

Experimental Approaches to the Study of Personality

**William Revelle
Northwestern University**

To appear in: Personality Research Methods by Robins, Fraley and Krueger

Experimental Approaches to the Study of Personality

Personality is an abstraction used to explain consistency and coherency in an individual's pattern of affects, cognitions, desires and behaviors. What one feels, thinks, wants and does changes from moment to moment and from situation to situation but shows a patterning across situations and over time that may be used to recognize, describe and even to understand a person. The task of the personality researcher is to identify the consistencies and differences within and between individuals (what one feels, thinks, wants and does) and eventually to try to explain them in terms of set of testable hypotheses (why one feels, thinks, wants and does).

Personality research is the last refuge of the generalist in psychology: it requires a familiarity with the mathematics of personality measurement, an understanding of genetic mechanisms and physiological systems as they interact with environmental influences to lead to development over the life span, an appreciation of how to measure and manipulate affect and cognitive states, and an ability to integrate all of this into a coherent description of normal and abnormal behavior across situations and across time.

Although the study of personality is normally associated with correlational techniques relating responses or observations in one situation or at one time with responses at in other situations and other times, it is also possible to examine causal relations through the use of experimental methods. This chapter will outline some of the challenges facing personality researchers and suggest that an experimental approach can be combined with more traditional observational techniques to tease out the causal structure of personality.

Central to our analysis is the distinction between personality traits and personality states. States can be thought of as the current values of one's affects, behaviors, cognitions and desires while traits have been conceptualized as either average values of these states or alternatively the rates of change in these states. In more cognitive terms, traits are measures of chronic accessibility or activation, and states are levels of current activation. It is perhaps useful here to think analogically and to equate states with today's weather and traits as long terms characteristics of weather, that is to say, climate. On any particular day, the weather in a particular location can be hot or cold, rainy or dry. But to describe the climate for that location is more complicated, for it includes among other aspects a description of the seasonal variation in temperature and the long term likelihood of draught or blizzards or hurricanes. Extending this analogy, climatologists explain differences in climate between locations in terms of variations in solar flux and proximity to large bodies of water, and changes in climate in terms of long term trends in greenhouse gases in the atmosphere. The role of the personality researcher is analogous to the meteorologist and climatologist, trying to predict someone's immediate states as well as understanding and explaining long term trends in feelings, thoughts and actions.

Integrating Two Alternative Research Approaches

Psychological research has traditionally been described in terms of two contrasting approaches: the correlational versus the experimental (viz., the influential papers by

Cronbach, 1957, 1975; and Eysenck, 1966, 1997). Francis Galton and his associate Karl Pearson introduced the correlation as a way to study how individual differences on one variable (e.g., the height of one's parents or one's occupation) could be related to individual differences in another variable (e.g., one's own height or to one's reaction time). Correlational approaches have been used in personality research since Galton to predict school achievement or job performance, and when combined with known family structures (e.g., parents and their offspring, monozygotic or dizygotic twins with each other, adopted and biological siblings) and structural equation models have allowed for an examination of the genetic basis of personality. Applying structural techniques such as factor analysis to covariance or correlation matrices of self and other descriptions has led to taxonomic solutions such as the Giant Three or Big Five trait dimensions. The emphasis in correlational research is on variability, correlation, and individual differences. Central tendencies are not important; variances and covariances are. The primary use of correlational research is in describing how people differ and how these differences relate to other differences. Unfortunately for theoretical inference, that two variables are correlated does not allow one to infer causality. (e.g., that foot size and verbal skill are highly correlated among preteens does not imply that large feet lead to better verbal skills, for a third variable, age, is causally related to both.)

A seemingly very different approach to research meant to tease out causality is the use of experimental manipulation. The psychological experiment, introduced by Wundt and then used by his students and intellectual descendants allows one to examine how an experimental manipulation (an Independent Variable) affects some psychological observation (the Dependent Variable) which is, in turn, thought to represent a psychological construct of interest. The emphasis is upon central tendencies, not variation, and indeed, variability not associated with an experimental manipulation is seen as a source of noise or error that needs to be controlled. Differences of means resulting from different experimental conditions are thought to reflect the direct causal effects of the IV upon the DV. Threats to the validity of an experiment may be due to confounding the experimental manipulation with multiple variables or poor definition of the dependent variables or an incorrect association between observation and construct.

One reason that correlational and experimental approaches are seen as so different is that they have traditionally employed different methods of statistical analysis. The standard individual differences/correlational study reports either a regression weight or a correlation coefficient. Regression weights are measures of how much does variable Y change as a function of a unit change in variable X. Correlations are regressions based upon standard scores, or alternatively the geometric mean of two regression slopes (X upon Y and Y upon X). A correlation is an index of how many *standardized* units does Y change for a *standardized* unit of X. (By converting the raw Y scores into standardized scores, $z_y = (Y - \bar{Y})/s.d._y$, one removes mean level as well as the units of measurement of Y.) Experimental results, on the other hand, are reported as the differences of the means of two or more groups, with respect to the amount of error within each group. Student's *t-test* and Fisher's *F test* are the classic way of reporting experimental results. Both *t* and *F* are also unit free, in that they are functions of the effect size (differences in means expressed in units of the within cell standard deviation) and the number of participants.

But it is easy to show that the t-test is a simple function of a correlation coefficient where one of the variables is dichotomous. Similarly, the F statistic of an analysis of variance is directly related to the correlation between the group means and a set of contrast coefficients. The recognition that correlations, regressions, t and F statistics are all special cases of the general linear model has allowed researchers to focus on the validity of the inferences drawn from the data, rather than on the seeming differences of experimental versus correlational statistics.

The use of meta-analysis to combine results from different studies has forced researchers to think about the size of their effects rather than the significance of the effects. Indeed, realizing that $r = \sqrt{F/(F+df)}$ or $\sqrt{t^2/(t^2+df)}$ did much to stop the complaint that personality coefficients of .3 were very small and accounted for less than 10% of the variance to be explained (Ozer, 2006). For suddenly, highly significant F statistics were found to be showing that only a small fraction of the variance of the dependent variable was accounted for by the experimental manipulation.

The realization that although the statistics seemed different but are actually just transformations of each other forces experimentalists and correlationalists to focus on the inferences they can make from their data, rather the way in which the data are analyzed. The problems are what kind of inferences one can draw from a particular design, not whether correlations or experiments are the better way of studying the problem.

Latent constructs, observed variables and the problems of inference

Fundamental to the problem of inference is the distinction between the variables we measure and observe and the constructs that we think about. This distinction between latent (unobserved) constructs and measured (observed) variables has been with us at least since Plato's Phaedra. Consider prisoners shackled in a cave and only able to see shadows (observed scores) on the cave wall of others (latent scores) walking past a fire. The prisoners attempt to make inferences about reality based upon what they can observe from the length and shape of the shadows. Individual differences in shadow length will correctly order individual differences in height, although real height can not be determined. To make this more complicated, as people approach the fire, their shadow lengths (the observed scores) will increase, even though their size (the latent score) has not changed. So it is for personality research. We are constrained to make inferences about latent variables based upon what we measure of observed variables. The problem may be shown diagrammatically (Figure 1) where boxes represent observed variables, circles latent constructs, and triangles experimental manipulations. From the observed pattern of correlations or t-tests we attempt to make inferences about the relationships between the latent variables as well as between the latent and observed variables.

Insert Figure 1 about here

There are at least three challenges that we face when making inferences about relationships between latent variables: the *shape* of the functional relationship between

observed and latent variables, the *strength* of the functional relationship between observed and latent variables, and the *proper identification of the latent variables* associated with observed variables and manipulations.

Consider the following two hypothetical studies. Both are field studies of the effect of education upon student outcomes. In study 1, students from a very selective university, a less selective university, and a junior college are given a pretest exam on their writing ability and then given a post test exam at the end of the first year. The same number of students are studied in each group and all students completed both the pretest and post test. Although there were differences on the pretest between the three student samples, the post differences were even larger (Figure 2a). Examining figure 2a, many who see these results conclude that students at the highly selective university learn more than students at the less selective university who change more than the students at the junior college. Some (particularly faculty members) like to conclude that the high tuition and faculty salaries at the prestigious and selective university lead to this greater gain. Others believe that the teaching methods at the more selective university are responsible for the gains, and if used at the other institutions, would also lead to better outcomes. Yet others (particularly students) point out that the students in the prestigious university were probably smarter and thus more able to learn than the students in the junior college.

Hypothetical study 2 was similar to study 1, in that it was done at the same three institutions during the first year, but this time the improvement on mathematics achievement was examined (Figure 2b). Here we see that students at the most selective school, although starting with very high scores, did not improve nearly as much as the students at the less selective university, who improved even less than the students at the junior college. Most faculty and students who see these results immediately point out that the changes for the selective university students were limited by a “ceiling effect” and that one should not conclude that the selective university faculty used less effective techniques nor that the students there were less able to learn.

The results and interpretations from these two hypothetical studies are interesting for in fact one is the inverse of the other. Scores in study 2 are merely the scores in study 1 subtracted from 100. The results from both study 1 and 2 can be seen as representing equal changes on an underlying latent score, but using tests that differ in their difficulty. Study 1 used a difficult test in which improvements of the students at the less selective institution were masked, study 2 used an easy test where improvements of students at the more selective institution were masked (Figure 2c). That differences in outcome are explained by ability in study 1 but scaling effects (in this case, a ceiling effect) in study 2 exemplifies the need to examine one’s inferences carefully and to avoid a confirmation bias of accepting effects that confirm one’s beliefs and searching for methodological artifacts when facing results that are disconfirming.

Insert Figure 2 about here

We will revisit this problem of scaling effects upon inferring differential effects of personality and situational manipulations when we consider the appropriate interpretation of interactions of personality and situations.

A second problem in inferring differences in latent scores based upon changes in observed score is the strength of the relationship between latent and observed. This is the problem of reliability of measurement. Although addressed more completely in chapter XX [note to editor- this is an assumption], the basic notion of reliability is that any particular observed score reflects some unknown fraction of the latent score as well a (typically much larger) fraction of random error. By aggregating observations across similar items or situations the proportion of the observed score due to the latent score will increase asymptotically towards 1 as a function of the number of items being used and the similarity of the items. Assuming that items are made up of a single latent score and random error, the proportion of latent score variance in a test with k items and an average inter item correlation of r is $\alpha = k*r/(1+(k-1)*r)$. Furthermore, while r is the average correlation between any two items and is equal to the ratio of latent score variance in an item to total item variance, α is equal to the percentage of total test variance that is due to latent score variance. More generally, the reliability of a measure of individual differences is a function of what we are trying to generalize across (e.g., items, people, raters, situations, etc.) and the structure of our measures (Zinbarg, et. al, 2005).

Confirmatory versus disconfirmatory designs

Although it is very tempting (and unfortunately extremely common) to test hypothesis by looking for evidence that is consistent with the hypothesis (e.g., “testing” the hypothesis “all swans are white” by looking for white swans), in fact disconfirming evidence is the only test of a hypothesis (even after seeing 1,000 white swans, seeing 1 black swan disconfirms the hypothesis.) The use of strong inference (Platt, 1964), to ask what hypothesis a finding can disconfirm, should be the goal of all studies. For science is the process of refining theories by excluding alternative hypotheses.

“I will mention one severe but useful private test – a touchstone of strong inference - that removes the necessity for third-person criticism, because it is a test that anyone can learn to carry with him for use as needed. It is our old friend the Baconian “exclusion,” but I call it “The Question.” Obviously it should be applied as much to one’s own thinking as to others’. It consists of asking in your own mind, on hearing any scientific explanation or theory put forward, “But sir, what experiment could disprove your hypothesis?”; or, on hearing a scientific experiment described, “But sir, what hypothesis does your experiment disprove?” Platt, 1964

Consider the following sequence of numbers that have been generated according to a certain rule: 2, 4, 8, X, Y, ... What is that rule? How do you know that is the rule? One can test the hypothesized rule by generating an X and then a Y and seeing if they fit the rule. Many people, when seeing this sequence will believe that the rule is successive powers of 2 and propose X=16 and then Y= 32. In both cases they would be told that

these numbers fit the rule generating the sequence. These people will think that they have confirmed their hypothesis. A few people will infer that the rule is actually increasing even numbers and test the rule by proposing $X = 10$ and then $Y = 12$. When told these numbers fit the rule, they will conclude they have confirmed their hypothesis. A few will suspect the rule is merely an increasing series of numbers (which is in fact the rule used to generate the numbers) and try $X = 9$ and $Y = 10.92$, and conclude that they have discovered the rule. Even fewer will propose that the rule is any series numbers and try to test the rule by proposing $X = 7$ or that $Y = \sqrt{43}$. These terms do not fit the rule and allow us to reject the hypothesis that any number will work. This simple example shows the need to consider many alternative hypotheses and to narrow the range of possible hypothesis by disconfirmation. For, as that great (but imaginary) scientist Sherlock Holmes reasoned “when you have eliminated the impossible, whatever remains, however improbable, must be the truth” (Doyle, 1929: *The study in scarlet*, chapter 6.)

Although it is nice to think of theories as mutually incompatible, with evidence for one disconfirming a second, in fact most theoretical descriptions of personality make predictions for only a limited range of phenomena and are silent about others. In this case, it is helpful to make a table with phenomena as rows, theories as columns, and entries as +, 0, -, or blank. Although many cells of this table will be empty, and some rows will all make the same prediction, there will be some rows that show real differences between the theories. These are the phenomena to test (e.g., Anderson and Revelle, 1982; Leon and Revelle, 1985; Zinbarg and Revelle, 1989).

Experimental manipulations as tests of theories of causality

In the mid 1500's, a revolutionary technique was added to the armamentarium of scientific reasoning. Rather than using arguments based upon assumptions and logical reasoning, the process of empirical observation and more importantly, experimental manipulation was introduced (see Shadish, Cook and Campbell, 2002, for a wonderful discussion of the development of experimentation and reasoning). By observing the results of experimental manipulations it became possible to tease apart alternative hypotheses and to address issues of causality. Although statistically, there is little to differentiate experimental and correlational data, the importance of experimental techniques is the ability to make statements about causality and to exclude possible explanations by experimental control. When applied to personality theory, experiments allow us to test the range of generalization of the relationships between individual differences and outcome variables of interest.

In addition to testing the range of generalization, experimental techniques allow for tests of causal hypotheses. That is, if we believe that X causes Y and that if and only if X occurs will Y occur, we can test this by doing both X and not X and seeing when Y occurs. To show that X leads to Y, it is not enough to merely observe that X and Y occur together, we also need to know what happens when we do not have X.

Consider the following table of possible outcomes for X and Y and consider which observations are necessary to test the hypothesis that X implies Y:

Hypothesis: $X \rightarrow Y$ (read, X implies Y)

Two states of X	Two states of Y	
	Y	Not Y
X	a	b
Not X	c	d

Four possible observations:

- a) X and Y
- b) X but not Y
- c) Not X, Y
- d) For any not Y is there is not X

Observing conditions a and c allow us to test the hypothesis, condition a is consistent with the hypothesis and c allows us to reject the hypothesis. Observing conditions b or d do not allow us to test the hypothesis.

Perhaps the easiest example of the problems of inferring and testing causality to consider is the relationship between sexual intercourse and pregnancy: If some one is pregnant, she has had intercourse (outcome a); if someone has not had intercourse, she will not become pregnant (outcome d). However, not being pregnant does not imply not having had intercourse (column 2) , nor does having intercourse necessarily imply pregnancy (row 1). Although intercourse is causally related to pregnancy, it is not sufficient. (It is interesting to note that even with such a well known causal relationship of sex with pregnancy, that with the reasonable assumption of frequency of intercourse twice a week for 20 years resulting in two children, the correlation coefficient (ϕ) is .026, and goes up to only .059 for the case of 10 children over 20 years.)

The fundamental requirement of an experiment is that there are at least two levels of a manipulated variable (the independent variable or IV). With the additional requirement that that assignment to those two (or more) levels is independent of any possible prior conditions, we have a true experiment. Observations on some outcome measure of interest (the dependent variable or DV) are then related to the states of the IV to detect if variation in the IV caused changes in the DV.

Subject variables or person variables (PV) reflect prior and stable characteristics of the individual and are not subject to manipulation. That is, one can not manipulate the sex, age, intelligence or extraversion of a participant. Although person variables are seen as nuisance variables to many cognitive and social psychologists, to personality researchers, person variables are of greatest import. Although excellent personality research can be done using correlations rather than experimental manipulation, with the use of experimental techniques, we are able to test the range of generality of our person variables.

It should be noted that experimental techniques do more than just limit the extent of inferences about personality. The use of personality in experimental designs allows one to achieve a greater range of the underlying constructs (e.g., arousal, fear, positive or negative affect) than would be achievable by simple manipulations. That caffeine increases arousal is well known, but the range of arousal can be increased by choosing subjects known to have high or low arousal in certain situations (evening people in the morning and morning people in the evening will have very low arousal, morning people in the morning and evening people in the evening will have very high arousal). Similarly, when studying mood effects upon memory, by choosing very depressed versus non-depressed participants, the range of negative affective state is greatly enhanced.

The generic experimental personality study can be thought of as examining how one or more stable personality traits (the PVs) combine (either additively or interactively) with one or more experimental manipulations (the EVs) to affect some hypothetical (but unobserved) states, the effects of which are then observed with on at least one measure (the OV). In some designs, measures of the intervening state are also taken and used either as manipulation checks or as part of the statistical model (Figure 1).

Experiments allow one to test the range of generality of a particular personality variable. In the most basic design of a single PV and a single EV, if the EV has an effect but does not interact with the PV, then we are able to extend the generality of the PV across at least the conditions of the EV. If the EV does interact with the PV, then the range of generalization of the PV is reduced. In both cases, we know more about the range of the PV-OV relationship. In the more complex case of multiple PVs or multiple EVs, interactions between the PVs or higher order interactions with several EVs constrain the limits of generalization even more.

A correlational study examines the relationship between a (presumably stable) person variable (PV) and some observed outcome variable of interest (OV). For instance, the hypothesis that trait extraversion related to positive mood is supported by a positive correlation of .4 between E as measured by scales from the International Personality Item Pool (IPIP, Goldberg, 2005) and positive mood as measured by items such as happy and cheerful. What is unclear from such an observed relationship is whether some people are chronically happy and cheerful and that this leads to the outgoing and energetic behavior of extraversion, or whether the behavioral approach exhibited by the more chronically extraverted results in higher positive mood.

With the introduction of an experimental manipulation of mood, where we find that showing a comedy increases positive affect more than a control movie, and that in both movie conditions, that extraversion is still related to positive affect, allows us to simultaneously increase the range of generalization of the E-PA relationship.

The range of potential person variables and potential manipulations is limited by one's imagination, but it is possible to list a number of the more common personality variables that have been examined in an experimental context (Table 1). The variables shown in Table 1 reflect the influence of two major proponents of experimental approaches to

personality research, J.W. Atkinson and H.J Eysenck. Both of these pioneers in personality research emphasized the importance of integrating studies of individual differences with experimental procedures. Their models also had the strength that with proper experimental design, hypotheses could be specified with enough detail that they could be rejected (e.g., Anderson and Revelle, 1992; Leon and Revelle, 1985; Zinbarg and Revelle, 1989).

Arousal based models of the Introversion/Extraversion dimension (Eysenck, 1967) made two strong theoretical statements: Introverts were chronically more aroused than were extraverts and arousal was curvilinearly related (with an inverted U shaped function) to cognitive performance. Thus, typical tests of the model would involve manipulations of arousal by giving stimulant or depressant drugs (e.g, amphetamine, caffeine, barbiturates), putting participants in arousing situations (e.g., under time pressure, noise, large groups) or varying the time of day. Confirmatory tests of this hypothesis showing interactive effects on complex cognitive performance of caffeine and introversion-extraversion (Revelle, Amaral, and Turrif, 1976) were followed up with studies that more precisely limited the effects by showing that these earlier results also interacted with time of day and were limited to a subcomponent of I-E, impulsivity (Revelle, Humphreys, Simon and Gilliland, 1980).

More recent conceptualizations of Introversion/Extraversion have focused on the relationship of Extraversion with positive affect and have examined the effects of positive mood inducing stimuli (e.g., movies, pictures, autobiographical memories) on subsequent mood (Larson and Kettalar, 199x) , performance (Rogers and Revelle, 1998) and psychophysiological (Canli et al, 2002) measures.

Early tests of theories of achievement motivation focused on the effect of perceived task difficulty and success and failure feedback upon task choice (Hamilton, 197x) persistence following failure, and changes in effort over time (Kuhl and Blankenship, 1974). More recent work has emphasized interactions between achievement goals and task feedback (Elliot, 200x).

Studies of Neuroticism and Anxiety have focused on ways of manipulating negative affect and inducing state anxiety. Manipulations similar to those used for inducing positive affect have been used for inducing negative affect and fear (e.g. sad or frightening movies, pictures of feared objects such as snakes or spiders) (Öhman and Mineka, 2002).

Avoiding confounds through experimental control, randomization and counterbalancing

Many seemingly different designs (EV with two levels, one EV with multiple levels, two EVs, etc.) all have similar requirements, the need to assign subjects to Experimental conditions with no relationship to other existing condition. That is, the only expected

variation between participants in the different conditions should be due to those conditions and not some other, confounded variable.

Perhaps the clearest way of thinking of the problem is consider a hypothetical data matrix where the rows are the participants, the columns include the Person Variables, Experimental Variables, and Observed Variables of interest, as well as other, extraneous Person and context variables (CVs). The inferential problem for the researcher is to know that the observed relationship between the PV, EV and OV is not due to any other source of variance. That is, that the effect is not due to the extraneous PV or CVs. These extraneous variables, if correlated with either the PV or the EV are confounded with the variables of interest and invalidate any causal inferences we try to make.

There is, of course, an infinite number of such possible confounding variables. Confounding person variables include individual differences in intelligence, SES, broad personality trait dimensions such as the Big 5, narrow trait dimensions such as impulsivity or anxiety or achievement motivation, or prior experience with the task. Confounding context variables range from the obvious effects of time of day, time of week, time of year, to effects of experimenter characteristics including gender, formality of clothing, friendliness, ethnicity, and age, as well as possible participant by experimenter interactions, of which among college students important ones to consider are participant gender and ethnicity by experimenter gender and ethnicity.

Quasi-experimental designs typically associated with field studies are rife with these potential confounds. Indeed, issues of differential selection, attrition, age, and experience are the topics of entire texts on quasi-experiments (Shadish et al., 2002). Our challenge as researchers is to eliminate the effect of these potential confounds. Unfortunately, we can not control for the effect of extraneous PVs by experimental design, but rather need to worry about them when trying to make inferences about the specific PVs of interest. We can, however, eliminate the effect of CVs by ensuring that their correlations with the EVs are 0.

It is possible to control explicitly for some confounding variables by making them part of the design. Thus, it is possible avoid time of day and experimenter characteristics as sources of variation by having the only experimenter run all participants at the same time of day. While avoiding confounds with time of day and experimenter characteristics, and reducing the amount of variation with experimental conditions, this procedure also reduces the generalizability of the findings to that particular combination of time of day and experimenter. Explicit control for confounds by restricting the experimental conditions thus increases the power of the design at the cost of reducing the generalization of the findings.

A statistically less powerful but much more generalizable control for confounds is to use random assignment. The technique of random assignment of participants to conditions will, on average, yield no correlation between the experimental conditions and extraneous, confounding variables. It is important to note that although randomization has an expected value of no relationship it does not guarantee no relationship in any

particular study. (I once found a significant interaction on cognitive performance between impulsivity and caffeine on a pretest, before the caffeine was administered. Either we had confirmed precognition, or had observed a failure of randomization to equate the groups.)

Random assignment of participants to conditions is easier to say than to do, for there are problems that will arise with randomization. For a given number of participants, statistical analysis will have the greatest power when an equal number of participants are assigned to each condition. But simple randomization (e.g., flipping a coin or rolling a die) will normally not achieve this goal, for there is random variation in the outcome. (Assume you want 10 participants in each of two cells, there is only about a 18% chance that a coin flip will result in equal size samples, and about a 26% chance that there will be 7 or fewer in one group. Indeed, as the sample size increases the probability of exact equal numbers per condition decreases, even though the chance of large variations from equality also decreases.)

A seeming solution to this problem is to randomly assign participants to conditions until the desired number is achieved in one condition and then to assign the remaining participants to the other condition. Unfortunately, this will normally result in an over abundance of later arriving participants in one condition rather than another. If there is any reason (and there are many, eg., early arriving subjects are likely to be more conscientious, anxious, and introverted than late arriving subjects) to expect that participant arrival is correlated with extraneous variables, then the condition effect is confounded with arrival time (which, for studies in a single day, is confounded with time of day as well).

A solution that guarantees equal numbers of subjects per condition but also has no expected correlation with other variables is to block randomize. That is, to divide the n participants into n/k blocks where k is the number of conditions. Then, within each block, randomly assign participants to condition by choosing the condition for a participant by sampling without replacement from the set of all conditions.

With random assignment of participants to conditions, the expected correlation of experimental manipulation with other possible confounds is 0. However, if not all participants complete the assigned conditions, problems can arise. For instance, high impulsive subjects tend to be not very wide awake in the morning and are much more likely to drop out of studies when assigned to morning versus evening conditions. Randomly assigning them to morning or evening avoids problems with differential volunteering but the problem of differential attrition still needs to be considered.

Basic experimental designs

There are two broad classes of experimental designs used in personality research. The first, between subjects, randomly assigns participants to conditions, where each person is in just one condition. The second, within subjects, assigns each person to all conditions. These two broad types of designs can be combined into mixed designs where participants

are assigned to multiple but not all conditions. Although the examples discussed below use just one PV and one EV, the generalization to multiple PVs and multiple EVs is straightforward.

Between subject

The classic experiment is to administer an experimental variable to two groups of randomly assigned participants. Unfortunately, by ignoring individual differences, the classic experiment can not test any hypothesis about personality. But, with the addition of a Person Variable to the study, we have the basic between subject PV x EV study. Until recently the Person Variable was some dichotomously scored trait, resulting in two levels of the PV and two levels of the EV and the analysis was a classic analysis of variance. With a greater understanding of the general linear model, more recent studies have treated the PV as a continuous variable and analyzed the data in terms of a moderated regression analysis. Some studies, in order to increase power to detect effects of the PV give a pre-test and then select participants with extreme scores on the PV.

Random assignment of participants to conditions will tend to eliminate confounds of the EV with extraneous variables but even the best of randomization and counterbalancing can not control for confounds introduced by the EVs. Avoiding such confounds requires a theoretical understanding of the task rather than just a mechanical consideration of design. Consider the interactive effect of task difficulty and anxiety on performance in list learning. Although more than 20 studies showed that anxious subjects learn easy lists more rapidly than do less anxious subjects, but learn difficult lists more slowly than do the less anxious, (e.g., Spence, Farber and McFann, 1956) this effect was shown to be due to a confound of task difficulty and implicit feedback (Weiner and Schneider, 1971). The traditional list learning task used a serial anticipation procedure in which participants would be shown a cue word, recite their answer, and then be shown the correct answer. Although no feedback was explicitly given, implicitly, participants could judge how well they were doing whenever they would make an incorrect response. Weiner and Schneider used the same procedure, but added explicit feedback (“compared to other students you are doing very well or not very well”). Feedback interacted with anxiety such that high anxious participants with either difficult or easy lists to learn did better when told they were doing better than average, but did less well when they were told they were doing worse than average. As is true with most interactions, the Weiner and Schneider study may also be interpreted as limiting the generality of the anxiety by task difficulty effect to situations where no feedback is given.

An example of a between subjects study is the examination of the relationship of extraversion and mood induction on positive affect. Contemporary models of introversion-extraversion claim that extraverts are either more likely to be in a positive mood or are more sensitive to positive mood inductions. These hypotheses may be examined by a simple PV x EV experiment where extraversion is indexed by some self report questionnaire and a mood induction such as showing a short film clip with humorous content (e.g., the birthday party scene from *Parenthood* versus a control film clip with neutral content (e.g., a National Geographic film about African wildlife).

Positive affect could be assessed with a short rating scale including items such as happy and pleased. Alternative measures and manipulations could include peer ratings of extraversion, and instructions to think about a positive event (finding \$20 while walking on a beach) or a neutral event (thinking about studying in the library).

Within subject

A way of controlling for large between subject variation in response, particularly when examining interactions with cognitive or psychophysiological variables, is the within subject design in which all levels of the experimental variable are presented to each subject. For instance, testing the hypothesis that there is cerebral lateralization of the response to positive stimuli and that this should depend upon levels of extraversion, Canli et al. (2002) examined the BOLD response (Blood Oxygen Level Dependent changes) in a functional Magnetic Resonance Imaging (fMRI) paradigm. Correlations of the BOLD response to positive versus negative stimuli as a function of Extraversion showed left lateralization of the response to positive stimuli as a function of Extraversion. The within subject design examined the response to affectively valenced stimuli compared to neutral stimuli to control for the very large variation between subjects in the amount of brain activation measured.

Similar within subjects design are necessary when using reaction time as the dependent variable. There are large between subject differences in reaction time that reflect both extraneous state variables (e.g., time of day, sleep deprivation) as well as extraneous trait variables (age, intelligence). The effects of these extraneous variables can be minimized by using each subject as their own control. Thus, when examining the relation of Anxiety or Neuroticism to the cognitive impairment induced by fearful or negative stimuli using the “dot probe” paradigm, or using the “emotional Stroop” paradigm, each participant serves as their own control by giving responses to valenced and non-valenced stimuli.

A potential confound in all such within subject studies is the effect of trial order for some the effect of some manipulations can persist well beyond the experiment. This requires running participants on multiple occasions rather than in rapid sequence. Examples of such possible carryover manipulations are the effect of alcohol, caffeine, or nicotine. To use any of these potent arousal manipulations in a within subjects design requires doing the study over multiple days rather than in one session.

If observations are collected within subject across many trials, block randomization can still be used to avoid confounding condition with order. If there are only a few (e.g. 4) observations per participant, then randomly assigning participants to one of two counterbalanced orders avoids possible order effects (e.g., if there are two experimental conditions, half the participants are given the conditions in the order ABBA, while the other half are given BAAB.) With three, four or more levels of a within subject variable, the use of Latin squares removes simple order and first order sequential effects: Participants can then be blocked randomized into each of the orders. E.g. for a study with four conditions:

Participant	Trial			
	1	2	3	4
1	A	B	C	D
2	B	D	A	C
3	C	A	D	B
4	D	C	B	A

Both ABBA and Latin square techniques force the correlation of order and experimental condition to be 0. Note how in the Latin Square not only does every condition appear in every trial, but first order sequential effects (e.g., the frequency with which A precedes versus follows B) is also controlled. (See Fisher and Yates, 1963 for tables of Latin Squares).

Examples of experiments and data analysis

This section will give a brief overview of how to analyze the data from several prototypical personality experiments. To allow the reader to compare alternative analytic strategies, the data are simulated using the R computer language (R Development Core Team, 2005) with the R code included as an appendix. Further details on R and more extensive analyses may be found in an online appendix at the Personality Project (<http://personality-project/r/simulating-personality.html>). The appendix includes the R code for each analysis discussed below as well as additional analyses. By showing the R code for generating simulated data as well as the code for analysis the reader is encouraged to try out some of the techniques for themselves rather than just passively reading yet another chapter.

In all the simulations, data are generated based upon a model that Positive Affect is an interactive and non-linear function of Extraversion and rewarding cues in the environment, that Negative Affect is an additive but non-linear function of Neuroticism and punishing cues in the environment, that arousal is an additive function of stimulant drugs and Extraversion, and that cognitive performance is a curvilinear function of arousal. This model is the “truth” that the analyses hope to uncover. Unlike real studies, in all simulations we have access to the latent (but unobserved) “true” variables as well as the observed Person and Outcome variables. (The simulations are based upon somewhat simplified summaries of a number of studies conducted over the past years at the Personality, Motivation and Cognition lab at Northwestern University but are idealizations and simplifications of the actual results of those studies.) Readers are encouraged to try the simulations and analyses for themselves, varying the sample sizes and the strength of the relationships. By presetting the seed for the random number generator to a memorable constant value (Adams, 1980) the results obtained in these simulations and reported below should match those carried out by copying the code in the appendix.

The first study considers a single Person Variable, e.g., Neuroticism, and a single manipulation, e.g., a negative versus neutral movie. The observed variable is negative

affect. In the simulation, the true model is that latent negative affect is a simple additive function of neuroticism and the movie condition, but that the observed score is a non-linear but monotonic effect of the latent score. That is, Negative Affect = $1/(1+\exp(-\text{Movie} - \text{Neuroticism}))$. This particular equation is the logistic function commonly used in Item Response Theory analyses of responses to personality and ability items.

The second study considers a single Person Variable, e.g. extraversion, and a single manipulation, e.g., a positive versus neutral movie. The observed variable is positive affect. In the simulation, positive affect is a monotonically increasing function of the interaction of extraversion and the mood induction. That is: Positive Affect = $1/(1+\exp(-\text{Extraversion} * \text{Movie}))$.

The third study is just a combination of the first two, and analyzes the data from studies 1 and 2 as part of one larger regression model.

The fourth study examines the effects of two levels of caffeine induced arousal on performance of introverts and extraverts in a within subjects design. The underlying model is that performance is an inverted U shape function of arousal and that arousal is a decreasing function of extraversion. (Ignoring the time of day effects that are most interesting, see Revelle, et al., 1980).

Study 1: The effect of Neuroticism and a negative mood induction upon Negative Affect. 100 participants, differing in Neuroticism are block randomly assigned to one of two movie conditions. Neuroticism was assessed by the Eysenck Personality Questionnaire (Eysenck and Eysenck, 1976) and the movie conditions were 9 minute selections from a *PBS Frontline* episode (May, 1985) depicting the allies' liberation of Nazi concentration camps and a *National Geographic* film depicting animals in their natural habitat, grazing. (See Rafaeli and Revelle, in press or Rogers and Revelle, 1998 for actual results using these manipulations.)

The analysis used the general linear model procedure from R with the model:
Negative affect = β_1 Neuroticism + β_2 Movie + β_3 (Neuroticism * Movie).

Movie was made a categorical “factor” and Neuroticism was “centered” around the mean. Centering is required when doing regression models with interaction terms for proper interpretation of the main effects. The three regression coefficients (β_1 , β_2 , β_3) were estimated using the Linear Model command and the magnitude of a t-statistic (the ratio of the estimate to the standard error of the estimate) gives us confidence in the estimates. The summary statistics for the model show that both the neuroticism slope (.72) with a standard error of .09 and that the movie slope (1.09) with a standard error of .11 are reliable effects (t 's > 8.0, $p < .001$), but that the interaction effect, with a negligible slope (.02) and a much larger standard error (.13) does not differ from 0. The model fit is shown graphically and is compared to the true model in Figure 3.

Insert Figure 3 about here

Study 2: The effect of Extraversion and a positive mood induction upon Positive Affect. 100 participants, differing in Introversion-Extraversion are block randomly assigned to one of two movie conditions. Extraversion was assessed by a short measure of the Big 5 using items from the International Personality Item Pool (Goldberg, 2006) and the movie conditions were 9 minute taken from the 1989 film Parenthood and a National Geographic film depicting animals in their natural habitat, grazing. (See Rafaeli and Revelle, in press or Rogers and Revelle, 1998 for actual results using these manipulations.)

The analysis used the general linear model procedure from R with the model:
Positive affect = b_1 Extraversion + b_2 Movie + b_3 (Extraversion * Movie).

Movie was made a categorical “factor” and Extraversion was “centered” around the mean. The three regression coefficients ($\beta_1, \beta_2, \beta_3$) were estimated using the Linear Model command and the magnitude of a t-statistic gives us confidence in the estimates. The summary statistics for the model show that there is no effect for extraversion ($b_1 = -.06$ with a standard error of .07) but there is a strong effect for the movie ($1.6, se = .11$) and for the interaction ($.37$ with $s.e. = .10$). As may be seen in Figure 4, the observed interaction suggests that the slopes in the two conditions go opposite directions, but this is due to sampling error.

Insert Figure 4 about here

Study 3: The effects of Extraversion and Neuroticism and Positive and Negative Mood inductions upon positive and negative affect. This are just the data from studies 1 and 2 combined into one analysis, noting that the affect measures are repeated within participants. Because of the within subjects design, the analysis is slightly more complicated and can be done either as a repeated measures ANOVA or as Mixed Effects analysis (Pinheiro and Bates, 2000).

The data are organized as a function of subject, movie conditions, extraversion and neuroticism. Although some statistical packages (e.g., SPSS and SYSTAT) treat repeated measures as separate columns in the data matrix, in R it is necessary to “stack” the repeated measures so that regressions with the categorical condition variable may be found. (See the online appendix for details.) The model is

$$\begin{aligned} \text{Response} = & \beta_1 \text{AffectMeasure} + \beta_2 \text{Extraversion} + \beta_3 \text{Neuroticism} + \beta_4 \text{PositiveMovie} + \\ & \beta_5 \text{NegativeMovie} + \beta_6 \text{Extraversion} * \text{PositiveMovie} + \\ & \beta_7 \text{Extraversion} * \text{NegativeMovie} + \beta_8 \text{Neuroticism} * \text{PositiveMovie} + \\ & \beta_9 \text{Neuroticism} * \text{NegativeMovie} + \beta_{10} \text{Extraversion} * \text{Neuroticism} * \text{PositiveMovie} + \\ & \beta_{11} \text{Extraversion} * \text{Neuroticism} * \text{NegativeMovie} + \beta_{12} \text{Extraversion} * \text{PositiveMovie} * \text{NegativeMovie} + \\ & \beta_{13} \text{Neuroticism} * \text{PositiveMovie} * \text{NegativeMovie} + \beta_{14} \text{Extraversion} * \text{Neuroticism} * \text{PositiveMovie} * \text{NegativeMovie} + \\ & \text{Error}(\text{subject}) \end{aligned}$$

As would be expected, the results show effects of the positive and negative movie conditions, neuroticism, and interactions of the positive mood condition with

extraversion, and interactions of type of affect with positive and negative movies, and a triple interaction of affect type with positive movies and extraversion. To show these effects graphically is somewhat more complicated, and the graph becomes a four panel graph.

Insert Figure 4 about here

Conclusion:

References

- Adams, D. (1979) *Hitchhikers Guide to the Galaxy*,
- Anderson KJ, Revelle W. (1982) Impulsivity, caffeine, and proofreading: a test of the Easterbrook hypothesis. *Journal of Experimental Psychology: Human Perception and Performance*. 1982 8, 614-24.
- Anderson, K. J. & Revelle, W. (1994) Impulsivity and time of day: is impulsivity related to the decay of arousal? *Journal of Personality and Social Psychology*, 67,334-344.
- Canli, T, Sivers, H. Whitfield, S.L. Gotlib, I.H. & Gabrieli, J.D.E. (2002) Amygdala Response to Happy Faces as a Function of Extraversion, *Science*, Vol. 296. no. 5576, p. 2191
- Cronbach, L.J. (1957) The two disciplines of scientific psychology. *American Psychologist*, 12, 671-684.
- Cronbach, L.J. (1975) Beyond the two disciplines of scientific psychology. *American Psychologist*,
- Eysenck, H.J. (1966) Personality and experimental psychology. *Bulletin of the British Psychological Society*, 19, 1-28.
- Eysenck, H.J. (1997) Personality and experimental psychology: The unification of psychology and the possibility of a paradigm. *Journal of Personality and Social Psychology*, 73 , 1224-1237.
- Doyle, A. C. (1929) *Sherlock Holmes: A study in scarlet; The sign of four; The hound of the Baskervilles; The valley of fear; the complete long stories by Arthur Conan Doyle*. London, J. Murray
- Elliot, A. J., & Thrash, T. M. (2002). Approach-avoidance motivation in personality: Approach-avoidance temperaments and goals. *Journal of Personality and Social Psychology*, 82, 804-818.
- Eysenck, H.J. 1967 *Biological Basis of Personality*. Springfield: Thomas.
- Fisher, R.A., & Yates, F. (1963) *Statistical Tables for Biological, Agricultural and Medical Research*, 6th Ed. Hafner Press (Macmillan), New York.
- Gilboa, E., & Revelle, W. (1994). Personality and the structure of emotional responses. In S. Van Goozen, N. E. Van de Poll, & J. A. Sargent (Eds.), *Emotions: Essays on current issues in the field of emotion theory* (pp. 135-159). Hillsdale, NJ: Erlbaum.
- Goldberg, L. R. (1999). A broad-bandwidth, public-domain, personality inventory

- measuring the lower-level facets of several five-factor models. In I. Mervielde, I. Deary, F. De Fruyt, & F. Ostendorf (Eds.), *Personality Psychology in Europe* (Vol. 7; pp. 7-28). Tilburg, The Netherlands: Tilburg University Press.
- Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Cloninger, C. R., & Gough, H. C. (2006). The International Personality Item Pool and the future of public-domain personality measures. *Journal of Research in Personality*, 40, 84-96.
- Kuhl, J., & Blankenship, V. (1979). The dynamic theory of achievement motivation: From episodic to dynamic thinking. *Psychological Review*, 86, 141-151.
- Larsen, R.L., & Ketelaar, T. (1989). *Extraversion, neuroticism and susceptibility to positive and negative mood induction procedures. Personality and Individual Differences*, 10, 1221-28.
- Leon, M.R., & Revelle, W. (1985) The effects of anxiety on analogical reasoning: a test of three theoretical models. *Journal of Personality and Social Psychology*, 49, 1302-1315.
- Ortony, A., Norman, D.A. & Revelle, W. (2005): Effective Functioning: A Three Level Model of Affect, Motivation, Cognition, and Behavior. in J. M. Fellous & M. A. Arbib (Eds.), *Who Needs Emotions? The Brain Meets the Machine*. New York: Oxford University Press.
- Ozer, D. (2006) Personality and the Prediction of Consequential Outcomes. *Annual Review of Psychology*, 57, 401-422.
- Pinheiro, J. C. & Bates, D.M. (2000) *Mixed-effects models in S and S-Plus*. Springer, New York.
- Platt, John R. (1964) Strong Inference *Science* 146: 347-353
- R Development Core Team (2005). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.
- Rafaeli, E. & Revelle, W. (in press) A premature consensus: Are happiness and sadness truly opposite affects? *Motivation and Emotion*.
- Revelle, W., Amaral, P., & Turriff, S. (1976). Introversion/extroversion, time stress, and caffeine: effect on verbal performance. *Science*, 192, 149-150.
- Revelle, W. and Anderson, K.J. (1992) Models for the testing of theory. In A. Gale and M.W. Eysenck (Eds.) *Handbook of Individual Differences: Biological Perspectives*. Wiley: Chichester, England.
- Revelle, W., Humphreys, M.S., Simon, L., and Gilliland, K. (1980) The interactive effect

- of personality, time of day, and caffeine: a test of the arousal model. *Journal of Experimental Psychology: General*, 109, 1-31.
- Rogers, G. and Revelle, W. (1998) Personality, mood, and the evaluation of *affective and neutral word pairs*. *Journal of Personality and Social Psychology*, 74, 1592-1605
- Shadish, W. R., Cook, T.D., & Campbell, D.T. (2002) *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston, Houghton Mifflin.
- Spence, K. W., Farber, I. E., & McFann, H. H. (1956). The relation of anxiety (drive) level to performance in competition and non-competition paired-associate learning. *Journal of Experimental Psychology*, 52, 296-305.
- Öhman, A. & Mineka, S. The malicious serpent: snakes as a prototypical stimulus for an evolved model of fear. *Current Directions in Psychological Science*. 12, 5-9.
- Weiner, B. & Schneider, K. (1971). Drive versus cognitive theory: A reply to Boor and Harmon. *Journal of Personality and Social Psychology*, 18, 258-262.
- Zinbarg, R. & Revelle, W. (1989) Personality and Conditioning: A test of four models. *Journal of Personality and Social Psychology*, 57, 301-314.
- Zinbarg, R.E., Revelle, W., Yovel, I., & Li, W. (2005). Cronbach's Alpha, Revelle's Beta, McDonald's Omega: Their relations with each and two alternative conceptualizations of reliability. *Psychometrika*. 70, 123-133.

Table 1: Examples of Experimental Personality Research

Person Variable	Experimental Variable	Hypothetical State	Observed Variable
Introversion/Extraversion	Caffeine	Arousal	Cognitive performance: total correct, attention decrements over trials, reaction time, accuracy, speed accuracy tradeoffs
	Time of day	Arousal	
	Time limits		
	Movies	Positive Affect	Cognitive Performance: reaction time to valenced words
	Autobiographical memory	Positive Affect	Mood ratings
	Affective pictures	Positive Affect	MRI activation
Emotional Stability /Neuroticism	Movies	Negative Affect	Cognitive Performance
	Affective Pictures		MRI activation
Impulsivity	Cues for reward/punishment	Behavioral Activation	Learning
	Time of day	Arousal	Cognitive Performance:
	Caffeine	Arousal	Cognitive Performance:
Achievement motivation	success/failure feedback	Achievement motive	Persistence Task choice
Chronic Promotion Focus/Prevention Focus	Cues for reward/punishment	Activation of Promotion Focus/Prevention Focus	
Anxiety	Success vs. Failure feedback	State anxiety	Learning
	Task difficulty	State anxiety	Learning
	Memory load	State anxiety	Cognitive performance: speed accuracy trade off
	Cues for reward/punishment	Behavioral Inhibition	Learning
	Autobiographical memory	State anxiety Negative Affect	Emotional “Stroop” task
	Fearful pictures	State anxiety	Illusory correlation
Conscientiousness	Forest vs. trees		Reaction time

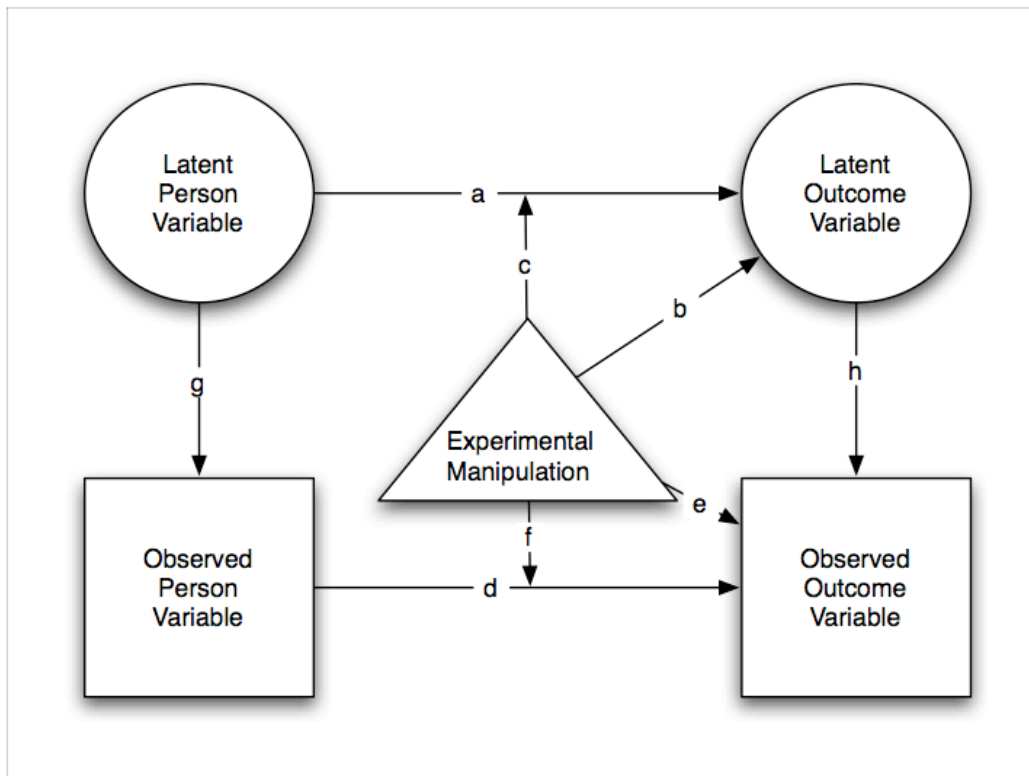
Figure Captions

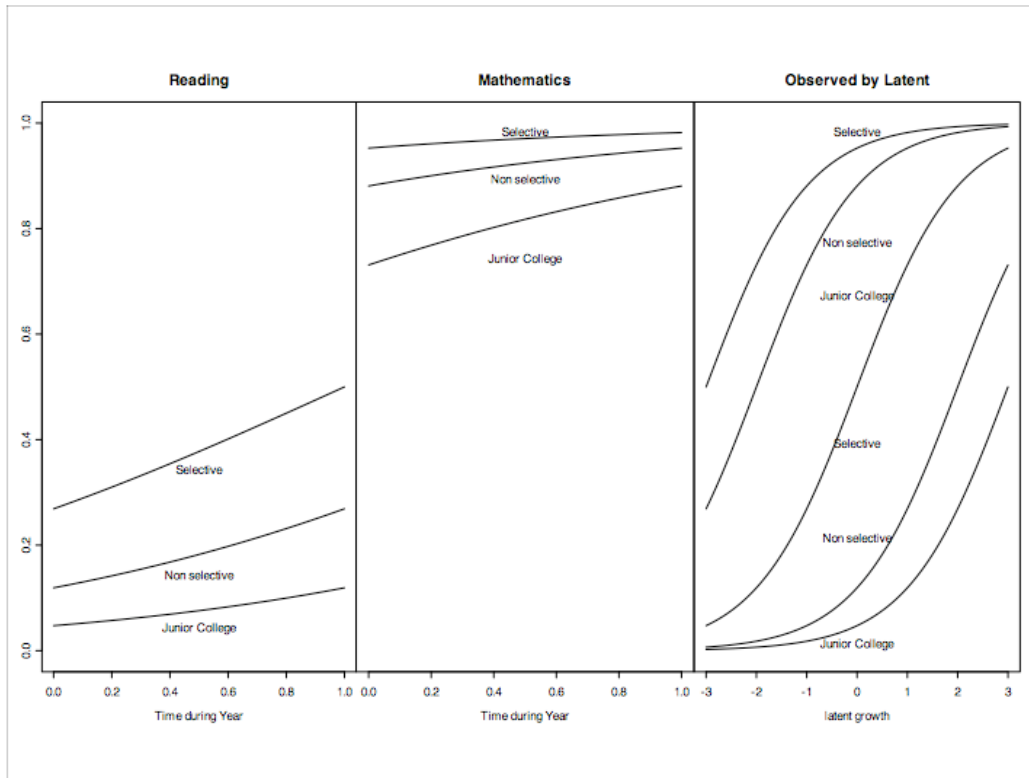
Figure 1: The problem of inference: Observed Person and Experimental Variables, Latent Person and Outcome Variables, Observed Outcome Variables. The strength of latent relationships (a,b,c) are estimated by the strength of observed relationships (d,e,f) and reduced by the validities (g,h) of the measures. (Paths c and f represent the interaction of the experimental manipulation with either the Latent (c) or Observed (d) variables. Paths g and h reflect the square roots of the reliabilities of the observed variables but also any non-linearities in the Latent to Observed variables.

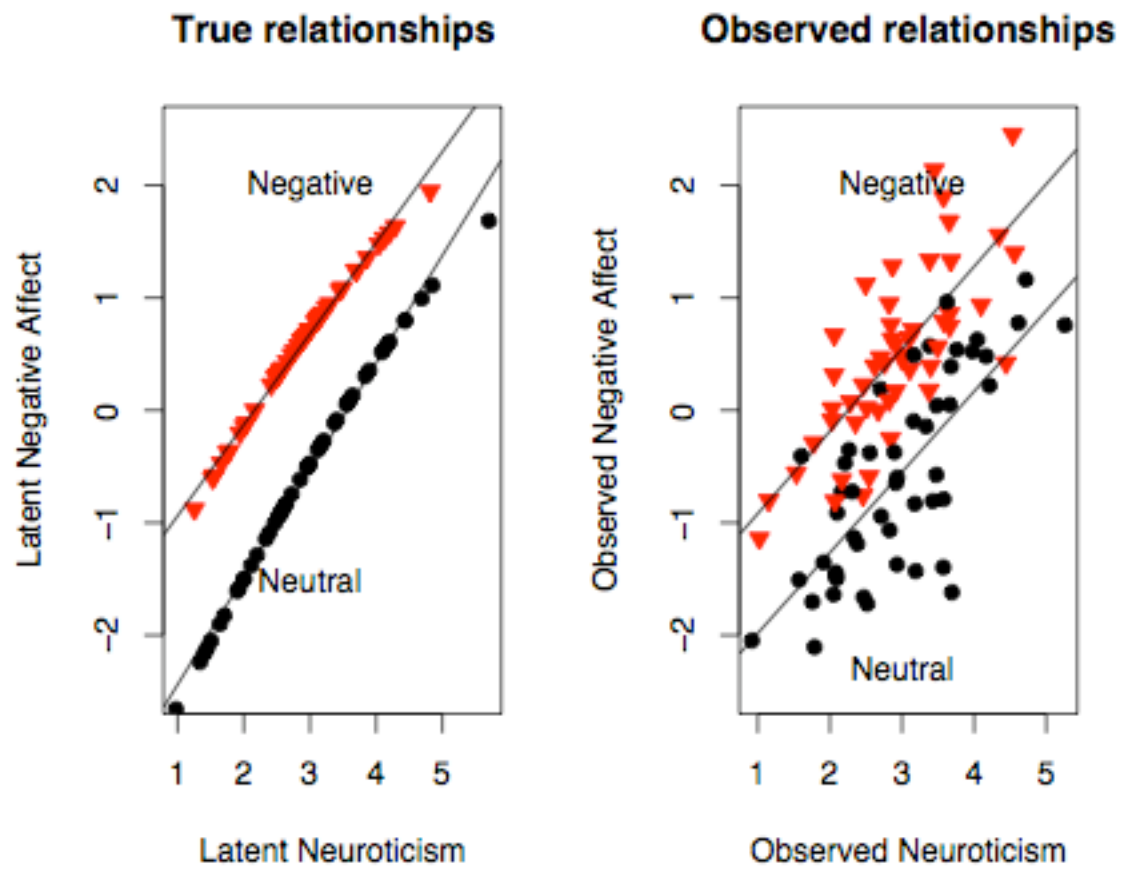
Figure 2: The problem of measurement: Panel 1) Observed scores: The effect of a year of schooling and college type upon writing performance; Panel 2) Observed scores: The effect of a year of schooling and college type upon mathematics performance; Panel 3) Latent scores: The effect of a year of schooling, college type and performance measure. The seeming interactions in panels 1 and 2 are due to the difficulty of the measures.

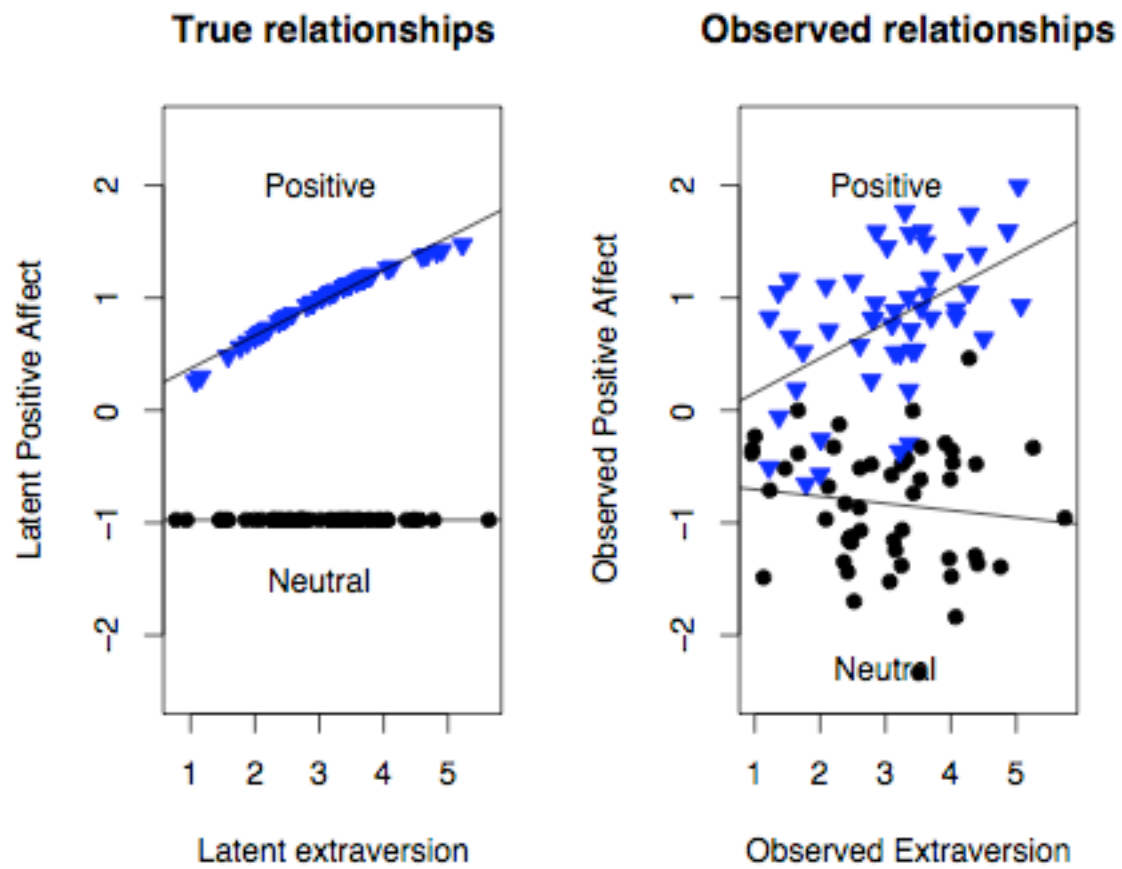
Figure 3: Analysis of simulated study 1. Negative affect is an additive function of Neuroticism and a Negative Affect manipulation. Panel 1) Latent scores. Panel 2) Observed scores.

Figure 4: Analysis of simulated study 2. Positive affect is an interactive function of Extraversion and a positive affect manipulation. Panel 1) Latent scores. Panel 2) Observed scores.









Appendix 1

```

#R code for simulating personality x situation effects
#a more extensive and annotated version of this may be found in the online supplement:
#also available at http://personality-project.org/r/simulating\_personality.html
set.seed(42) #random number seed is fixed to produce identical "random" sequences

#first set some parameters of the model
num <- 100 #number of people to simulate
weightreward <- .4 #an index of how strong is the effect of reward on positive affect
weightpunish <- .4 #how strong is the effect of punishment on negative affect
weight_e <- .0 #how strong is the effect of extraversion on positive affect
weight_n <- .3 #how strong is the effect of neuroticism on negative affect?
weight_er <- .4 #how strong is the interaction of e x reward on positive affect
weight_np <- 0 #how strong is the interaction of n * punish on negative affect
reliability_e <- .7 #the reliability of the observed extraversion score
reliability_n <- .8 #the reliability of the observed neuroticism score
reliability_P <- .7 #the reliability of observed measures of Positive Affect
reliability_N <- .8 #the reliability of observed measures of Negative Affect

#generate the data using the random normal distribution

#first simulate true (latent) scores
true_e <- rnorm(num) #true trait extraversion is normally distributed
true_n <- rnorm(num) #true trait neuroticism is normally distributed

#observed E and N are a mixture of true and error scores
extraversion <- sqrt(reliability_e)*true_e + sqrt(1-reliability_e)*rnorm(num)
neuroticism <- sqrt(reliability_n)*true_n + sqrt(1-reliability_n)*rnorm(num)

#experimental conditions are block randomized
reward <- rep(c(0,1),num/2) #reward vector of 0 or 1
punish <- rep(c(0,0,1,1),num/4) #punishment vector
block <- sort(rep(seq(1:(num/4)),4)) #block the data to allow for block randomization
condition <- data.frame(reward,punish,block,rand =rnorm(num)) #experimental conditions ordered by block
condition <- condition[order(condition$block,condition$rand),] #conditions are now block randomized

#true affect measures are a function of a trait, a situation, and their interaction
TruePosAffect <- 1/(1+exp(-(weight_e*true_e + weightreward*condition$reward +weight_er*true_e*condition$reward)))
TrueNegAffect <- 1/(1+exp(-(weight_n*true_n + weightpunish*condition$punish +weight_np*true_n*condition$punish )))

```

```

TruePosAffect <- TruePosAffect/sd(TruePosAffect) #standardize them
TrueNegAffect <- TrueNegAffect/sd(TrueNegAffect) #standardize them

#observed affect is a function of true affect and errors in measurement
PosAffect <- sqrt(reliability_P) * TruePosAffect+sqrt(1-reliability_P)*rnorm(num)
NegAffect <- sqrt(reliability_N) * TrueNegAffect+sqrt(1-reliability_N)*rnorm(num)

#organize all the data in a data frame to allow for analysis
affect.data <-
data.frame(extraversion,neuroticism,PosAffect,NegAffect,condition,true_e,true_n,TruePosAffect,TrueNegAffect)
centered.affect.data <- data.frame(scale(affect.data[,1:10],scale=FALSE),reward =
as.factor(condition$reward),punish=as.factor(condition$punish))
#do the analyses using a linear model
#just look at main effects
mod1 <- lm(NegAffect ~ neuroticism+punish,data = centered.affect.data)
mod2 <- lm(PosAffect ~ extraversion+reward,data= centered.affect.data) #don't examine
interactions

#include interactions
# note that mod3 and mod4 are two different ways of specifying the interaction
mod3 <- lm(PosAffect ~ extraversion*reward,data = centered.affect.data) #look for
interactions
mod4 <- lm(NegAffect ~ neuroticism+ punish + neuroticism*punish,data =
centered.affect.data)
#go for the full models
mod5 <- lm(PosAffect ~ extraversion*neuroticism*reward*punish,data =
centered.affect.data)
mod6 <- lm(NegAffect ~ extraversion*neuroticism*reward*punish,data =
centered.affect.data)

#show the results of the analyses
summary(mod1,digits=2)
summary(mod2,digits=2)
summary(mod3,digits=2)
summary(mod4,digits=2)
summary(mod5,digits=2)
summary(mod6,digits=2)

```

Outtakes

The basic experiment in personality examines the effect of some manipulated variable (the IV) on an outcome variable (the DV) while at the same time examining the effect of person variable (PV) upon the DV. However, merely showing that both the PV and IV have an effect is not as useful for theory testing as is finding that the PV by DV relationship depends upon (interacts with) the IV.

We need to distinguish between experiments, in which a variable is manipulated, from observations that require sophisticated technology (e.g., Magnetic Resonance Imaging) but do not actually manipulate the state of the subject. That is, to show that the amygdala of subjects shows more activation following fear cues than neutral cues, and that neurotics have more amygdalar activation than more emotionally stable subjects is interesting, but is really merely a correlation between the person variable and the outcome variable.

Broadly speaking, there are three possible classes of personality theories: those that describe stable trait differences that result in constant differences in behavior across situations, those that propose stable trait differences across situations that have additive effects with environmental variations, and those that interact with the environment.