

## Why correlation is not causation, and why it matters.

It has often been reported on investigations that certain events or measurements seem to happen in conjunction with haunting type experiences or at locations where these experiences are said to occur. A change in EMF readings, temperature drops, seemingly whispered voices, knocks, bangs and a feeling of being touched are commonly reported. Additionally, it is also common to research the history related to the location. This becomes especially salient if it reveals tragedy or deaths during its habitation or occupation.

Let's leave aside any appraisal of whether these actually do correlate with haunt type experiences. Let us assume that it has been found that the factors mentioned do indeed correlate with locations and reports of paranormal activity. What is wrong then with the common following assumption, that because these factors are present when experiences occur, there is a causal relationship? That is, either such factors are causal in experiences or events (such as a tragic history or even human habitation itself) or paranormal activity causes reported factors (such as temperature drops, noises or sensing a presence).

I'll make the case that, even if certain factors do correlate with paranormal experience, it is an error to claim them as causal factors for it. I'll also say why it matters.

*...so what is correlation?*

With correlation data, you are essentially measuring the strength and direction of relationships that already exist between variables. These relationships can be strong or weak, negative or positive. We see or come across examples every day. For example gender and height (males tending to be taller than females), socioeconomic status and smoking (higher proportion of smokers in lower SES categories) or traffic and air pollution (higher levels of air pollution found in areas of higher volumes of traffic).

From these examples we can see not only the strength and direction of relationships between variables, but could also the make claims at to the causal relationships between variables. But is it that simple?

*...so what is causality?*

In experimental terms, we can infer causality when manipulating variable A has an effect on variable B. Well, if we want to be precise, what we actually do is reject the null hypothesis that manipulating variable A has no effect on variable B. That's a discussion for another time however. But in plain English, we can say variable A causes something to happen to variable B.

The issue arises when one is presented with correlation data, and then assumes a causal relationship based only on the presented data. The following example should illustrate the point;

Suppose you have a data set that shows a correlation between wet weather and the number of traffic accidents on a particular stretch of road. Using everyday logic, we might be confident in claiming that the wet weather caused more accidents. It seems reasonable, and based on our own experience of driving in wet weather, it indeed seems likely.

But what if we dig a little deeper behind the numbers and pose a few questions? Are there other factors that may indicate possible causes?

Consider the following regarding the collection of the data set;

- Time of day - Was it during a peak time when traffic was heaviest? Could the number of vehicles itself have contributed?
- Period of data collection – Was the data collected only on one day or over a number of days? If over a number days, did both wet and dry days contribute to the data?
- Particular day – If collected on one day, what was the context of that day? For example, suppose the data was collected on a Tuesday following a Bank Holiday Monday. Lots of people drink all day on a Bank Holiday. Could driver blood/alcohol levels the following morning have an effect on the number of accidents?
- Weather – How long had it been raining? Was it currently raining when the data collection took place? Had the rain stopped and it was now sunny? How long was the road dry before it became wet?
- Road conditions – What was the state of the road? Was there more oil than usual on the road? Was the road recently resurfaced or in need of repair?
- Effect of weather – Were the accidents due to road surface conditions affecting braking distances? Were the accidents due to impaired visibility for the drivers?
- Types of vehicles – Were more heavy goods vehicles present at the time of the accidents? Were more buses present? Did the accidents involve only cars or motorcycles or both?
- Types of accident – Were there different types of accidents? Were different categories of accident more prevalent than others? Were different factors responsible for the different categories of accident?

From these we can see that if we look deeper, the original correlation data may not necessarily imply causation. At the very least, we find good grounds on which to question any assumption that the wet weather, at least on its own, caused the accidents. There is a correlation certainly. But to claim a causal relationship, we would need specific experiments designed to not only test the effect of wet weather on likelihood of accidents, but also to control for the other factors listed.

After all, bear in mind that if the correlation data shows a positive relationship between wetter weather and number of traffic accidents, we can also interpret it as showing a positive relationship between number of traffic accidents and wetter weather. Would we be so quick to say that accidents cause it to rain?

*...In a paranormal context*

So in the context of paranormal investigations does it matter? Is “correlation is not causation” such an important issue?

Yes it is.

If we want to investigate claims of paranormal experiences or events, we need to account for what it is not before we can claim what it is. Many groups simply don't do this. Quite often we see groups claiming one thing must be caused by another, simply by virtue of them happening at the same time at the same location. This then is interpreted as being evidence for paranormal activity.

However, if we think critically and objectively about these situations we can see this is a highly flawed conclusion. Let's look at a common element of many groups “evidence” from investigations.

*...Orbs*

Many groups post photos of orbs captured by digital cameras. Many groups go so far as to claim these as evidence of the paranormal. Leaving aside the orbs as paranormal debate, let's look at this from the perspective of correlation vs. causation.

From the start, we can say with some confidence that people who post orb photos will have a prior belief not only in the reality of the paranormal, but also that orbs are manifestations of it. There's a correlation right there. Would those who support the notion that orbs are paranormal manifestations say that their belief causes the orbs to manifest, rather than the spirit of a deceased human? It's actually an interesting thought, but I've personally not seen it proposed.

Of course, believers in orbs can come back and say that, just because there was dust or moisture in the air when the orbs were captured, it doesn't mean this caused them to manifest or be captured. And that's a valid point.

And herein lays the difference between correlation and causation...and why it matters.

The fact that dust, moisture or insects are present when orbs are captured does not inherently mean that they are the cause. Similarly, the fact that orbs were captured increasingly frequently with the increased use of digital cameras does not mean this photographic process was the cause. To make

any kind of claim for this, you would need to conduct a specifically designed experiment to test these hypotheses.

...and so some researchers did.

I'm not going to discuss these experiments or the findings here. You can read about them yourself by following the enclosed links;

<http://www.assap.ac.uk/newsite/articles/Orb%20Zone%20Theory.html>

<http://www.parascience.org.uk/articles/orbkill.htm>

But in essence what the experiments found was evidence supporting the theory that orbs were an artefact of the photographic process in conjunction with the presence of particles in the air.

So we have a correlation from observation, and causation from experiments.

...savvy?

What about from the perspective that orbs are paranormal manifestations? Not so much.

We can see that there may be a number of elements that correlate with this. Orbs being captured in locations with reported hauntings. Orbs captured during drops in ambient or even localized temperature drops. Orbs captured at the same time as people report feeling being touched or sensing a presence.

Is there any experimental evidence to suggest these correlations are causal factors in orbs being captured, and therefore suggestive of a paranormal origin? To the best of my knowledge there is none. We have self-reported and anecdotal evidence, but if we want to make claims of causation we need to go beyond this...quite far beyond it.

*...necessary, but not sufficient*

We can (and should) apply the correlation vs. causation argument to many factors reported in relation to allegedly paranormal experiences or locations. Examples include a history of tragedy or death related to the location, cold spots, capturing orbs in photos, responses to questions during ghost box sessions, feelings of being touched and a host of other reported experiences. These are commonly reported to occur during investigations, and are often used to bolster the claim of a location's haunting or that an investigator's experience had a paranormal source. However, as with the traffic accident example earlier, if we dig a little deeper we find the picture is a little more complicated.

Looking at factors that may correlate with reported anomalous experiences or events is undoubtedly useful and necessary. It can provide associated avenues of inquiry and generate new questions. We just need to be aware of the strengths and limitations of correlation data.

If we are going to make claims regarding paranormal causes for experiences or events, or at least propose that prosaic ones won't suffice, these need to be supported with transparent and verifiable data, preferably from experiments designed to test specific or related hypotheses.