Integrating experimental and observational personality research – the contributions of Hans Eysenck

William Revelle and Katherine Oehlberg
Northwestern University

Abstract

A fundamental aspect of Han Eysenck's research was his emphasis upon using all the tools available to the researcher to study personality. This included correlational, experimental, physiological, and genetic approaches. 50 years after Cronbach's call for the reunification of the two disciplines of psychology (Cronbach, 1957) and 40 years after Eysenck's plea for experimental approaches to personality research (H. J. Eysenck, 1966), what is the status of the unification? Are personality researchers taking advantage of experimental techniques to tease out causality, and are we communicating the advantages of combining experimental with multivariate correlational techniques?

Introduction

A central theme of Hans Eysenck's research and writings was the integration of the scientific study of personality into the field of psychology as a whole, as well as the rest of the natural sciences (H. J. Eysenck, 1966, 1997; H. J. Eysenck & Eysenck, 1985). Genetic and physiological questions were as much a part of Eysenck's theoretical framework as were basic findings in learning and motivation (H. J. Eysenck & Eysenck, 1985). He pioneered the use of the most recent developments in psychological measurement and psychometrics and the application of these techniques to self-report and behavioral observations. Unsatisfied with merely trying to utilize classic experimental psychology as a guide for personality theory, Eysenck also emphasized the contribution that personality theory and research could make to the seemingly unrelated research questions of experimental psy-

Prepared for a special issue of the *Journal of Personality* on Eysenckian Themes, edited by Kate Walton and Robert Krueger. Revised version submitted March 20, 2007.
contact: William Revelle: revelle@northwestern.edu

chology[1] (H. J. Eysenck, 1966, 1983, 1997). In this article we evaluate the degree to which current work in personality theory and research has aimed at and reached Eysenck's lofty goals of the integration of these two fields. In order to do so, we attempt to map some of the recent personality literature unto a four dimensional taxonomy of levels of analysis, degrees of generality, quality of functioning, and research method.

Personality and Experimental Psychology

Ever since Wundt introduced experiments into psychology (Wundt, 1874/1999, 1904) and Galton (1892) studied individual differences in genius, there has been a persistant tension between the experimental and correlational methodological and statistical approaches taken by experimental and personality psychology, respectively. Cronbach (1957, 1975), H. J. Eysenck (1966, 1997) and Vale and Vale (1969), however, highlighted the strengths and weaknesses of the alternative approaches and argued for the reunification of the two disciplines. They believed that the field of psychology would be improved if experimentalists and correlationalists could share methods, theories, and findings. Eysenck's most impressive statement of the need to combine the two disciplines was his (posthumous) 1997 paper contending that personality researchers should adapt a paradigmatic approach (H. J. Eysenck, 1997) in order to make progress. Following Kuhn (1970), he used paradigm (H. J. Eysenck, 1997) to refer to a coherent theoretical and methodological model within which a scientific field conducts its work. He suggested that personality psychology, insofar as it resisted the integration of experimental methods, remained pre-paradigmatic; that is, it lacked an explicit framework that related constructs via causal mechanisms, and moreover lacked the ability to test hypothesized causal relationships. Most importantly, he suggested that a research agenda combining experimental and correlational techniques to develop and test causal theories of personality was crucial for the field to develop a paradigm within which progress can be made.

As many readers will recognize, dichotomizing research approaches into the experimental and correlational confounds research design with the method of data analysis. The traditional statistical tool for the experimentalist has been the comparison of means using the t-test or its generalization, the analysis of variance (ANOVA). This is in contrast to the analysis of variability and covariance using the correlation coefficient and multivariate procedures. However, because ANOVA and the correlation coefficient are both special cases of the general linear model, it is better to consider the distinction to be between *experimental* and *observational* methods rather than *experimental* and *correlational* analysis.

Perhaps Eysenck's greatest strength was his commitment to developing personality psychology into a mature scientific field of inquiry. By that, he meant one in which we have gone beyond observations and hunches to the development and testing of causal models. He

[1]The terms used by Cronbach (1957) and H. J. Eysenck (1966) seem somewhat quaint in that now most psychologists refer to cognitive psychology or cognitive-neuro psychology for what used to be the domain of "experimental" psychology.

observed that scientific inquiry in general, and personality theory in particular, ranges from inspired hunch to formal theory and hoped that it was possible to develop formal theory that was subject to rigorous test. In addition to his concern with developing good measures of personality traits, he was an advocate of experimental and physiological techniques to tease apart the intricacies of personality, for he recognized that it was impossible to *test* causal theories from even the best of observational analysis. To Eysenck, factor analysis and structural equation modeling were tools to *describe* structure, but not tools to *explain* structure or process. For explanation, experiments were required.

Importance of Individual Differences for experimental psychology

According to Eysenck, the failure to integrate experimental with observational evidence was not just an oversight of observationalists who do not consider experimental evidence; it was also a weakness of experimentalists who treat all subjects as if they were the same. Eysenck argued that experimental psychologists need to consider how individual differences affect their findings just as chemists need to consider how different elements react differently (H. J. Eysenck, 1966). For instance, no chemist would say "stuff dissolves in water" or even "some stuff dissolves in water, other stuff doesn't," but rather would examine the properties of molecules that lead to water solubility. Most experimentalists do appreciate that individuals differ in their response to experimental conditions; however, they tend to view these differences as nuisances that must be controlled for by using proper (usually within-subject) experimental designs.

There are, of course, exceptions to the generalization that experimentalists ignore individual differences. For example, Underwood (1975) considered individual differences to be the crucible of psychological theory. After years of fruitfully investigating learning using experimental methods, he realized that the theoretical inferences drawn from his findings implied a basic assumption that people differed in their learning experiences, and that these differences mediated the effects that he observed. Unless there were systematic individual differences in response to the manipulations, his theoretical explanations would be false (note that he was primarily concerned with individual differences in states rather than stable differences in traits). For the cognitive research program of Broadbent (1971), individual differences were a source of hypotheses that led to elegant generalizations of models of decision processes. For example, the similarity of the effects of sleep deprivation and extraversion on vigilance performance led him to search for a common cause (arousal) to both the experimental and observational variables. To yet other experimentalists, individual differences are interesting extensions of cognitive (M. W. Eysenck & Calvo, 1998; M. W. Eysenck & Mathews, 1987) or drive theory (Spence, Farber, & McFann, 1956) as applied to real world problems such as anxiety. Individual differences in state anxiety are thought to influence the working memory capacity of participants involved in cognitive processing (M. W. Eysenck & Calvo, 1998) or the excitatory potential while learning easy versus difficult lists in a serial anticipation task (Spence et al., 1956). For those of us who

include intelligence as an aspect of personality, the work of Hunt has always been an example of the integration of experimental cognitive psychology with the study of individual differences (Hunt, 1995; Waller, Knapp, & Hunt, 2001).

Individual differences as sources of variation to be controlled

Most experimental psychologists treat individual differences not as a source of theory, but rather as a nuisance to be controlled for by good experimental design. The easiest way to control for individual differences is, of course, merely to increase the sample size. This increases statistical power because the standard errors have been reduced to allow for "statistical significance" for the particular population effect size of interest (see Harlow, Mulaik, and Steiger (1997) for a critique of this approach of conventional null hypothesis testing). Given the size limitations of undergraduate subject pools, it is more typical to use within-subject designs that effectively remove the between individual effects. That participants differ in ability, age, arousal, and motivation, all large sources of variance in reaction time, is irrelevant if one is concerned with measuring reaction time (RT) differences associated with semantic priming or perceptual interference in a global-local task. RT paradigms are particularly sensitive to the power of within-subject designs: the between conditions effects might be of the order of 10-20 ms and the within subject standard deviations are of the order of 50 ms. Even worse, the stable between subject standard deviations are of the order of several hundred ms. Thus, using participants as their own control increases the power of the design.

Interactions with individual differences can mask effects

Unfortunately for experimentalists, systematic interactions of individual differences with many experimental variables can mask some very important findings relating situational manipulations to performance. H. J. Eysenck (1966, 1967, 1997) reviewed excellent examples of cross-over interactions of personality variables (specifically extraversion, impulsivity, and neuroticism) with situational manipulations, and a comprehensive review of the power of integrating experimental approaches with personality was the landmark 1985 text (H. J. Eysenck & Eysenck, 1985). For example, Shigehisa and Symons (1973) investigated personality effects on multimodal stimulation and showed that for more introverted participants, auditory sensitivity was an inverted U shape function of light intensity (that is, it was a positive function of light intensity for low levels of intensity, but a negative function for high intensities). This was in contrast to the finding that the auditory sensitivity of more extraverted participants increased monotonically as a function of light intensity. Howarth and Eysenck (1968) found that verbal recall was an interactive function of extraversion and recall interval with more introverted participants recalling more as the recall interval increased but more extraverted participants recalling less as the recall interval increased. This result is consistent with examinations of arousal effects on memory where low arousal seems to facilitate immediate recall but hinder later recall and high arousal hinders

immediate but facilitates delayed recall (Revelle & Loftus, 1990). In another experiment, H. J. Eysenck and Levey (1972) showed that eyeblink conditioning was better for more introverted participants under weak UCS conditions but better for more extraverted subjects under strong UCS conditions. This paper also demonstrated the power of experiments to tease out more subtle interactions: the effect of enhanced conditioning was much larger for the impulsivity rather than sociability component of extraversion as measured by the Eysenck Personality Inventory (H. J. Eysenck & Eysenck, 1964).

Inspired by these demonstrations of the importance of considering individual differences in the context of experimental manipulations, we have shown in a series of experiments at Northwestern that the complex cognitive performance of more introverted subjects is hindered, but that of more extraverted subjects is facilitated, by caffeine and time stress (Revelle, Amaral, & Turriff, 1976), with no main effects of either personality or caffeine and time stress on performance. Followup studies showed that this effect was even more complicated and showed a systematic three way cross-over interaction between personality (impulsivity), caffeine, and time of day with no main effects of either personality, drug, or time of day (Revelle, Humphreys, Simon, & Gilliland, 1980). Consistent with the earlier findings of H. J. Eysenck and Levey (1972), impulsivity rather than sociabiilty was the component of extraversion with the most systematic effect.

The importance of individual differences in physiological responses in brain imaging and genetic paradigms has become increasingly recognized in the past few years. Reminiscent of the suggestions by Underwood (1975), Kosslyn et al. (2002) showed the power of individual differences in understanding physiological processes and how, if ignored, individual differences can mask important findings. A review by Canli (2004) in this journal and chapters (Canli, 2006b; Lesch & Canli, 2006) in a recent volume on the biological basis of personality (Canli, 2006a) also make this point very well.

Importance of experimental psychology for personality theory

H. J. Eysenck (1997) answered the question of whether personality had a paradigm with an optimistic "yes" and emphasized the importance of experimental techniques and theory for personality research:

> ... purely taxonomic studies, inevitably correlational in kind, and using factor analytic, multidimensional scaling, and similar methods of analysis, cannot achieve paradigmatic status because of the inevitable subjectivity involved in such studies. What is required is a more theoretical approach seeking causal connections and using experimental tests of deductions from the theories in question. Existing theories have already shown the possibility of this approach in the field of personality and intelligence, enabling researchers to answer questions that a purely correlational approach cannot answer. (H.J. Eysenck, 1997, p 1234)

H. J. Eysenck (1997) emphasized the importance of experiments as ways of testing causal theory. He did not believe that pure observational approaches could be anything more than mere descriptions and *sources* of hunches in a preparadigmatic science. He also believed that personality theory could gain a great deal by taking the finest theories from experimental psychology and specifying how individual differences acted as either parameter settings in these models or as process variables. He wanted to integrate the two approaches into a mature, unified field. Thus, his early work (H. J. Eysenck, 1957) attempted to explain differences in introversion-extraversion in terms of the drive theory models of the day (Hull, 1952), while he later revised these explanations in terms of arousal systems (Broadbent, 1971; H. J. Eysenck, 1967) and then integrated cognitive (H. J. Eysenck & Eysenck, 1985) and molecular genetic (H. J. Eysenck, 1997) findings.[2]

Individual Differences and the range of generalizability

In addition to providing theoretical foundations and allowing causal tests of theory, experiments are capable of better refining our knowledge of personality traits, both extending and limiting the range of their generalizibility. If a personality trait variable interacts with an experimental manipulation, this limits the scope of generality both of the manipulation and of the personality trait. Thus, although discussed above as an example of how interactions can mask effects, H. J. Eysenck and Levey (1972)'s examination of the conditioning theory of socialization (introverts condition more readily, and are thus better socialized), also may be seen as defining the limits of the conditioning theory. In particular, only when the situation was relaxing did introverts condition more rapidly than extraverts; when the situation was stressful, the reverse was the case.

That the performance on complex cognitive tasks of more introverted subjects is hindered but that of more extraverted participants is facilitated by caffeine is interesting and consistent with predictions made by H. J. Eysenck (1967). This prediction followed from his hypotheses that introverts are chronically more aroused than are extraverts and (based upon a generalization of Yerkes and Dodson (1908)) that arousal has an inverted U relationship to performance. But that this effect interacts with time of day such that the effect reverses in the evening limits the generalization that introverts are chronically more aroused than extraverts (Revelle et al., 1980). Moreover, that this effect is mainly due to impulsivity and not to sociability speaks to issues in the measurement of extraversion far better than factor analytical arguments (Rocklin & Revelle, 1981).

In addition to limiting the extent of inferences about personality (Revelle, 2007), experimental designs incorporating personality variables allows the elicitation of a greater range of underlying psychological states (e.g., arousal, fear, positive or negative affect) than would be achievable by simple manipulations that do not take personality into account. That caffeine increases arousal is well known, but the range of arousal can be increased by

[2]In his last talk to the International Society of Individual Difference in 1997, a few months before he died, he said that if he were younger he would try to learn molecular genetics.

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, the selection of depressed versus non-depressed participants greatly enhances the range of negative affective states.

The Person x Situation Debate – revisited

An unfortunate detour for much of American personality research in the 1970s and 1980s was the debate about the relative importance of the person or the situation in predicting behavior (Magnusson & Endler, 1977). Eysenck and his colleagues tended to ignore this controversy, mainly because his work had consistently shown the significance of the situation and its interaction with individual differences. (H. J. Eysenck, 1967; H. J. Eysenck & Eysenck, 1985)

Individual differences, situational profiles and dimensions of situations. What was missing from the personality x situation debate was an emphasis upon defining the situation in a manner that could lead to theoretical predictions about the patterning of individual differences in behavior across situations. Does extraverted behavior lead to positive affect or does positive affect lead to extraverted behavior? Fleeson and his colleagues are starting to tease these questions apart by examining within subject variability across situations, and by inducing people to act extraverted or introverted (Fleeson, Malanos, & Achille, 2002; Fleeson, 2004).

Carving nature at its joints

Some have proposed that personality taxonomies based upon the lexicon (Norman, 1963, 1969) allow us to follow Plato's dictum to "carve nature at its joints," and that by using taxometric methods we can find "natural kinds" of individuals (Gangestad & Snyder, 1985; Asendorf, Borkenau, Ostendorf, & Van Aken, 2001), (but see Zachar (2000) for a contrary view). It is likely, however, that the use of experimental methods will allow us to discriminate between patterns of individual responding as a function of situational manipulations more effectively than by taxometric techniques alone. That the differential patterning of brain activation observed in response to pictures inducing positive or negative mood differs further as a function of extraversion (for positive pictures) and neuroticsm (for negative pictures) (Canli, 2006a) makes an important distinction between the affects and the personality dimensions that can not be done by self report of affect by itself.

Theoretical inference

If we are to do paradigmatic and progressive research in personality, it is important to attend carefully to the logic of the scientific method as it applies to our field. The power of good experimental technique is that sound theoretical progress can be made by

empirically pruning the tree of possible hypotheses and eliminating those inconsistent with experimental results. That is, by following the inductive reasoning procedures advocated by Platt (1964) we follow the method proposed by that great (but mythical) investigator, Sherlock Holmes, who reasoned that when you have eliminated the impossible, whatever remains, no matter how improbable, must be the truth (Doyle, 1929). Although others have argued that Platt misrepresents the process of science (Davis, 2006; O'Donohue & Buchanan, 2001), the emphasis upon asking "the question" of which specific hypothesis a finding disconfirms, or asking what finding would disconfirm a specific hypothesis, nevertheless makes for more critical researchers and logically rigorous theoreticians. It is deceptively easy to claim to test the hypothesis that all swans are white by the confirmatory procedure of looking for white swans rather than seeking to disconfirm by looking for black swans (Popper, 1935)

Theory testing versus confirmatory studies

H. J. Eysenck (1997) worried that the current status of most theory in personality does not allow for the kind of Popperian falsification advocated by Platt (1964), and that an undue reliance on disconfirmatory results would slay theories before they had the chance to mature. He suggested that most personality theories were weak in that they require long chains of assumptions to make their predictions; indeed, he considered most nomological networks of psychology to be composed of more hypothesis than theory. Because of the many inferential steps needed, he suggested that our studies emphasize theory verification rather than theory disconfirmation. At the early stages of theory and construct development, an emphasis on falsification may be problematic, partly because any study on latent psychological variables has many more ways of being wrong (poor theory, poor measurement, poor experimental design) then of being correct (Revelle, 2007; Revelle & Anderson, 1992). However, exclusive emphasis upon confirmatory studies can hinder theory development, in that it fails to prune the tree of alternative hypotheses. As personality theories become stronger and our confirmatory evidence more reliable, the ability to make clearer predictions should be accompanied by increased use of disconfirmatory studies.

The hallmark of good theory, Eysenck's being among the best, is that it is possible to make and test predictions that are direct challenges to it. At least five studies from our lab have been direct tests of two (or more) competing hypotheses derived from Eysenck's 1967 and 1985 theories. One of these was a test of Eysenck's hypothesis that introverts are always more aroused than extraverts (Revelle et al., 1980), a second tested predictions from Eysenck versus Gray (1982) with respect to the role of cues for reward and punishment (Zinbarg & Revelle, 1989), a third tested competing models of the effects of anxiety on cognitive performance (Leon & Revelle, 1985), a fourth compared two hypotheses about the relationship between impulsivity and the decay of arousal (Anderson & Revelle, 1994), and a fifth examined competing explanations for the Yerkes-Dodson "Law" (Anderson, Revelle, & Lynch, 1989; Yerkes & Dodson, 1908).

H. J. Eysenck (1967) claimed that a primary reason that extraverts seek more social

stimulation, smoke more, and engage in more sex and at an earlier age than do introverts is due to their basal level of arousal. Extraverts were thought to be compensating for a low internal level of arousal by seeking more externally induced arousal. In addition, he proposed that there was an optimal level of arousal for performance, with higher or lower levels leading to decrements in performance. But he had also reviewed findings suggesting the introverts and extraverts differed in the phase of their diurnal arousal rhythm, at least as assessed by body temperature. He did not seem to notice that this latter finding was incompatible with his basic hypothesis. The results of seven studies showed that caffeine facilitated the performance on complex cognitive tasks (similar to the Graduate Record Exam) of more extraverted participants in the morning but hindered their performance in the evening (Revelle et al., 1980). This result would be consistent with the hypothesis of greater arousal for extraverts if the detrimental effect of caffeine on the performance of introverts in the morning was even larger in the evening. In direct violation of the assumption of a constant difference in arousal, the performance of the more introverted participants was enhanced in the evening. This results was consistent with the hypothesis that arousal varied diurnally and that the introvert extravert difference was more of phase than of level. Such a finding was, however, at complete odds with the arousal seeking explanation of extraversion, which implies that extraverts would become more "introverted" (e.g., not interested in being sociable or seeking sexual companionship) in the evening, when their arousal is at its highest and most extraverted at dawn, when their arousal level is lowest. We know of no evidence to support this prediction.

Competing versus complementary hypotheses

Psychological theories differ both in breadth and depth. A theory is broader insofar as it incorporates predictions and explanations of more diverse phenomena, and it is deeper according to the detail of the causal explanations in the mechanisms evoked. Theory development, then, consists of increasing the breadth of the theory by extending it into new domains, as well as clarifying the fundamental mechanisms. Eysenck's theory of personality (H. J. Eysenck, 1967; H. J. Eysenck & Eysenck, 1985) was both broad and deep. By integrating self reports with observational and physiological measures it had a breadth far beyond the taxonomic descriptions of the Big 5 (Goldberg, 1990; McCrae & Costa, 1999), and by attempting to attribute cause to genetic predispositions and physiological mechanisms, it had greater depth as well.

A useful heuristic to compare alternative theories is forming a table crossing phenomena as rows and theories as columns. While many cells in the table will be empty as theories may be complementary and speak to different phenomena, some rows will have identical entries across all the columns as alternative theories will all make the same prediction. Theory generalization studies will attempt to add new rows to the matrix. Verification studies will test whether or not a particular phenomena predicted by a particular theory can be observed; failure to verify can happen due to a lack of power, bad design, or an earlier fault in the inferential chain. What can lead to a study of competitive theory test-

ing is a row in which different theories make different predictions. Examples of competing theoretical predictions include those of H. J. Eysenck (1967) versus Gray (1982) in the role of individual differences in conditioning during a go/no go paradigm to cues for reward and punishment (Zinbarg & Revelle, 1989). Intriguingly enough, these results were also relevant for later distinguishing between the original "Gray model" and later refinements of "Reinforcement Sensitivity Theory" by Gray and McNaughton (2000), (Corr, 2007).

Another example of using the matrix of competing theories with multiple phenomena was the examination of anxiety induced decrements on cognitive performance (Leon & Revelle, 1985). Competing explanations for the detrimental effects of anxiety on performance include a narrowing of attention (Easterbrook, 1959), limitations on working memory (M. W. Eysenck, 1979; M. W. Eysenck & Mathews, 1987), and distraction due to off task thoughts (Mandler & Sarason, 1952; Wine, 1971). Using a complex geometric analogies task developed to compare memory and attentional load, Leon and Revelle (1985) found mixed support for the distraction hypothesis, and no support for the effects of anxiety on working memory.

Theory development and theory testing–anomalous findings

However, theory development and testing involves more than simply disconfirming a theory in one study and immediately moving on to do something new and different. The degree to which a particular theory has already received empirical confirmation should inform the interpretation of disconfirmatory findings. H. J. Eysenck (1997) recognized that not all results will be compatible with predictions; in fact, some will even explicitly contradict certain theoretical predictions. Although to Gray (1981) the time of day findings relating extraversion and caffeine induced stress (Revelle et al., 1980) were "a dagger in the heart of Eysenckian theory," to H. J. Eysenck and Eysenck (1985) they were anomalies that required theory modification, but not necessarily theory rejection. Approaching such disconfirmatory results with cautious skepticism, as anomalies indicative of problems with either theory or method, we can avoid abandoning very useful theories that may need only minor modification.

The time of day results (Revelle et al., 1980) did not lead to a complete rejection of the basic model (H. J. Eysenck, 1967), for they were shown to be primarily a function of impulsivity rather than sociability, two components of what was then called extraversion. Psychometric refinement of the scales and some modest modification of the theory led to the revised model (H. J. Eysenck & Eysenck, 1985) that it was the sociability component of extraversion that was more related to stable differences in arousal across the day, and the phase differences in the arousal rhythms observed for impulsivity were no longer incompatible with the revised theory.

Model fitting in structural equation modeling

On the surface, the ability to test alternative structural equation models may seem very similiar to the process of theory testing described above. Although certainly following

the form of hypothesis testing, with statistical tests of the change in model residuals as a function of relaxing one or more model parameters, these procedures do not allow for tests of causal structure for all the same reasons that correlational patterns do not imply causality (Glymour, 2003; Scheines, Spirtes, Glymour, Meek, & Richardson, 1998). Even if a temporal component is added to the model, the structural equations can not show causality. Consider the observation that yellow fingers, yellow teeth, and bad breath at time one are predictors of subsequent lung cancer at time two. Even if we had a structural model that fit these covariances perfectly, we should not conclude that better dental hygiene would protect from lung cancer.

The example of genetic modeling

There is one research area in which experimental and observational data coincide. Nature, by randomly recombining our genes from generation to generation, and by "experimentally" assigning some participants to two alternative twin conditions, provides data that be analyzed as if they were experimental. Structural equation modeling of these "experiments of nature" allow one to tease out genetic effects that would otherwise be untestable. That very complex social behaviors have moderate to strong genetic components is without question; and that these heritabilities do not follow the OGOSH model (One Gene One System Hypothesis) is equally without question. It is implausible that evolutionary pressures have led to particular brain systems for divorce or television viewing, two complex behaviors that are as heritable as most non-cognitive traits (Bouchard, 2004; McGue & Bouchard, 1998).

A taxonomy of psychological research

H. J. Eysenck (1997) organized personality research along a dimension ranging from DNA through physiological structures to social behavior. When we combine these five levels of causal analysis with the distinction of how people are both similar and different made by Kluckhohn and Murray (1953), we have a simple way to organize the psychological literature relevant to personality (Revelle, 1995). We have modified the mid levels of analysis to include recent analyses of the determinants of effective functioning (Ortony, Norman, & Revelle, 2005) as well as studies of the effects of culture; additional dimensions to this taxonomy include the use of experimental versus observational approaches and the level of effective versus ineffective functioning. We have categorized the 64 articles apearing in the 2006 volume of the *Journal of Personality* (JoP) and the 140 articles appearing in the second of the two 2006 volumes of *Personality and Individual Differences* (PaID) in terms of these dimensions (Table 1). In addition, we have performed computer searches for the use of the words "random, experiment, experimental, condition, assigned, iq, ability, or intelligence" for all of the 2005 and 2006 volumes of JoP, PaID as well as the *Journal of*

Research in Personality (JRP), the *Journal of Personality and Social Psychology* (JPSP)[3] and the *European Journal of Personality* (EJP).

Perhaps the most obvious finding from this classification is the infrequency of experimental work published in the last two years in any of the journals (Table 2). 0% of the articles in the EJP, <6% of the articles in JoP, ≈ 12% in the journal that Eysenck edited for 20 years (PaID) and 16% of the articles in JRP contained some experimental study; the journal with the highest percentage of experimental studies of personality was the personality section of JPSP with 28%. The other noticeable difference is that although infrequent, studies of intellectual ability do appear in PaID (≈15%), somewhat less frequently in EJP (3%), and less than 1% in JoP, JRP, or JPSP.

Of the studies that are experimental in PaID, the majority are tests of hypotheses derived from Reinforcement Sensitivity Theory (Gray & McNaughton, 2000). In addition to studies where there were actual experimental manipulations there were a few studies using tasks more typically seen in experimental psychology (e.g., the wholistic-analytic or the "forest-trees" perceptual task developed by Navon (1977)).

The unfortunate conclusion from this brief review of publication practices is that the use of experimental techniques is underemployed in current research. This suggests that the desired unification of the correlational/observational with the experimental disciplines called for by Cronbach and Eysenck has not yet occurred.

Recommendations

H. J. Eysenck (1997) suggested that if we are to progress to the level of paradigmatic research in personality we should address several issues. First and foremost is an emphasis upon programmatic research. That is, more progress can be made working within (and criticizing) a particular common framework rather than repeatedly relabeling old work as "new" and mistaking novelty for progress. If the success of the Big 5 and Five Factor Theory (Goldberg, 1990; McCrae & Costa, 1999) has taught us anything, it is that we can make progress by sharing measurements and constructs across laboratories and research programs. The introduction and availability of the shared item pool as part of the International Personality Item Pool (Goldberg et al., 2006) is an amazing contribution. Another example, particularly relevant for the theme of experimental theory testing, is the progress made in testing what was originally known as "Gray's Theory" (Gray, 1981, 1982) but has become known as "Reinforcement Sensitivity Theory" (Corr, 2007; Gray & McNaughton, 2000; Smillie, Pickering, & Jackson, 2006). Organized around a set of hypotheses about the biological bases of individual differences in anxiety, impulsivity, extraversion, and neuroticism, researchers have focused on improving the measurement model (Smillie et al., 2006), the implications for pathology (Zinbarg & Yoon, 2007), as well as our understanding of the

[3]For JPSP we included only those articles that also had the word "personality" either as a keyword or in the abstract. In the analysis of JPSP we report both the total of articles published as well as that subset having to do with personality.

Table 1: Psychological research may be organized along four dimensions: levels of analysis, levels of generality, quality of functioning, and research design. Using two of these dimensions, levels of analysis and generality, allows for a simple taxonomy of published research. (Adapted from Eysenck (1997), Ortony et al. (2005), and Revelle (1995)).

Levels of Generality	Levels of analysis					
Species Typical	Genome Mapping	Mechanisms: BIS/BAS/FFFS, Neuro-transmitters		Mechanisms: ProtoAffect, Conditioning, Memory, Preparedness	Mood & "full blown" Emotion Cognition & Self regulation	Evolutionary issues, Life preservation, Reproduction, Eating
Individual Differences	allelic variation of specific genes, e.g., 5HTT	Parameters of BIS/BAS/FFFS, Neuro-transmitters	Extraversion Neuroticism Psychoticism Intelligence	Affective Sensitivities & Information Processing Biases	Attributional & Interpretive Biases	Identity & Reputation, Self other reports of "Big 5" traits
Groups and Culture	Genomic Diversity Project	Differential Brain Activation Patterns		Ideological impact on processing	Ideology	Ethnology
Unique Patterning	genetic "fingerprinting"	Specific Lesions		Fears & Phobias	Narrative Identity	Life History
	Genetic Personality Determinants	Biological Systems		Reactive & Routine (Automatic) Processing	Reflective (Conscious) Controlled Processing	Observable Behavior
	Distal Antecedents	Proximal Antecedents	Basic Traits	Proximal Consequences		Distal Consequences

Table 2: Frequency of experimental research in personality published in 2005 or 2006. * For JPSP, we have included all articles and then just the ones with personality in the abstract or as a keyword.

Journal	Total	Experimental Personality	% Experimental Personality	IQ	% IQ
EJP	68	0	0	2	3
JoP	125	7	6	1	1
JPSP	280	26	9	3	1
JPSP*	92	26	28	3	3
JRP	102	16	16	1	1
PaID	586	73	12	47	8
Total*	1161	148	71	13	6

genetic and physiological bases (Reuter, 2007) of personality.

Quality of measurement

> I often say that when you can measure what you are speaking about, and express it in numbers, you know something about it; but when you cannot measure it, when you cannot express it in numbers, your knowledge is of a meagre and unsatisfactory kind; it may be the beginning of knowledge, but you have scarcely in your thoughts advanced to the state of Science, whatever the matter may be. (Thomson, 1889-1891)

It is not surprising that our journals are heavily biased with correlational/observational studies given that most personality researchers receive more training in measurement techniques than in experimental design. Courses in psychometrics, Item Response Theory, Structural Equation Modeling, hierarchical linear or mixed effects models are without question important; good science requires good measurement. While experimental psychologists could benefit from more training in psychometrics, observational researchers need to better understand how measurement issues affect the theoretical inferences drawn from experiments. It is not just the poor benighted experimentalists who need to focus on the metric properties of their measures. Observationalists do as well.

As the quality of measurement improves, the ability to falsify hypotheses via disconfirmatory studies increases. Although measurement is invariably weak early in the process of theory building, as competing hypotheses are teased apart, improvements in measurement become an essential focus. The use of structural equation modeling, with its emphasis upon evaluating both the measurement and the structural components of the model can make a strong addition to our theoretical tool kit at this point. The measurement component of the model, by emphasizing multiple indicators for proposed constructs, and

evaluating the adequacy of the constructs to fit the covariances of the indicator variables forces us to specify models more precisely than has been done in the past.

What is sometimes overlooked in the quest for structural fits is the basic metric quality of our measures, particularly as interactions with situational manipulations are interpreted. For example, some interactions with experimental variables are likely due to measurement artifacts rather than interactions at the latent level. Non-linearities of the mapping between the latent construct and the observed indicator are tolerable only if the mere direction or magnitude of the effect are of interest. But, such non-linearities, when combined with experimental manipulations, can lead to interaction patterns at the observed score level that do not reflect interactions at the underlying latent construct level (Revelle, 2007).

Reliability. Experimentalists should recognize that the quality of measurement is vitally important. The number of participants has a direct impact upon the statistical power to detect an effect; however, it does not allow us to correctly estimate the magnitude of the effect. Although increasing sample size can compensate for the attenuation of effect sizes due to lower reliability, it is better to improve the reliability to properly estimate the strength of a relationship.

Validity. Clearly, reliability is not enough. Just as SEM forces us to focus on the measurement model, so does it force us to focus on the structural model relating the constructs. The incorporation of experimental techniques can provide essential clarity as well. By finding particular manipulations that affect one scale but not another we can resolve issues that can not be solved by psychometrics alone. Even with thousands of participants, Rafaeli and Revelle (2006) were unable to conclusively argue against the bipolar nature of happiness versus sadness, but by showing that these two affects respond differently to experimental manipulations of mood, the argument was much more compelling.

The integration of experimental and observational approaches

H. J. Eysenck (1997) summarized a career's worth of research in a brief article in an effort to integrate the two disciplines of scientific psychology. The articles in this special issue of the *The Journal of Personality* address how well we have progressed in the past ten years. Experimental techniques and findings have much to offer the field of personality as we move towards a stronger science by doing paradigmatic and programmatic research. This is important not just for those of us in personality, but for the entire field of psychology, for personality is the one subdiscipline of the field that requires a knowledge of the entire field and has the opportunity to provide integrative findings from genes to society.

References

Anderson, K. J., & Revelle, W. (1994). Impulsivity and time of day: Is rate of change in arousal a function of impulsivity? *Journal of Personality and Social Psychology, 67*(2), 334-344.

Anderson, K. J., Revelle, W., & Lynch, M. J. (1989). Caffeine, impulsivity, and memory scanning: A comparison of two explanations for the yerkes-dodson effect. *Motivation and Emotion*, *13*(1), 1-20.

Asendorf, J. B., Borkenau, P., Ostendorf, F., & Van Aken, M. A. (2001). Carving personality description at its joints: Confirmation of three replicable personality prototypes for both children and adults. *European Journal of Personality*, *15*(3), 169-198.

Bouchard, J., Thomas J. (2004). Genetic influence on human psychological traits: A survey. *Current Directions in Psychological Science*, *13*(4), 148-151.

Broadbent, D. (1971). *Decision and stress*. London: Academic Press.

Canli, T. (2004). Functional brain mapping of extraversion and neuroticism: learning from individual differences in emotion processing. *Journal of Personality*, *72*(6), 1105-1132.

Canli, T. (2006a). *Biology of personality and individual differences*. xv, 462 pp. New York, NY: Guilford Press Guilford Press.

Canli, T. (2006b). Genomic imaging of extraversion. In T. Canli (Ed.), *Biology of personality and individual differences* (p. 93-115). New York, NY: Guilford Press Publisher Information Guilford Press.

Corr, P. J. (2007). The reinforcement sensitivity theory. In P. J. Corr (Ed.), *The reinforcement sensitivity theory*. Cambridge: Cambridge University Press.

Cronbach, L. (1957). The two disciplines of scientific psychology. *American Psychologist*, *12*, 671-684.

Cronbach, L. (1975). Beyond the two disciplines of scientific psychology. *American Psychologist*, *30*, 116-127.

Davis, R. H. (2006). Strong inference: rationale or inspiration. *Perspectives in biology and medicine*, *49*(2), 238-250.

Doyle, A. (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.

Easterbrook, J. (1959). The effect of emotion on cue utilization and the organization of behavior. *Psychological Review*, *66*, 183-20.

Eysenck, H. J. (1957). *The dynamics of anxiety and hysteria: An experimental application of modern learning theory to psychiatry*. London: Routledge and Kegan Paul.

Eysenck, H. J. (1966). Personality and experimental psychology. *Bulletin of the British Psychological Society*, *19*, 1-28.

Eysenck, H. J. (1967). *The biological basis of personality*. Springfield: Thomas.

Eysenck, H. J. (1983). Is there a paradigm in personality-research. *Journal Of Research In Personality*, *17*(4), 369-397.

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*(6), 1224-1237.

Eysenck, H. J., & Eysenck, M. W. (1985). *Personality and individual differences: a natural science approach*. New York: Plenum Press.

Eysenck, H. J., & Eysenck, S. B. G. (1964). *Eysenck personality inventory*. San Diego, California: Educational and Industrial Testing Service.

Eysenck, H. J., & Levey, A. (1972). Conditioning, introversion-extraversion and the strength of the nervous system. In V. Nebylitsyn & J. Gray (Eds.), *Biological basis of individual behavior* (p. 206-220). New York: Academic Press.

Eysenck, M. W. (1979). Anxiety, learning, and memory: A reconceptualization. *Journal of Research in Personality*, *13*(4), 363-385.

Eysenck, M. W., & Calvo, M. (1998). Cognitive bias to internal sources of information in anxiety. *International Journal of Psychology*, *33*, 287-299.

Eysenck, M. W., & Mathews, A. (1987). Trait anxiety and cognition. In H. J. Eysenck & I. Martin (Eds.), *Theoretical foundations of behavior therapy* (p. 73-86). New York: Plenum.

Fleeson, W. (2004). Moving personality beyond the person-situation debate: The challenge and the opportunity of within-person variability. *Current Directions in Psychological Science*, *13*(2), 83-87.

Fleeson, W., Malanos, A. B., & Achille, N. M. (2002). An intraindividual process approach to the relationship between extraversion and positive affect: Is acting extraverted as "good" as being extraverted? *Journal of Personality and Social Psychology*, *83*(6), 1409-1422.

Galton, F. (1892). *Hereditary genius: an inquiry into its lawas and consequences* (2nd ed.). London: Macmillan and Co.

Gangestad, S., & Snyder, M. (1985). "to carve nature at its joints": On the existence of discrete classes in personality. *Psychological Review*, *92*(3), 317-349.

Glymour, C. (2003). Learning, prediction and causal bayes nets. *Trends in Cognitive Sciences*, *7*(1), 43-48.

Goldberg, L. R. (1990). An alternative "description of personality": The big-five factor structure. *Journal of Personality and Social Psychology*, *59*(6), 1216-1229.

Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Cloninger, C. R., et al. (2006). The international personality item pool and the future of public-domain personality measures. *Journal of Research in Personality*, *40*(1), 84-96.

Gray, J. A. (1981). A critique of eysenck's theory of personality. In H. J. Eysenck (Ed.), (p. 246-277). Berlin: Springer.

Gray, J. A. (1982). *Neuropsychological theory of anxiety: An investigation of the septal-hippocampal system*. Cambridge: Cambridge University Press.

Gray, J. A., & McNaughton, N. (2000). *The neuropsychology of anxiety: An enquiry into the functions of the septo-hippocampal system* (2nd ed.). Oxford: Oxford University Press.

Harlow, L. A., Mulaik, S. A., & Steiger, J. H. (Eds.). (1997). *What if there were no significance tests?* Mahwah, NJ: Lawrence Erlbaum Associates.

Howarth, E., & Eysenck, H. J. (1968). Extraversion, arousal, and paired-associate recall. *Journal of Experimental Research in Personality*, *3*(2), 114-116.

Hull, C. L. (1952). *A behavior system*. New Haven: Yale University Press.

Hunt, E. (1995). *Will we be smart enough? a cognitive analysis of the coming workforce*. New York: Russell Sage Foundation.

Kluckhohn, C., & Murray, H. A. (1953). Personality formation: The determinants. In C. Kluckhohn, H. A. Murray, & D. M. Schneider (Eds.), *Personality in nature, society, and culture* (p. 53-67). New York: Alfred A. Knopf.

Kosslyn, S. M., Cacioppo, J. T., Davidson, R. J., Hugdahl, K., Lovallo, W. R., Spiegel, D., et al. (2002). Bridging psychology and biology: the analysis of individuals in groups. *American Psychologist*, *57*(5), 341-351.

Kuhn, T. (1970). *The structure of scientific revolution* (2nd ed.). Chicago: University of Chicago Press.

Leon, M. R., & Revelle, W. (1985). Effects of anxiety on analogical reasoning: A test of three theoretical models. *Journal of Personality and Social Psychology*, *49*(5), 1302-1315.

Lesch, K.-P., & Canli, T. (2006). 5-ht-sub(1a) receptor and anxiety-related traits: Pharmacology,

genetics, and imaging. In T. Canli (Ed.), *Biology of personality and individual differences* (p. 273-294). New York, NY: Guilford Press Guilford Press.

Magnusson, D., & Endler, N. S. (1977). *Personality at the crossroads: current issues in interactional psychology.* Hillsdale, N.J.: Erlbaum.

Mandler, G., & Sarason, S. (1952). A study of anxiety and learning. *Journal of Abnormal and Social Psychology*, *47*, 166-173.

McCrae, R. R., & Costa, J., Paul T. (1999). A five-factor theory of personality. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality: Theory and research (2nd ed.)* (p. 139-153). New York, NY: Guilford Press Guilford Press.

McGue, M., & Bouchard, J., Thomas J. (1998). Genetic and environmental influences on human behavioral differences. *Annual Review of Neuroscience. Vol 21 1998, 1-24. Annual Reviews. Annual Reviews.*

Navon, D. (1977). Forest before trees: The precedence of global features in visual perception. *Cognitive Psychology*, *9*(3), 353-383.

Norman, W. T. (1963). Toward an adequate taxonomy of personality attributes: replicated factors structure in peer nomination personality ratings. *Journal of Abnormal and Social Psychology*, *66*(June), 574-583.

Norman, W. T. (1969). "to see ourseles as ithers see us]": Relations among self-perceptions, peer-perceptions, and expected peer-perceptions of personality attributes. *Multivariate Behavioral Research*, *4*(4), 417-443.

O'Donohue, W., & Buchanan, J. A. (2001). The weakness of strong inference. *Behavior and Philosophy*, *29*, 1-10.

Ortony, A., Norman, D. A., & Revelle, W. (2005). Effective functioning: A three level model of affect, motivation, cognition, and behavior. In J. Fellous & M. Arbib (Eds.), *Who needs emotions? the brain meets the machine.* (p. 173-202). New York: Oxford Univeristy Press.

Platt, J. R. (1964). Strong inference. *Science*, *146*, 347-353.

Popper, K. (1935). *Logik der forschung.* Vienna: Springer.

Rafaeli, E., & Revelle, W. (2006). A premature consensus: Are happiness and sadness truly opposite affects? *Motivation and Emotion*, *30*(1), 1-12.

Reuter, M. (2007). Neuroimagining and genetics. In P. J. Corr (Ed.), *The reinforcement sensitivity theory.* Cambridge: Cambridge University Press.

Revelle, W. (1995). Personality processes. *Annual Review of Psychology*, *46*.

Revelle, W. (2007). Experimental approaches to the study of personality. In B. Robins, C. Fraley, & R. Krueger (Eds.), *Personality research methods.* New York, New York: Guilford.

Revelle, W., Amaral, P., & Turriff, S. (1976). Introversion-extraversion, time stress, and caffeine: the effect on verbal performance. *Science*, *192*, 149-150.

Revelle, W., & Anderson, K. J. (1992). Models for the testing of theory. In A. Gale & M. Eysenck (Eds.), *Handbook of individual differences: Biological perspectives.* Chichester, England: John Wiley and Sons.

Revelle, W., Humphreys, M. S., Simon, L., & Gilliland, K. (1980). Interactive effect of personality, time of day, and caffeine - test of the arousal model. *Journal of Experimental Psychology General*, *0109*, 1-31.

Revelle, W., & Loftus, D. A. (1990). Individual-differences and arousal - implications for the study of mood and memory. *Cognition and Emotion*, *4*(3), 209-237.

Rocklin, T., & Revelle, W. (1981). The measurement of extraversion: A comparison of the eysenck

personality inventory and the eysenck personality questionnaire. *British Journal of Social Psychology*, *20*(4), 279-284.

Scheines, R., Spirtes, P., Glymour, C., Meek, C., & Richardson, T. (1998). The tetrad project: Constraint based aids to causal model specification. *Multivariate Behavioral Research*, *33*(1), 65-117.

Shigehisa, T., & Symons, J. R. (1973). Effect of intensity of visual stimulation on auditory sensitivity in relation to personality. *British Journal of Psychology*, *64*(2), 205-213.

Smillie, L. D., Pickering, A. D., & Jackson, C. J. (2006). The new reinforcement sensitivity theory: implications for personality measurement. *Personality and Sociality Psychology Review*, *10*(4), 320-335.

Spence, K., Farber, I., & McFann, H. (1956). The relation of anxiety (drive) level to performance in competitional and non-competitional paired-associates learning. *Journal of Experimental Psychology*, *52*, 296-305.

Thomson, W. (1889-1891). *Popular lectures and address*. London: Macmilland and Company.

Underwood, B. J. (1975). Individual differences as a crucible in theory construction. *American Psychologist*, *30*, 128-134.

Vale, J., & Vale, C. (1969). Individual differences and general laws in psychology: a reconciliation. *American Psychologist*, *24*, 1093-1108.

Waller, D., Knapp, D., & Hunt, E. (2001). Spatial representations of virtual mazes: The role of visual fidelity and individual differences. *Human Factors: The Journal of the Human Factors and Ergonomics Societ*, *43*(1), 147-158.

Wine, J. (1971). Test anxiety and direction of attention. *Psychological Bulletin*, *76*(2), 92-104.

Wundt, W. (1874/1999). *Grundzge der physiologischen psychologie*. Leipzig/Bristol: Wilhelm Engelmann/Thoemmes Press.

Wundt, W. (1904). *Principles of physiological psychology* (Translated from the Fifth German Edition (1902) ed.). London: Swan Sonnenschein.

Yerkes, R., & Dodson, J. (1908). The relation of strength of stimuli to rapidity of habit-information. *Journal of Comparative Neurology and Psychology*, *18*, 459-482.

Zachar, P. (2000). Psychiatric disorders are not natural kinds. *Philosophy, Psychiatry, and Psychology*, *7*(3), 167-182.

Zinbarg, R. E., & Revelle, W. (1989). Personality and conditioning: A test of four models. *Journal of Personality and Social Psychology*, *57*(2), 301-314.

Zinbarg, R. E., & Yoon, K. L. (2007). Rst and clinical disorders: Anxiety and depression. In P. J. Corr (Ed.), *The reinforcement sensitivity theory.* Cambridge: Cambridge University Press.