Integrating Experimental and Observational Personality Research—The Contributions of Hans Eysenck

William Revelle and Katherine Oehlberg
Northwestern University

ABSTRACT A fundamental aspect of Hans 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. Fifty 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? Should personality researchers use experimental techniques? Do experimental techniques allow us to tease out causality, and are we communicating the advantages of combining experimental with multivariate correlational techniques? We review the progress made since Cronbach's and Eysenck's original papers and suggest that although it is still uncommon to find experimental studies of personality, psychology would benefit from the joint use of correlational and experimental approaches.

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 & M. W. 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 & M. W. Eysenck, 1985). He pioneered the use of the most recent developments in psychological measurement

We gratefully acknowledge the help of Josh Wilt, Katherina Hauner, three anonymous reviewers, and the editors for the suggestions they made to improve this manuscript.

Correspondence concerning this article may be sent to William Revelle. E-mail: revelle@northwestern.edu.

Journal of Personality 76:6, December 2008

DOI: 10.1111/j.1467-6494.2008.00526.x

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 psychology[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.

Personality and Experimental Psychology

Ever since Wundt introduced experiments into psychology (Wundt, 1874/1999; Wundt & Titchener, 1904) and Galton (1892) studied individual differences in genius, there has been a persistent 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" 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 preparadigmatic; 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 would be

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-neuropsychology for what used to be the domain of "experimental" psychology.

crucial for the field to develop a paradigm within which progress could 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 generalized linear model, it is better to consider the distinction to be between *experimental* and *observational* methods rather than *experimental* and *correlational* analysis.

Perhaps Hans 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 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.

The Current State of Integration of Experimental Methods and Personality Research

In order to evaluate the current level of usage of experimental techniques in research on individual differences, we analyzed all 2005 and 2006 volumes of the five major personality journals—*European Journal of Personality* (EJP), *Journal of Personality* (JoP), the *Journal of Personality and Social Psychology* (JPSP),[2] *Journal of Research*

2. 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
Frequency of Experimental Research in Personality Published in 2005 or 2006

Journal	Total	Experimental Personality	% Experimental Personality
EJP	68	0	0
JoP	125	7	6
JPSP	280	26	9
JPSP*	92	26	28
JRP	102	16	16
PaID	586	73	12
Total*	1161	122	11

*For JPSP, we have included all articles and then just the ones with personality in the abstract or as a keyword.

in Personality (JRP), *and the Personality and Individual Differences* (PaID)—by performing computer searches for the use of the words *random, experiment, experimental, condition,* or *assigned* (Table 1).

Perhaps the most obvious finding from this classification is the infrequency of experimental work published in the last 2 years in any of the journals. Zero percent of the articles in the EJP, $<6\%$ of the articles in JoP, $\approx 12\%$ 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%.

In the journal that had the most experimental studies (PaID), most were tests of hypotheses derived from Reinforcement Sensitivity Theory (Corr, 2008; 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 holistic-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 uncommon 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. In the rest of this

article we address why we believe that it remains important to unify these two approaches.

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 the fact 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.

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. If one is concerned with measuring reaction time (RT), differences associated with semantic priming, or perceptual interference in a global-local task, that participants differ in ability, age, arousal, and motivation, all large sources of variance in reaction time, is irrelevant. 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 controls increases the power of the design enough to get reliable between condition effects.

Unfortunately for experimentalists, the use of within-subject designs in itself is not able to obviate entirely the need to attend to individual differences because 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 crossover 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 his landmark 1985 text (H. J. Eysenck & M. W. 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. Ignoring individual differences would have diminished the cross modal effect. 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 (W. Revelle & Loftus, 1990). Once again, by ignoring the individual differences in introversion-extraversion, the consistency of the arousal by recall interval interaction where arousal is either manipulated or is the result of stable between subject differences (introversion-extraversion) would have been missed. In another experiment, Hans Eysenck and Levey (1972) showed that eye-blink conditioning was better for more introverted participants under weak UCS conditions but better for more extraverted subjects under strong UCS conditions. This paper, in addition to clarifying the effect of stimulation on eye-blink conditioning, also demonstrated the power of experiments to tease out more subtle interactions: the effect of enhanced conditioning was much larger for the impulsivity than the sociability component of extraversion as measured by the

Eysenck Personality Inventory (H. J. Eysenck & S. B. G. Eysenck, 1964).

Inspired by these demonstrations of the importance of considering individual differences in the context of experimental manipulations, we showed that the complex cognitive performance of more introverted subjects is hindered, but that of more extraverted subjects is facilitated, by caffeine and time stress (W. Revelle, Amaral, & Turriff, 1976), with no main effects of either personality or caffeine and time stress on performance. Even with the enormous sample sizes associated with studies done by the Educational Testing Service, prior experimental work had failed to show any effect of stress upon complex reasoning tasks (the Graduate Record Exam) similar to the ones we used. This was, of course, because our effects were crossover interactions, with no main effects of stress or of personality. Follow-up studies showed that this effect was even more complicated and showed a systematic three-way crossover interaction between personality (impulsivity), caffeine, and time of day with no main effects of either personality, drug, or time of day (W. Revelle, Humphreys, Simon, & Gilliland, 1980). Consistent with the earlier findings of H. J. Eysenck and Levey (1972), impulsivity rather than sociability was the component of extraversion with the most systematic effect. The complex, but systematic pattern of person $\times$ situation $\times$ task interactions shown in these and subsequent studies provided strong support for the need to integrate individual differences into more standard cognitive paradigms and theory (Humphreys & W. Revelle, 1984).

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 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). In a subsequent study, Underwood used the power of individual differences by examining the correlations and factor structure of a number of measures of episodic and semantic

memory to distinguish beyond attributes of memory and of response (Underwood, Boruch, & Malmi, 1978). 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, 1983, 1995; Waller, Knapp, & Hunt, 2001).

The importance of individual differences in physiological responses in brain imaging and genetic paradigms has become increasingly recognized in the past few years (Canli, Sivers, Whitfield, Gotlib, & Gabrieli, 2002). 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; Depue, 2006; Lesch & Canli, 2006) in a recent volume on the biological basis of personality (Canli, 2006a) and others also make this point very well. Examining individual differences in physiological reactions to situations. Depue (2006) discusses how it is possible to test biological models of Extraversion and dopamine sensitivity by manipulating affect using film cues with and without opiate antagonists for participants who differ in trait affiliation. This work goes beyond demonstrations of differences in brain structure or functioning and tests individual differences in neural responses to environmental cues.

Although not typically considered experimental psychology per se, randomized clinical trials of psychological interventions are

another area in which concern for individual differences has led to significant theoretical advances in an otherwise experimental setting. For example, attention to individual differences over time in the response to cognitive-behavioral therapy (CBT) for depression led Tang and DeRubeis (1999) to discover the important treatment phenomenon of "sudden gains," in which some (but not all) patients undergoing a psychological treatment will experience dramatic improvements in outcome measures within a very short period of time, rather than a gradual, constant improvement. There had been a longstanding discrepancy between anecdotal clinical experience (in which clinicians often reported seeing major improvements in a patient in between two consecutive sessions while change was otherwise very slow), and the analyses of outcome studies, which indicated that groups of patients improved gradually over the course of treatment. Rather than analyzing the pattern of symptom improvement of whole treatment groups, Tang and DeRubeis instead looked at the patterns of change in individual patients and discovered that more than a third experienced these rapid bursts of improvement. This finding turned out to have significant consequences, particularly with respect to the study of the active elements of psychological interventions: By identifying the therapeutic elements introduced immediately prior to such sudden gains, it is possible to understand better what makes a particular treatment effective. Moreover, it has been found that those who experience sudden gains are less likely to relapse following treatment (Tang, DeRubeis, Hollon, Amsterdam, & Shelton, 2007), suggesting that closer study of the individual differences (in addition to the treatment elements) responsible for sudden gains is critical for understanding the causal mechanisms of psychotherapy and therefore for designing more effective and efficacious treatments.

Importance of Experimental Methods for Personality Theory

Hans Eysenck (1997) answered the question of whether personality research could be paradigmatic with an optimistic "yes" and emphasized the importance of experimental techniques and theory for continued progress in the field:

> [P]urely 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. (Eysenck, 1997, p. 1234)

Eysenck (1997) stressed 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. It should be noted that by emphasizing the need for experimentation in science, he over looked the substantial progress in observational sciences such as astronomy, meteorology, or oceanography. However, the lack of the ability to do experiments in these fields has delayed the acceptance of hypotheses such as the anthropogenic causes of global warming (R. Revelle & Seuss, 1957).

Eysenck 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 & M. W. Eysenck, 1985) and molecular genetic (H. J. Eysenck, 1997) findings.[3]

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. But how can we do experiments with personality? We cannot assign a person to the female condition, to the extravert condition, or the intelligent condition; these traits are

3. In his last talk to the International Society for the study of Individual Difference in 1997, a few months before he died, Hans Eysenck said that if he were younger, he would try to learn molecular genetics.

stable between-individual differences that are not subject to random assignment. However, 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, Hans Eysenck and Levey's (1972) 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. This result called into question the simple notion that introverts were simply more readily conditioned and thus more likely to be socialized to conventional rules.

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 Hans Eysenck (1967). This prediction followed from his hypotheses that introverts are chronically more aroused than are extraverts and (based upon a generalization of Yerkes & Dodson, 1908) that arousal has an inverted U relationship to performance. This interaction with caffeine allows us to reject the hypothesis that introverts are simply smarter than extraverts, for while their scores are higher in the relaxed, placebo condition, their scores were lower in the time-stress and caffeine condition. 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 (W. 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 & W. Revelle, 1981). Interaction patterns constrain our generalizations about personality variables and force us to specify the particular conditions in which trait X is related to phenomena Y. By constraining the effect to particular conditions we are actually strengthening our causal models. For example, through the use of random assignment of "morning types" and "evening types" to morning or evening conditions, Bodenhausen (1990) was able to show that the tendency to stereotype was a judgmental heuristic associated with a lack of cognitive resources rather than a broader trait variable.

In addition to limiting the extent of inferences about personality (W. 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 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.

Our emphasis upon the theoretical power of interactions might remind some of our readers about the person $\times$ situation controversies of the 1970s. This is not our intent. We view the debate about the relative importance of the person or the situation in predicting behavior (Magnusson & Endler, 1977) as an unfortunate detour for much of American personality research in the 1970s and 1980s. 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 & M. W. Eysenck, 1985). That person variables interacted with situational and task variables was so obvious to those working in the Eysenckian tradition that it hardly needed to be elaborated. Interactions were just assumed to be a natural part of the process of theory building and testing.

What was missing from the personality $\times$ 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. Unfortunately, this still seems to be the case. Not enough effort has been applied to the problem of what is a situation and how situations differ. As a methodological advance, it would probably be useful to think of situations as we think of items in the ability domain: To what extent is the patterning of responses across situations different than the patterning of ability items across ability domains? Just as some people are better at spatial than verbal material, are some people more sociable at parties and others in small groups? Is it possible to generalize Item Response

Theory techniques from psychometrics to the study of social situations? That is, are some situations more difficult than others for eliciting particular behaviors? Can we scale situations in terms of the effect they have upon individuals?

A serious challenge for correlational examinations of the effects of situations on individuals is that people do not randomly choose situations. Extraverts, for instance, seek out lively parties and tend to experience more positive affect. But does the lively party induce the positive affect, does being in a good mood increase the likelihood of attending a party, do people with positive affect make the party lively, or are extraverts (who tend to go to parties) just more likely to have positive affect? By experimentally assigning people to situations or to acting in a particular way, Fleeson and his colleagues have been able to tease these phenomena apart and to show that at least some of the affective state—personality trait association—is due to the differences in behavior associated with trait extraversion (Fleeson, 2004; Fleeson, Malanos, & Achille, 2002; McNiel & Fleeson, 2006).

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 (Asendorf, Borkenau, Ostendorf, & Van Aken, 2001; Gangestad & Snyder, 1985, 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. Canli's demonstration (2004) 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 neuroticism (for negative pictures) makes an important distinction between positive and negative affect and the personality dimensions representing sensitivity to environmental cues. It is difficult to imagine the self-report item that could make this distinction as well. Differential patterns of responding to different situations can be used as a way to discriminate between traits far more powerfully than can factor analytic techniques. (Consider the different patterning of responses to caffeine or time of day for sociability and impulsivity [W. Revelle et al., 1980] compared to the analysis of scales for impulsivity and sociability [Rocklin & W. Revelle, 1981].)

Theoretical Inference

The relative lack of experimental studies in personality research reflects that much of the field is descriptive rather than causal. This is, of course, the separation of the two disciplines that Cronbach and Eysenck wanted to unify. The importance of detecting relationships as found by correlations is not to be denied, but 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 (W. Revelle, 2007; W. 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, Hans 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. Two of these studies were direct tests; the remaining three were tests of derivative hypotheses. A direct test of Eysenck's hypothesis that introverts are always more aroused than extraverts examined the effect of time of day on the caffeine by introversion interaction (W. Revelle et al., 1980); a second direct test compared predictions from Eysenck versus Gray (1982) with respect to the role of cues for reward and punishment (Zinbarg & W. Revelle, 1989). A third study tested competing models of the effects of anxiety on cognitive performance partially derived from H. J. and M.W. Eysenck (1985; Leon & W. Revelle, 1985), a fourth compared two hypotheses about the relationship between impulsivity and the decay of arousal (Anderson and Revelle & W. Revelle, 1994), and a fifth examined competing explanations for the Yerkes-Dodson "Law" (Anderson, W. Revelle, & Lynch, 1989; Yerkes & Dodson, 1908).

Hans 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. In an early study, we reported evidence supporting (compatible with) this hypothesis by showing that when taking a complex reasoning test similar to the Graduate Record Examination the performance of introverts was hindered but that of extraverts was facilitated by the combination of time stress and caffeine (W. Revelle et al., 1976).

But Eysenck (1967) 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 (Blake, 1967). He did not seem to notice that this latter finding was incompatible with his basic hypothesis. For if the introvert-extravert difference was one of phase rather than level of arousal, it would be difficult to claim this led to differences in stimulation seeking. 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* (W. Revelle et al., 1980). This result would be consistent with the hypothesis of greater arousal for introverts if the detrimental effect of caffeine on the performance of introverts in the morning were 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. These results were consistent with the hypothesis that arousal varied diurnally and that the introvert-extravert difference in arousal was one of phase rather than of level. Such a finding was, however, at complete odds with the arousal-seeking explanation of extraversio for it would imply 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. Hans Eysenck's theory of personality (H. J. Eysenck, 1967; H. J. Eysenck & M. W. 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 Five (Goldberg, 1990; McCrae & Costa, 1999), and by attempting to attribute cause to genetic predispositions and phys-

Table 2
Theoretical Differences Between Eysenck and Gray Relevant to the Zinbarg and Colleagues' Studies (1989, 1998)

Phenomenon	Eysenck	Gray
Conditioning		
	Introverts ++	0
Appetitive Conditioning		
	Introverts ++	*Introvert* −
	Impulsive −	Impulsive ++
Aversive Conditioning		
	Introverts ++	Introverts +
	Anxious +	*Anxious* ++

Note. Italicized entries reflect derived hypotheses.

iological mechanisms, it had greater depth as well. A useful heuristic to compare alternative theories is to consider phenomena and theories as rows and columns of a matrix in which the cells represent whether the theory predicts a positive or a negative relationship or does not address a particular phenomenon. While many cells in the matrix will be empty because theories may be complementary or speak to different phenomena, some rows will have identical entries across all the columns since 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 phenomenon 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 testing is a row in which different theories make different predictions. Examples of competing theoretical predictions include those of Hans 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 (Corr et al., 1997; Zinbarg & Mohlman, 1998; Zinbarg & W. Revelle, 1989). We compare a few of the contrasting predictions made for overall differences in conditioning, and for appetitive and aversive conditioning (Table 2). Intriguingly enough, the Zinbarg and W. Revelle (1989) 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, 2008). A more thorough comparison of the breadth and depth of the Eysenck and Gray models is seen in the work of Matthews and Gilliland (1999).

Another example of using the matrix of competing theories with multiple phenomena was the examination of anxiety-induced decrements on cognitive performance (Leon & W. 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 W. Revelle (1985) found mixed support for the distraction hypothesis and no support for the effects of anxiety on working memory.

Although it would be reasonable to apply correlational techniques to compare competing predictions such as seen in Table 2, such studies tend not to be done. The emphasis in most correlational studies is descriptive and confirmatory—that is, to demonstrate a non-zero correlation between a trait and an outcome—and is less likely to be disconfirmatory or a comparison of two correlations.

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. Hans Eysenck (1997) recognized that not all results would 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 (W. Revelle et al., 1980) were ''a dagger in the heart of Eysenckian theory,'' to Hans Eysenck and M. W. 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 (W. 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 (Rocklin & W. Revelle, 1981). Psychometric refinement of the scales and some modest modification of the theory led to the revised model (H. J. Eysenck & M. W. 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 (SEM)

On the surface, the ability to test alternative structural equation models may seem very similar 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 do not show causality. Consider the observation that yellow fingers, yellow teeth, and bad breath at Time 1 are predictors of subsequent lung cancer at Time 2. Even if a structural model fit these covariances perfectly, we should not conclude that better dental hygiene would protect from lung cancer. "No analysis void of experimental data can possibly defend causal assumptions" (Pearl, 2003, p. 289). For, as compelling as the path diagrams of an SEM are, "It is important to note that, in path diagrams, causal assumptions are encoded not in the links but, rather, in the missing links" (Pearl, 2003, p. 293).

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 can be analyzed as if they were experimental. Structural equation modeling of these "experiments of nature" allows one to tease out genetic effects that would otherwise be untestable. That very com-

plex 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. That is, it does not follow that just because something has a high heritability that it reflects a single biological system (Revelle, 1995). 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 noncognitive traits (Bouchard, 2004; McGue & Bouchard, 1998).

Recommendations

Hans 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 than repeatedly labeling old work as "new" and mistaking novelty for progress. If the success of the Big Five and the Five-Factor Model (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, 2008; Gray & McNaughton, 2000; W. Revelle, 2008; 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 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 meager 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 (W. Revelle, 2007).

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 toolkit 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. Nonlinearities of the mapping between the latent construct and the observed indicator are tolerable only if the mere direction or magnitude of the effect is of interest. But such nonlinearities, 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 (W. 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 W. Revelle (2006) were unable to argue conclusively 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

Hans 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 *Journal of Personality* address how well we have progressed in the past 10 years. Experimental techniques and findings have much to offer the field of personality as we move toward 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 because 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.

We reiterate Eysenck's point that personality psychologists should take advantage of the theories of cognitive, social, and neuropsychology. We should also borrow the best of experimental methodologies in order to pit causal theories against each other. At the same

time, personality psychologists should communicate to their cognitive, social, and physiological colleagues how to take advantage of the power of an analysis of individual differences.

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. G. (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.

Blake, M. J. (1967). Relationship between circadian rhythm of body temperature and introversion-extraversion. *Nature*, **215** (5103), 896–897.

Bodenhausen, G. V. (1990, Sep). Stereotypes as judgmental heuristics: Evidence of circadian variations in discrimination. *Psychological Science*, **1** (5), 319–322.

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

Broadbent, D. E. (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*. New York: Guilford Press.

Canli, T. (2006b). Genomic imaging of extraversion. In T. Canli (Ed.), *Biology of personality and individual differences* (pp. 93–115). New York: Guilford Press.

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*, **296** (5576), 2191.

Corr, P. J. (2008). Reinforcement sensitivity theory (RST). In P. J. Corr (Ed.), *The reinforcement sensitivity theory*. Cambridge: Cambridge University Press.

Corr, P. J., Pickering, A. D., & Gray, J. A. (1997). Personality, punishment, and procedural learning: A test of J. A. Gray's anxiety theory. *Journal of Personality and Social Psychology*, **73** (2), 337–344.

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*, **30**, 116–127.

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

Depue, R. A. (2006). Interpersonal behavior and the structure of personality: Neurobehavioral foundation of agentic extraversion and affiliation. In T. Canli (Ed.), *Biology of personality and individual differences* (pp. 60–92). New York: Guilford Press.

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.

Easterbrook, J. A. (1959). The effect of emotion on cue utilization and the organization of behavior. *Psychological Review*, **66**, 183–120.

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, IL: 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**, 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, CA: Educational and Industrial Testing Service.

Eysenck, H. J., & Levey, A. (1972). Conditioning, introversion-extraversion and the strength of the nervous system. In V. D. Nebylitsyn & J. A. Gray (Eds.), *Biological basis of individual behavior* (pp. 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. G. (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* (pp. 73–86). New York: Plenum.

Fleeson, W. (2004, Apr). 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, Dec). 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 laws and consequences* (2nd ed.). London: Macmillan and Co.

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

Glymour, C. (2003, Jan). 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.), *A Model for Personality* (pp. 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, CT: Yale University Press.

Humphreys, M. S., & Revelle, W. (1984). Personality, motivation, and performance: A theory of the relationship between individual differences and information processing. *Psychological Review*, **91** (2), 153–184.

Hunt, E. (1983). On the nature of intelligence. *Science*, **219**, 141–146.

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

Kosslyn, S. M., Cacioppo, J. T., Davidson, R. J., Hugdahl, K., Lovallo, W. R., Spiegel, D., et al. (2002, May). 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* (pp. 273–294). New York: Guilford Press.

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

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

Matthews, G., & Gilliland, K. (1999). The personality theories of H. J. Eysenck and J. A. Gray: A comparative review. *Personality and Individual Differences*, **26** (4), 583–626.

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

McGue, M., & Bouchard, T. J., Jr. (1998). Genetic and environmental influences on human behavioral differences. *Annual Review of Neuroscience*, **21**, 1–24.

McNiel, J. M., & Fleeson, W. (2006). The causal effects of extraversion on positive affect and neuroticism on negative affect: Manipulating state extraversion and state neuroticism in an experimental approach. *Journal of Research in Personality*, **40** (5), 529–550.

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** (6), 574–583.

Norman, W. T. (1969). To see ourselves as others 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.

Pearl, J. (2003). Statistics and causal inference: A review. *Test Journal*, **12** (2), 281–345.

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

Popper, K. (1935). *Logik der forschung* [The Logic of Scientific Discovery]. Vienna: Springer.

Rafaeli, E., & Revelle, W. (2006, Mar). 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, R., & Seuss, H. E. (1957). Carbon dioxide exchange between atmosphere and ocean and the question of an increase of atmospheric (CO2) during the past decades. *Tellus*, **9** (1), 18–27.

Revelle, W. (1995). Personality processes. *Annual Review of Psychology*, **46**, 295–328.

Revelle, W. (2007). Experimental approaches to the study of personality. In B. Robins, C. Fraley, & R. Krueger (Eds.), *Handbook of research methods in personality psychology* (pp. 37–61). New York: Guilford.

Revelle, W. (2008). The contribution of reinforcement sensitivity theory to personality theory. In P. J. Corr (Ed.), *The reinforcement sensitivity theory* (pp. 508–527). Cambridge: Cambridge University Press.

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. W. Eysenck (Eds.), *Handbook of individual differences: Biological perspectives*. Chichester, UK: 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*, **109**, 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, May). 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. W., Farber, I. E., & McFann, H. 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.

Tang, T. Z., & DeRubeis, R. J. (1999). Sudden gains and critical sessions in cognitive-behavioral therapy for depression. *Journal of Consulting and Clinical Psychology*, **67** (6), 894–904.

Tang, T. Z., DeRubeis, R. J., Hollon, S. D., Amsterdam, J., & Shelton, R. (2007). Sudden gains in cognitive therapy of depression and depression relapse/recurrence. *Journal of Consulting and Clinical Psychology*, **75** (3), 404–408.

Thomson, W. (1889-1891). *Popular lectures and addresses* (3 vols). London: Macmillan and Company.

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

Underwood, B. J., Boruch, R. F., & Malmi, R. A. (1978). Composition of episodic memory. *Journal of Experimental Psychology: General*, **107** (4), 393–419.

Vale, J. R., & Vale, C. A. (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 Society*, **43** (1), 147–158.

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

Wundt, W. (1874/1999). *Grundzüge der physiologischen psychologie* [Principals of physiological psychology]. Leipzig/Bristol: Wilhelm Engelmann/Thoemmes Press.

Wundt, W. (1904). *Principles of physiological psychology* (Titchener, E. B. Trans., Translated from the 5th German ed. [1902]) London: Swan Sonnenschein. (Original work: *Grundzüge der physiologischen Psychologie*).

Yerkes, R. M., & Dodson, J. D. (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., & Mohlman, J. (1998). Individual differences in the acquisition of affectively valenced associations. *Journal of Personality and Social Psychology*, **74** (4), 1024–1040.

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.