Experiment: Scientific inquiry is all about asking what would happen when a unit (like a person) is exposed to a treatment (like a drug). Ideally, you want two units that are exactly the same, one of which you give the treatment to, and one of which you give nothing to, and then you measure the difference. That difference can be said to be *caused* by the treatment.

Unfortunately, nothing in the social world has an exact copy; no two humans are the same, no two political systems are the same, and certainly no two countries are the same. One way around this problem is to look at a large number of cases and to *randomly assign* cases two either a treatment group or a control group. If we properly randomize these groups, we shouldn't expect there to be any systematic differences between them *on average*. Thus, when you treat one group and leave the other as a control group, the differences you measure, on average, can be said to be caused by your treatment. Again, this is because the two groups are the same *in expectation*. Of course we know that the two groups won't be exactly equal, but the larger the groups you have, the more likely the differences will average out.

Because countries are so unique, and because there are so few of them (relative to other social units, like people), it becomes more difficult to apply the above scientific method. In the real world you can never know the *counterfactual* (i.e. what would have happened if Egypt was a major oil exporter?), and when your sample is small and your units are extremely *heterogeneous*, it's even harder to know. Still, in all scientific endeavors, in order to try to gain knowledge, some manifestation of the above method should be applied. This is what led Rustow to claim, "Any country study nevertheless sacrifices the advantages of comparison, the social scientist's nearest substitute for a laboratory." Rustow wants to avoid the pitfalls of only studying one case, because the counterfactual essentially becomes impossible to know.

When you write your paper, you should try to think in these terms when analyzing any causal claims that Brownlee et al makes. But the degree to which a research project mimics a laboratory experiment is not the only concern one should have when analyzing a causal claim (though it is extremely important). The following concepts can and should also guide you when making your judgments.

Necessity and Sufficiency: A necessary condition is one which is required for a claim to be realized. For example, Rustow claims that the only necessary condition for a transition to democracy is that the people in a polity have a sense of national unity. A sufficient condition, on the other hand, is one which mandates an outcome. Sufficient claims are more difficult to make because they *always* claim that an outcome will happen. An example would be something like McFaul's claims about elite preferences for democratic change. For McFaul, having a power *imbalance* is a necessary condition for democratic regime change, and having the winners of that imbalance embrace democratic principles is a sufficient condition (all other things being equal). When you are assessing Brownlee et al, determine whether the arguments are necessary, sufficient, both, or neither for regime stability.

Reverse Causation: Reverse causation means that a stated causal relationship actually goes in the wrong direction. An example may be something like gun laws and illegal firearms use. One might think that as illegal firearms use increases, you will see an increase in gun laws. At the same time, however, it's easy to recognize that as more activities become illegal, more behaviors are defined as crimes, whether or not the behavior changes at all. In other words, making more things illegal *causes* them to become crimes.

Omitted/Confounding Variable (Bias): Social scientists like to use terms like Independent Variable and Dependent Variable. You can just consider the definitions of the words Independent and Dependent to figure out what these mean. The Independent Variable (IV) is the one that doesn't (or shouldn't) change. In experimental terminology, this is simply the treatment. If you are running a drug trial, it makes sense to give everyone the same drug if you want to test that drug against a control group. The drug is Independent because it is (or should be) unaffected by anything else. In other words, it is administered equally and in an unbiased way. The Dependent Variable (DV), on the other hand, does change. This is exactly the change you want to measure. In experimental terminology, this would just be the difference in the effect between the experimental and the control group. This outcome *depends* on whether or not someone received the treatment. Additionally, if you didn't get variance in this outcome, you would have measured nothing.

So in an experimental study, you would have a model that looks something like this:

IV → DV

This is simple and straightforward. Receiving your treatment causes a change in your measurement of interest. Another way to think of this is the following:

Headache Medicine → Reduced pain

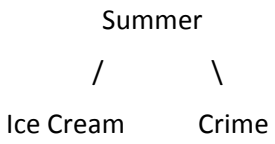
Simple, elegant, publishable, great work! This is easy to claim in a randomized control experiment in a lab. This is much less easy to claim in the social world. Let's try to take a real world example. There is a strong correlation between ice cream sales and incidence of crime. You might think there's a relationship here. Maybe eating too much ice cream makes people crazy:

Ice cream → Crime

Or maybe once people are criminalized they seek solace in a nice bowl of ice cream:

Crime → Ice Cream

While there may be some truth to these claims (probably not), there's a much more likely explanation for this apparent relationship; some other unmeasured variable is causing both crime rates and ice cream sales. It just so happens that more crime happens in the summer and more ice cream is purchased in the summer. Instead of there being a relationship between ice cream and crime, each of those things is actually related to summer, and not directly to each other:



This is the omitted or <u>*confounding*</u> variable bias. This happens when you *think* you see a relationship but it doesn't exist in reality. The relationship is *spurious*. This is exactly why the phrase "<u>*correlation is not causation*</u>" exists. When you analyze Brownlee et al, you definitely want to determine whether what they show is a correlational or a causal relationship.

To illustrate this point even further, consider this obviously ridiculous example:

When reading Wood, I noticed a relationship between the two countries she studied; both El Salvador and South Africa have two words in their names. Based on this similarity, I claim that all countries with two words in their names will lead to democracy. When I look at the real world, I also notice that The United States of America has *all kinds* of words in it, so I further conclude that more words leads to

more democracy. Here we clearly see that the relationship we observe is completely spurious, and is likely an issue that would wash away (average out) in a much larger sample size.

Tautology: Another potential problem with a causal argument is that it might be *tautological*, or in other words, the IV is the same thing as the DV. For example, if one of the authors is suggesting that having a strong regime leads to stability, is that really a relationship, or is that just a definition? Tautological arguments do not demonstrate a relationship because, as stated, the DV is the same as the IV. This doesn't mean the authors are incorrect; in point of fact all tautologies are axiomatically correct. It just means that they haven't said anything new.

Statistical Significance: Statistical significance is a somewhat amorphous term given that there are various levels of significance, and all of them are arbitrarily chosen. Social science likes to use the probability .05 level of significance as the minimum acceptable, but this is not a hard and fast rule. What this means is that social scientists tend to believe results if they are only likely to happen by chance 1 out of 20 times. This is a somewhat more complicated argument to make since it involves an understanding of statistics that you might not yet have acquired, but if this is a skill set you have you can bring it to bear. Relating back to the experimental language, the more units (i.e. cases) you have to measure, the more likely the differences between those cases average out. Without this kind of confidence, researchers must do a much more in depth level of *qualitative* analysis to make their point. Because country studies are quite often limited to very few cases, even when they do obtain statistical significance it's often hard to believe, and we demand a much more rigorous study of the processes, systems, institutions, actors etc. before we accept the results. If a study is missing both numbers and an in depth analysis, you should worry about their findings.

Internal Validity: Internal validity asks the question of whether or not a research model is actually measuring what it intends to measure. I could claim that I am going to study the effects of high partisanship on regime stability, but if my measure of partisanship is how strongly people prefer chocolate ice cream over vanilla (i.e. ice cream partisans), my study would lack internal validity. When analyzing Brownlee et al, you certainly want to make sure that they are measuring what they claim to measure.

External Validity: On the other hand, you also want to get a sense of the external validity of a research project. External validity concerns the question of whether and how well a project's findings generalize. Some studies don't claim that their findings generalize at all. Several of the authors this week suggest that their finding might only apply to a particular set of institutions that exist in a particular time in history and that we would be unlikely to find again. You might think of McFaul in this way. McFaul is studying a specific group of once-communist countries that slowly collapsed as their primary benefactor/coercive hegemon also collapsed (Russia). Do we expect to see what McFaul has shown at any point in the future? Probably not. When reading Brownlee et al, you will want to consider whether the claims they are making are generalizable to other countries, and if so, are these claims believable.

Face Validity: Finally, there's the more informal idea of face validity. Face validity simply asks the question "Does this argument make any sense just looking at the face of it?" The ice cream/crime relationship severely lacked face validity.