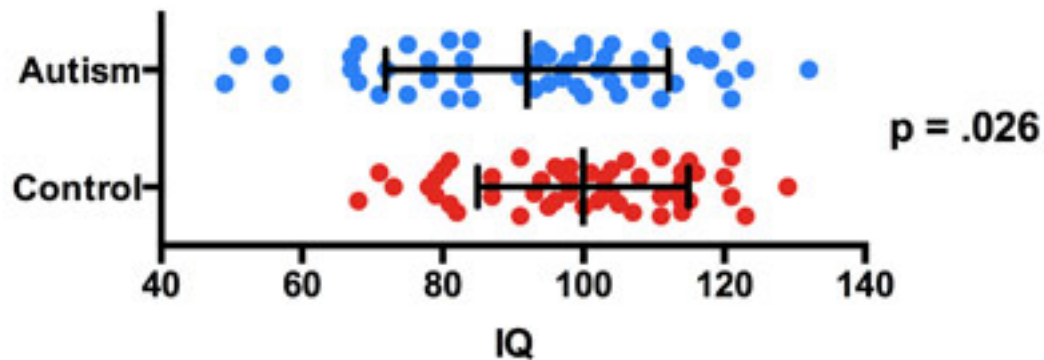
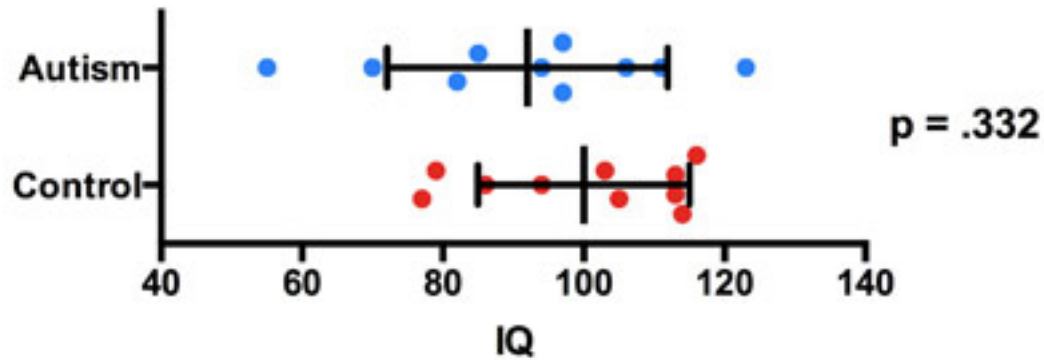
The logic is simple: If two groups are matched on one measure, such as intelligence or age, then this can't explain differences on another measure, such as performance on an emotion recognition test, that is under investigation.

Despite its widespread use, there are many issues to consider when **designing an experiment** with **matched controls** or when reading and attempting to evaluate such a study. Who should be in the control group? On what measures should they be matched? And how do we decide if the groups are truly matched?

In an article published in the January issue of the *American Journal of Intellectual and Developmental Disabilities*, **Sara Kover** and **Amy Atwood** consider **this last question**. Most researchers aim to convince the reader that their groups are well matched by reporting a statistical test, called a t-test, and showing that there is no significant difference between the two groups on the matching measure.

The problem is that t-tests really aren't designed for this purpose.

If a t-test shows a significant difference, then you can be fairly confident that the groups really are different. But the **reverse doesn't apply**: If the t-test doesn't find the groups to be significantly different, you can't necessarily conclude that they are similar.



Fact or fiction: These made-up studies illustrate the importance of reporting multiple measures of the differences between study groups.

To make matters worse, the result of the t-test depends on the number of people in the study. For small groups, it's much harder to get a significant difference — meaning that it's much easier to have groups that appear matched.

To illustrate the problem, Figure 1 shows some made-up data from two fictitious studies of different sizes. Both studies show the same distribution of intelligence quotients (IQs).

In the first study, which has ten participants in each group (represented by colored circles), the differences in IQ are not even close to being statistically significant. So we would consider the two groups to be well matched on IQ.

In the second, larger study, with 50 participants per group, the difference is highly significant. This time we'd say they are clearly not matched, even though the distribution of scores is exactly the same as in the first study.

Kover and Atwood recommend that when describing the matching procedure, researchers should always report the effect size. This statistic indicates how big a difference there is between the two groups and, crucially, does not depend on the size of the groups.

They also recommend reporting the ratio of the variance, which shows how similar the spread of scores is in the two groups.

In our two made-up studies, the effect size is similar (0.46 and 0.48) and the variance ratio is identical (1.78), indicating that the matching is equivalent in both studies.

However, there are currently no agreed-upon guidelines for matching, so researchers would have to get together and decide on cut-off values to define acceptable degrees of matching. For now, reporting the effect size and variance ratio would at least make it easier to judge how well the groups in a study are matched.

In fact, I've just added them to the paper I'm about to submit.

Jon Brock is a research fellow at Macquarie University in Sydney, Australia. He also blogs regularly on his website, [Cracking the Enigma](#). Read more Connections columns at SFARI.org/connections »