

Available online at www.sciencedirect.com
SCIENCE DIRECT®
International Journal of Psychophysiology xx (2006) xxx – xxx
INTERNATIONAL JOURNAL OF PSYCHOPHYSIOLOGY
www.elsevier.com/locate/ijpsycho

The role of theory in the psychophysiology of personality: From Ivan Pavlov to Jeffrey Gray

Philip J. Corr *, Adam M. Perkins

Department of Psychology, University of Wales Swansea, Singleton Park, Swansea SA2 8PP, UK

Received 16 June 2005; received in revised form 24 October 2005; accepted 5 January 2006

Abstract

Psychophysiological approaches to personality have made significant progress in recent years, partly as a spin-off of technological innovation (e.g., functional neuroimaging) and partly as a result of an emerging theoretical consensus regarding the structure and biology of basic processes. In this field, Jeffrey Gray's influential psychophysiological theory of personality – now widely known as Reinforcement Sensitivity Theory (RST) – owes much to Pavlov, who devoted a large proportion of his later life to personality differences and their implications for psychiatry. In this article, we trace the influence of Pavlov on Hans Eysenck's and Jeffrey Gray's work, and then provide a brief description of RST in order to highlight some of the central problems – as well as some tentative solutions – in the psychophysiology of personality. Specifically, the importance of theory in personality research is stressed by the contrast of Gray's theoretically driven model with less fertile atheoretical (i.e., exploratory–inductive) approaches. The fecundity of RST, which has been in continual development over a period of thirty years, is discussed in the light of Karl Popper's views on the nature of science, especially the formulation of the 'problem situation', which sets up the theoretical and operational conditions under which hypotheses may be challenged and tested to destruction. In this respect, we see the truth of Lewin's [Lewin, K., 1951. Field theory in social science: selected theoretical papers. In: Cartwright, D., (Ed.). Harper & Row, New York] famous phrase, "There is nothing so practical as a good theory".

Keywords: Personality; Psychophysiology; Theory; Fear; Anxiety; Pavlov; Gray; Reinforcement

1. Introduction

Personality research has been likened to the Cinderella of psychology, with its scientific potential thwarted by a number of factors, including statistical indeterminism based on atheoretical data (Corr, 2004). This unsatisfactory state of affairs has now changed with the emergence of a neuroscience of personality, which in large measure has been a fortunate offspring of the opportunities afforded by new psychophysiological technologies (e.g., functional neuroimaging), as well as significant theoretical developments. In this latter respect, the work of Jeffrey Gray has been especially important.[1]

Gray's neuropsychological theory of personality – now known as *Reinforcement Sensitivity Theory* (RST; Pickering et al., 1995) – has been continually refined over the past thirty years (e.g., Gray, 1970, 1975, 1982, 1987; Gray and McNaughton, 2000; for a summary, see McNaughton and Corr, 2004). As shall become evident in this paper, the success of Gray's approach owes much to the scientific legacy of Ivan Pavlov (e.g., Gray, 1964), specifically his work on conditioning, neurophysiology, personality and psychopathology. These themes dominated Gray's scientific life, as they had Pavlov's later scientific career — today Pavlov is often remembered for little more than his work on the conditioned reflex, but this was only part of his much broader scientific interests.

Following in the footsteps of Hans Eysenck – who was also inspired by Pavlov in his theory of personality (Eysenck, 1957) – Gray advocated two complementary approaches: the *conceptual nervous system* (cns), and the *central nervous system* (CNS). Like Pavlov, Gray formulated cns components to personality (learning theory; see Gray, 1975) and located these

* Corresponding author. Tel.: +44 1792 205678x5081.
E-mail address: p.j.j.corr@swan.ac.uk (P.J. Corr).

[1] In this paper we focus on Gray's theory as an example of the importance of theory in psychophysiology, but we acknowledge the importance of other psychologists who have worked in the general field of personality inspired by Pavlov (e.g., Brebner, Claridge, Rusalov, Strelau, Zuckerman, and especially Robinson's approach which retains close links with Pavlov's theories — see this Special Issue).

0167-8760/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.ijpsycho.2006.01.005

cns components in the brain systems (CNS) underlying behavior. As noted by Gray (1972), save Cartesian ghosts in the machine, these two levels of explanation *must* be compatible. This Janus-faced strategy hallmarks the *neuropsychology* of behavior and personality (cf. Hebb, 1955).

In this article we trace the influence of Pavlov's ideas through Gray's neuropsychology of emotion and personality and argue that the formulation of rigorous theory – or, at least, rigorous formulation of the central problems to be addressed – is a precondition to scientific advance in general psychology as well as in the psychophysiology of personality. Although this argument may seem trite – perhaps even too obvious to mention to fellow scientists – it is obvious that, at least in the field of personality, much research is still atheoretical, relying upon data-driven strategies.

2. Ivan Pavlov

Ivan Pavlov is acknowledged as one of the founding fathers of modern experimental psychology, world famous for his work on the conditioned reflex. Less well known is his fundamental work on personality and its extension to psychiatry (Pickering, 1997). It is, therefore, somewhat ironic that Pavlov had a scathing view of psychology, seeing his own contribution fitting into the emerging discipline of the "physiology of higher nervous activity" (Pavlov, 1927; see Gray, 1979). Pavlov was first and foremost a physiologist and only became a psychologist when he was dragged across the gulf from physiology to psychology by a phenomenon revealed during studies of digestion. This phenomenon was, of course, the now-famous conditioned reflex that manifested itself as the tendency of the dogs that were the subjects of Pavlov's digestion experiments to start salivating at signs that predicted imminent feeding. (The term conditioned reflex was a translation error: according to Gray (1979) what Pavlov actually wrote was "conditional reflex" which makes a lot more sense as it describes a reflex that is conditional upon learning having taken place. The phrase 'conditioned reflex' has, however, become "hallowed by use" (Gray, 1979, p.42) and so is retained in the psychological literature.) Initially labeled "psychic secretion" – or "appetite-juice" – to highlight the psychological state of desire (Pavlov, 1897) – the conditioned reflex was a well known occurrence in Pavlov's laboratory for many years prior to his first official interest in 1897; the earliest research being performed by one of Pavlov's colleagues, Anton Snarsky, who was the first person to show that a conditioned reflex can be elicited by any arbitrary stimulus — the precise details of Snarsky's contribution may never be known because he advocated a mentalist approach couched in terms of the dog's desires and expectations with which Pavlov disagreed and so became persona non grata in Pavlov's laboratory.

The change from using stimuli that act on the mouth (as in Pavlov's original digestion studies) to stimuli that act on the distance receptors (eyes and ears) is, as noted by Gray (1979, p. 33), "…not a big step to take. But, if there is a divide between physiology and psychology, this step takes you across it. Pavlov was well aware of this." This small practical step thus represented a major theoretical leap for Pavlov, moving him away from a purely physiological account of digestive behavior to one based on psychology (i.e., a conceptual analysis of behavior) and so he hesitated for several years before he finally began the systematic investigation of conditioned reflexes that would one day make him famous and overshadow his Nobel-prize winning work on digestion (Gray, 1979).

In the light of these facts about Pavlov, it is interesting to speculate whether, if he were miraculously resurrected today, he would be any more comfortable with being labeled a psychologist now than he was during his own lifetime? Of course this is an impossible question to answer; but we may speculate that, in the field of personality research at least, it would depend on the type of articles he read. Pavlov's reading of articles from the "psychophysiological" school of personality research, epitomized by the work of Hans Eysenck, would probably prompt a favorable response; however, his reading of articles from the "trait descriptive" school of personality research, characterized by the Big Five model (Goldberg, 1981), might elicit a different response.

The difference between these two schools of personality research is clear: the first is driven by causal biological desiderata, amenable to empirical testing in objective experiments; the second is based on observations of behavior, sifted using statistics, with a view to revealing the *manifest* structure of personality. This second approach may, perhaps uncharitably, be dubbed "naming, but not explaining", and arguably the 'factors' uncovered are no closer to explaining the causal basis of personality traits than the ancient peoples who first described them.

Multivariate statistical techniques, the most important in personality research being factor analysis, can provide only a preliminary guide to the biological processes underlying sources of the most common variations in a population. Factor analysis simply cannot differentiate, for example, separate causes that are conflated in development, and nor is it able to identify primary causes that have become conflated in their expression. It works on measures of the phenotype that may be (and often are) the end product of a long chain of causal, and interacting, factors. Sometimes, underlying causal and phenotypic factors may be so similar as to allow a one-to-one correspondence, but this outcome is one of serendipity not of the logic of factor analysis. Therefore, 'discovering' factors of personality and then searching for their causal bases may, in many instances, be a flawed strategy, and this is nowhere more evident than in the psychophysiology of personality (Corr and McNaughton, submitted for publication).

The limitations of the descriptive approach to personality does not imply that observation has no role in theory development; quite the contrary: just as Pavlov's study of conditioned reflexes grew out of chance observations made during research into digestion, so his research into personality grew out of chance observations of individual differences in behavior in the dogs that served as subjects in his conditioning experiments (see Fig. 1). The difference between Pavlov and the descriptive psychologists is that, instead of resting content with the collation and analysis of his dog's personality in purely descriptive terms, he sought to advance causal theories to explicate the mechanisms responsible for their varied behaviors. Interestingly, he would challenge his own theories if a single dog's behavior did not conform to theoretical expectation, rather than relegate this dog's behavior to a statistical outlier.

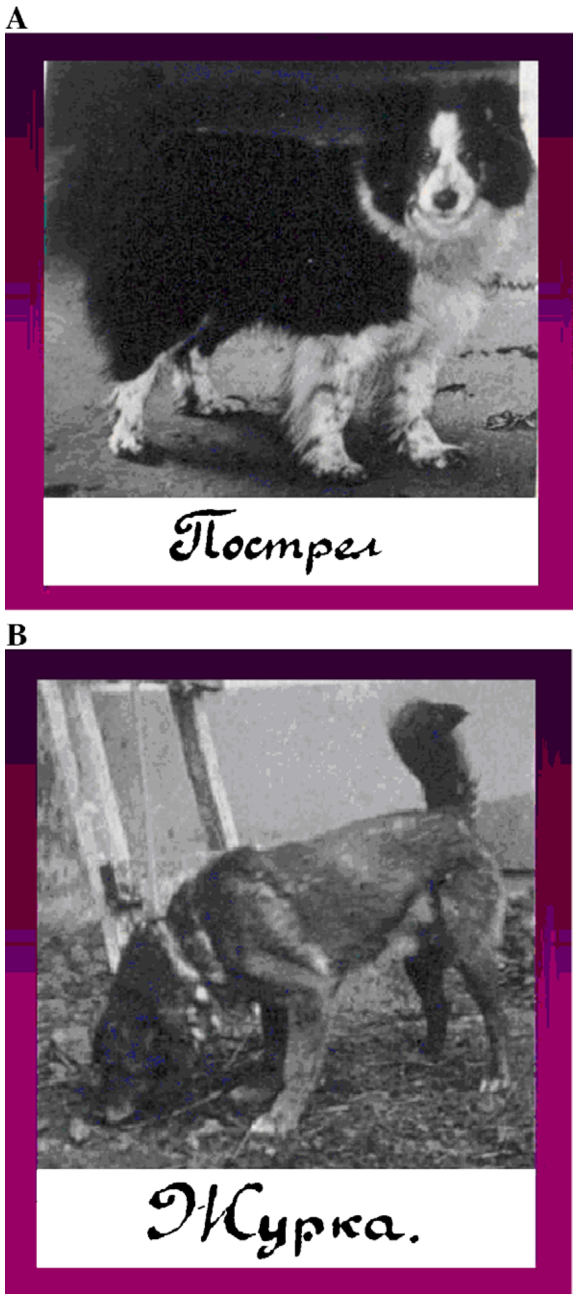

Fig. 1. Pavlov noticed systematic differences in the behavior of his experimental dogs, which through careful experimentation suggested distinct personalities; indeed, it was these differences that often led to theoretical insights into the nature of conditioning phenomena. It was scientifically courageous of Pavlov to take seriously such individual differences, regarding them as empirical data and not experimental 'error' and an inconvenient nuisance in his experimental studies. Pavlov's methodology also holds important lessons for modern-day personality psychologists. He would use mongrel dogs raised to maturity outside the laboratory, which were then kept under humane conditions for long periods of time during which extensive study took place. For example, the dog shown in (A) had a bold and lively personality and so was named "Postrel", which is colloquial Russian for gunshot; conversely, the dog shown in (B), had a timid and cowardly personality and so was named "Jurka", which is colloquial Russian for moaner/grumbler.

Less well known about Pavlov was his interest in psychiatry; in his later years he would attend ward rounds in hospitals and apply his concepts to neurosis in human beings; he also developed the field of 'experimental psychopathology' – for example presenting to dogs increasingly similar discriminative stimuli eliciting conflicting responses, to induce experimental neurosis – which was to influence modern psychiatry between the World Wars and lead to research on war neurosis (e.g., Slater, 1943).

3. Hans Eysenck

Amongst the most prominent psychologists to be influenced by Pavlov's ideas and findings was Hans Eysenck who started his work in personality with the investigation of the statistical structure of medical symptoms observed in war neurotics (700 soldiers invalided out the British Army in World War 2). Employing factor analysis, he isolated two major dimensions of variation which, on the basis of existing theories concerning the relationship between psychiatric disorders (especially dysthymia and hysteria) and personality, he conceptualized as Extraversion (E) and Neuroticism (N) (Eysenck, 1944, 1947). Then later he used criterion groups to sharpen his operational definitions of E and N as well as to select items used in the development of scales, once again using factor analysis (see Eysenck, 1960). In 1957, he published a causal theory of personality, incorporating Pavlov's concepts of excitation–inhibition and mobility: introversion–extraversion was aligned with the processes of excitatory and inhibitory processes, respectively, and neuroticism was aligned with mobility.

Eysenck tested his theory using both behavioral and psychophysiological methodologies (e.g., eye blink conditioning; see Eysenck and Levey, 1972). The rigor with which Pavlov's theory was elaborated in Eysenck's personality theory allowed firm conclusions to be drawn from experimental data, and allowed hypotheses to be tested, often to destruction. In 1967, the theory of personality based on the excitatory–inhibitory balance and mobility was reformulated in modern terms of cortical arousal and limbic activation (Eysenck, 1967). It was at this very juncture that Gray's work focused, exposing the theoretical and empirical cracks in Eysenck's personality edifice (see Gray, 1981), leading to the formulation of an alternative theory of personality based on sensitivities to rewarding and punishing stimuli. Without Pavlov's pioneering work, and then Eysenck's incorporation of it into modern accounts of personality, it is difficult to see how Gray's highly influential theory would be developed.

4. Jeffrey Gray

At the time of the transition between Eysenck's 1957 inhibition–excitation and 1967 arousal–activation theories, Gray was undergoing doctoral training in Eysenck's

Department, at the Institute of Psychiatry (Maudsley Hospital in London). Knowing that Gray was also a modern languages scholar, and having only recently been trained in Russian by the British military during national service, Eysenck asked the young Gray to translate some psychological books written in Russian. These books included the work of Teplov, and the translation formed part of his PhD thesis, awarded in 1964.

The remainder of Gray's PhD work was concerned with Hullian 'drive' in rats, like Pavlov using purely experimental (behavioral) data to infer central processes of arousal, drive-reduction, etc. This work was strongly influenced by then recent neurophysiological work on the 'pleasure centers' in the brain (e.g., Olds and Milner, 1954). Stemming from this work, Gray (1975) was later to argue that, instead of the Hullian based notion of a single dimension of drive (or arousal in Eysenck's theory), behavioral, lesion and pharmacological evidence pointed to two separate mechanisms of reward and punishment (see Corr et al., 1995). This theory was further supported by Mowrer's (1960) two-factor theory of learning, as well as Konorski's (1967) work on reward and punishment systems in the brain. Thirty years of research attest to the existence of two such systems in personality (Corr, 2004).

As noted by Gray (1983), an objection sometimes raised against the psychophysiological approach to behavior is that the existence of individual differences renders the development of a lawful science of behavior impossible — or, at least, impractical. In answer to this criticism, Pavlov's general approach contends that there are general laws that apply to all individuals within a given species, and these laws arise from the general functioning of the systems that control behavior, and that differences between members of a given species reflect differences in the parameters of the functioning of these systems. Thus, as Gray (1983, p. 32) notes, "It follows from this analysis that the study of personality is inextricably related to the study of the general laws of behavior: the two enquiries are opposite sides of the same coin." An important consequence of this approach is that the statistical observation of individual differences can point to the existence of the systems that control behavior; but this is only possible within a biological model that recognizes that personality is a reflection of different functioning in general causal systems of behavior.

In marked contrast to this Pavlov–Eysenck–Gray tradition, the descriptive trait approach is still debating the 'best' factor solution; a debate that is, potentially at least, not resolvable in statistical terms for the reasons given above.

5. The 'Problem' of the psychophysiology of personality

Before discussing the details of Gray's *Reinforcement Sensitivity Theory* (RST) – which here is presented as a paradigmatic example of a scientific theory of personality, or at least the beginnings of one (i.e., a *prolegomenon*[2]; Corr and McNaughton, submitted for publication) – it would be wise first to stand back to consider the nature of theory and theoretical problems.

A theory is a nomological network of concepts, with its own set of principles, premises and conclusions, cemented by the mortar of internal logical consistency, and rendered scientifically fecund by its predictive power to explain previously known phenomena and, most importantly, yet to be discovered phenomena. The power to predict is the gold standard of a scientific theory: according to standard Popperian wisdom, a scientific theory, as opposed to prescientific thinking (e.g., religion) and quasi-science (e.g., astrology), is one vulnerable to falsification. However, long before a set of concepts and hypotheses can be elevated to the status of a 'theory', the nature of the scientific problem needs to be clarified. This point is especially germane in the atheoretical world of much of personality psychology, including the empirically grounded type of research that is often found within the confines of the psychophysiology laboratory.

Opposing the scientific method of induction, Popper (1972/1999) noted that the starting point for each new development in science is the 'problem' or 'problem situation' — that is, the appearance of a problem in certain state of current knowledge: 'without a problem, no observation' (p. 6). The method to solve problems represents a form of trial and error — "To be more precise, it is the method of *trying out* solutions to our problems and then discarding the false ones as erroneous" (p. 3). Although this quote may appear little more than an innocuous statement of the obvious, it holds important implications for the formulation of theory, and especially for its empirical test. The method of *expectation* (hypothesis), *test* and *rejection* (involving the elimination of hypothesis; refutation) depends critically upon the formulation of the scientific problem in the first instance: an inadequately formulated hypothesis is either immune to refutation, or open to refutation but without a means by which the process of refutation suggests alternative hypotheses. Thus, the 'problem situation' needs to be set up in an operational, experimental form that lends itself to rejection of false hypotheses. However, all too often in psychophysiology a hypothesis may fail, but it is extremely difficult to decide whether this falsification is a rejection of a fundamentally flawed hypothesis or some inadequacy in the experimental setup (e.g., a failure to relate physiological processes to the key psychological processes of interest). Indeed, all too often, it is not clear which physiological variables are of most interest, and why.

Occasionally at psychophysiology conferences, Gray would make a joke at the expense of psychophysiologists (Tony Gale, personal communication, 11th June, 2005; see Gray, 1994 for another version of this story and its relevance for RST).

> "A man is looking on the ground under a street lamp on a dark night. A policeman asks him what is he looking for? The man replies 'My car keys'. The policeman then asks, 'You lost them here, did you?' 'No', the man replies, 'I lost them over there.' Confused, the policeman asks, 'Then why

[2] After reading a draft, Jeffrey Gray himself suggested that the Corr and McNaughton (submitted for publication) paper, that explores the implications of the Gray and McNaughton (2000) theory in terms of personality, was more a prolegomenon than a fully fledged theory.

are you looking on the ground over here then?' 'Well', replies the man, 'that's where the light is!'

The 'light' provided by technology may seem to suggest solutions to theoretical problems – and psychophysiology all too often seems technology-driven – but there is no substitute for properly articulating the theoretical problem in the first place. We now turn to Gray's personality theory which represents, among many others, an attempt to clarify the 'problem situation'. We discuss how this theory guides the psychophysiological study of individual differences in basic systems of emotion, motivation and behavior, and the problems this theory raises for more traditional accounts of the psychophysiology of personality. (A full account of the development, and empirical test, of RST has already been provided; Corr, 2004.)

6. Gray's Reinforcement Sensitivity Theory (RST)

Gray's theory has been in development over a period of thirty years. The revised version of the theory (Gray and McNaughton, 2000) is described below, and contrasts are drawn with the original theory (Gray, 1982). At this point it is important to note that the continual revisions in Gray's RST reflect the spirit of clarification and development in response to new data and thinking: desirable signs of a progressive science.

In brief, RST views substantive affective events as falling into just two distinct types, positive and negative (McNaughton and Corr, 2004). It also considers the absence of an expected positive event as functionally equivalent to the presence of a negative event, and vice versa. The most recent revision of RST represents a major revision and clarification: the new theory has a more elaborate neurophysiology, and it makes new predictions especially with regard to the elicitation of fear and anxiety, which are seen as related, although at times opposing emotions.

Revised RST postulates three systems.

1. The *fight–flight–freeze system* (FFFS) is responsible for mediating reactions to *all* aversive stimuli, conditioned and unconditioned. A hierarchical array of neural modules comprises the FFFS, responsible for avoidance and escape behaviors. Importantly, the FFFS mediates the "get me out of this place" emotion of fear, not anxiety. The associated personality comprises a combination of fear-proneness and avoidance, which clinically is mapped onto such disorders as phobia and panic. (In contrast, the original, 1982, theory assigned the FFFS to reactions to *unconditioned* aversive (pain) stimuli only.)
2. The *Behavioral Approach System* (BAS) mediates reactions to *all* appetitive stimuli, conditioned and unconditioned. This generates the appetitively hopeful emotion of 'anticipatory pleasure'. The associated personality comprises a cluster of optimism, reward-orientation and impulsivity, which clinically maps onto addictive behaviors (e.g., pathological gambling) and various varieties of high-risk, impulsive behavior, and possibly the appetitive component of mania. (The BAS is largely unchanged in the revised version of RST.)
3. The *Behavioral Inhibition System* (BIS) is responsible for the resolution of goal conflict in general (e.g., between BAS-approach and FFFS-avoidance, as in foraging situations — but it is also involved in BAS–BAS and FFFS–FFFS conflict). The BIS generates the "watch out for danger" emotion of anxiety, which entails the inhibition of pre-potent conflicting behaviors, the engagement of risk assessment processes, and the scanning of memory and the environment to help resolve concurrent goal conflict. The BIS resolves conflicts by increasing, by recursive loops, the negative valence of stimuli (these are adequate inputs into the FFFS), until behavioral resolution occurs in favor of approach or avoidance. Subjectively, this state is experienced as worry and rumination. The associated personality comprises as combination of worry-proneness and anxiety, with a high BIS person constantly on the look-out for possible signs of danger, which clinically maps onto such conditions as generalized anxiety and obsessive–compulsive disorder (OCD), which reflects a lack of adequate goal conflict resolution appropriate to local environmental conditions — e.g., the door handle really does not hold cancer-inducing viruses. (In contrast, the original, 1982, theory assigned the BIS to reactions to *conditioned* aversive stimuli only.)

7. Fear and anxiety

The Gray and McNaughton (2000) revised theory views fear (FFFS) and anxiety (BIS) as distinct processes. This distinction, which is made explicit in the revised theory, was only implicit in the original (Gray, 1982) theory. This categorical separation derives from detailed analysis of defensive responses by Robert and Caroline Blanchard (e.g., Blanchard et al., 1989). The Blanchards link to a state of fear a set of behaviors elicited by a predator. These behaviors, originally defined ethologically, turn out to be sensitive to drugs that are panicolytic, but not to those that are only anxiolytic. The Blanchards link to a state of anxiety a quite different set of behaviors (especially 'risk assessment'). These behaviors, again defined ethologically, are elicited by the potential presence of a predator and turn out to be sensitive to anxiolytic drugs. The Blanchards' detailed analysis, and its pharmacological validation, provides a basis for coherent conceptualization of a vast animal literature. For example, their analysis of fear predicts the well-demonstrated insensitivity to anxiolytic drugs of active avoidance in a wide variety of species and of phobia in humans (see below).

7.1. Defensive direction

Because of the detailed effects of anxiolytic drugs on behavior (see Gray and McNaughton, 2000), it is argued that the key factor distinguishing fear and anxiety is not that posited by the Blanchards, namely immediacy (or certainty = fear) versus potentiality (or uncertainty = anxiety) of threat but 'defensive direction': fear operates when leaving a dangerous situation (active avoidance; "get me out of here"), anxiety when entering it (e.g. cautious 'risk assessment' approach behavior; "watch out for danger") or withholding entrance (passive avoidance; "reduce behavior to avoid detection").

7.2. Defensive distance

Revised RST contends that defensive behavior results from the superimposition on defensive direction (i.e., approach or avoid) of what is known as 'defensive distance' As noted by McNaughton and Corr (2004), for a particular individual in a particular situation, defensive distance equates with real distance; but, in a more dangerous situation, the perceived defensive distance is shortened. In other words, a defensive behavior (e.g., active avoidance) will be elicited at a longer (objective) distance with a highly dangerous stimulus (corresponding to short perceived distance), as compared to the same behavior with a less dangerous stimulus. According to the theory, neurotic individuals have a much shorter perceived defensive distance, and thus react more intensively to relatively innocuous (real distance) stimuli. For this reason, weak aversive stimuli are sufficient to trigger a neurotic reaction in highly defensive individuals; but for a less defensive individual, aversive stimuli would need to be much closer to elicit a comparable reaction. This set of relations is shown below.

System state	Defensive distance	Real distance sufficient to elicit reaction
High defensive individual	Perceived distance < actual distance	Long
Normal defensive individual	Perceived distance = actual distance	Medium
Low defensive individual	Perceived distance > actual distance	Short

Defensive distance thus operationalizes an internal cognitive construct of intensity of perceived threat. It is a dimension controlling the type of defensive behavior observed. In the case of defensive avoidance, the smallest defensive distances result in explosive attack, intermediate defensive distances result in freezing and flight, and very great defensive distances result in normal non-defensive behavior. Thus, defensive distance maps to different levels of the FFFS (see McNaughton and Corr, 2004).

In human beings, the psychological state at very small defensive distance would be labeled panic. The commonly associated cognition in panic, "I'm going to die", would seem similar to whatever cognitions may be attributed to a rat when it is nose-to-nose with a cat (it is the comparable emotion, we would feel if we were trapped in a car in the path of an oncoming high-speed train). Intermediate defensive distances can be equated with phobic avoidance. With the opposite direction, defensive approach, defensive quiescence occurs at the closest defensive distances (and, in rats, can be distinguished from freezing only by minor postural features). At intermediate distances, risk assessment behavior occurs and, at very great distances, defensive behavior disappears and normal pre-threat behavior reappears.

McNaughton and Corr (2004) view individual differences in defensive distance for a fixed real distance as a reflection of the personality dimension underlying punishment sensitivity. Anxiolytic drugs alter (internally perceived) defensive distance relative to actual external threat. As we shall see below this theory holds important implications for the psychophysiological test of personality hypotheses.

8. Goal conflict

Revised RST theory holds that anxiety results from conflicts between competing available goals. The classic form of such conflict is approach–avoidance (Miller, 1944), and this is the most familiar for those studying anxiety. However, in principle, approach–approach (e.g., which equally desirable job offer to accept?) and avoidance–avoidance (e.g., to escape or freeze?) conflicts would also involve activation of the same system and have essentially the same effects as approach–avoidance conflict. Approach–approach conflict is not likely normally to generate high levels of anxiety; but the aversive component of the conflict rests in the frustration that could result from the relative loss incurred if the wrong choice is made — arguably much of the psychological malaise associated with modern life resides in this very type of frustration.

Once conflict is detected, there is a selective potentiation of the cognitive power of affectively negative current perceptions and affectively negative remembered consequences. Affectively positive ones (although increased by simple drive summation) are not potentiated by conflict. In simple approach–avoidance, this will favor avoidance over approach. While fear and anxiety are distinct, there will be many cases where anxiety (as indexed by anxiolytic action) involves an amplification of fear (i.e., the BIS provides adequate inputs into the fear-FFFS in its conflict resolution process). There will also be cases where anxiety involves an amplification of frustration.

9. Individual differences in reinforcement sensitivity: personality

The key feature of the McNaughton and Corr (2004) view of RST is that defensive distance maps onto a series of distinct neural modules, to each of which is attributed generation of a particular symptomatology (e.g., panic, phobia, obsession). These "symptoms" may be generated in several different ways:

1. As a normally adaptive reaction to their specific eliciting stimuli (e.g., mild anxiety before an important examination);
2. At maladaptive intensity, as a result of excessive sensitivity to their specific eliciting stimuli (e.g., sight of a harmless spider = fearful avoidance);
3. At maladaptive intensity, as a result of excessive activation of a related structure by its specific eliciting stimuli but where the "symptoms" are not excessive given the level of input (e.g., oncoming train = panic).

Normal variation in personality entails variation in sensitivity in either separate modules of the hierarchical defense system and/or general modulatory influences on the overall defense system (for a discussion of RST and personality, see Corr and McNaughton, submitted for publication). As a preliminary model of personality, the comorbidity of neurotic disorders may reside in general modulatory influences (or, relatedly, the ease of coupling of separate neural modules), whereas the specificity of clinical disorders seems to reside in the sensitivity of specific neural modules (e.g., periaqueductal gray and panic). At the very least in RST, fear and anxiety are separate emotions, and the FFFS related to general sensitivity to threat — which, often leads to anxiety due to goal conflict or the need to approach the dangerous environment. In these different situations, different behaviors are elicited and, therefore, different psychophysiological processes.

10. Implications of theory for psychophysiological personality research

Let us tackle a standard question in the psychophysiology of emotion: what are the psychophysiological correlates of fear and anxiety, including its normal and clinical variants? The conventional psychophysiological approach to personality is to take some (often theoretically motivated) psychophysiological measures (e.g., EMG startle or heart rate) and relate these to psychometric traits (e.g., trait anxiety), usually within an appropriate experimental design with control over relevant independent variables. At best, approximate relations may be found, for example between arousal and the BIS and heart rate and the BAS (e.g., Fowles, 1980, 2000).

The problem with this approach is the atheoretical nature of the relationship between personality and psychophysiological parameters. As shown by the discussion of 'defensive distance', a threat stimulus of a fixed intensity leads to different behavioral reactions depending on the individual's *perceived* defensive distance, and with each distinct defensive behavior (e.g., avoidance versus freezing) different psychophysiological processes are engaged. With psychophysiological measures that may measure whole defensive system functioning (e.g., skin conductance), this may not be too much of a problem, but it is altogether a different matter when we want to measure activation of specific neural modules, or to even distinguish between fear and anxiety. The widely reported 'fractionation' (Lacey, 1967) of psychophysiological measures may be a result of the activation of different neural modules at different defensive intensities.

11. RST and theoretical psychophysiology

Now let us turn to the RST model of psychophysiological research. An important conclusion of this theory is that it should be possible to separate different syndromes of defensive disorders by using theoretically based challenge tests, and by so doing circumvent the problem that different syndromes can present with much the same symptoms (which occurs due to secondary activation of the entire defensive hierarchy). Indeed, a key feature of these tests is that they should seldom be directed towards the most obvious symptoms and should be administered when state anxiety and hence symptoms are minimal. The same would of course be true of any challenges used to activate the brain for imaging (for a discussion of this matter, see McNaughton and Corr, 2004).

To be more precise, the central idea of differential diagnosis is that the specific nodes of the defense system should be selectively challenged to determine whether they are functioning normally. Such challenges should be designed to produce *minimal* reactions from the rest of the defense system, otherwise, anxiety (or fear or panic) will automatically spill over into activation of much of the remainder of the system, so making it impossible to determine at which point excessive reactions begin. What challenge tests could be used to test some aspects of the theory and form a diagnostic scheme? Three types of challenge tests are discussed below.

11.1. Periaqueductal gray and panic

At the bottom of the defense system is the periaqueductal gray, which mediates fight/flight reactions to impending danger, pain, or asphyxia. What is required is a stimulus maximally activating this region accompanied by minimal activation of other parts of the defense system. With such a challenge healthy individuals and anxious patients could be tested for the extent to which the periaqueductal gray itself is over-reactive, as opposed to being secondarily triggered by excessive activity elsewhere in the defense system. To detect only clinical panic disorder, the *threshold* level of CO_2 required to elicit an attack would need to be determined (as soon as panic is elicited, other parts of the defense system would contribute to the attack). Threshold measurements should detect supersensitivity in the periaqueductal gray independent of other abnormalities in the defense system.

11.2. Amygdala and emotional arousal

Amygdala dysfunction is associated with the arousal component of anxiety, and one of the most relevant challenges is 'fear'-potentiated startle, since this is sensitive to anxiolytic drugs (including when injected into the amygdala). Affective modulation of the startle reflex is a widely validated and employed psychophysiological measure of emotional responding, which, according to RST, is not mediated by the FFFS and thus is not primarily associated with the emotion of fear (despite its name). Simply discussing affective modulation in terms of negative valence obscures these issues, as does the psychometric overlap of, so-called, 'fear' (associated with *leaving* a dangerous situation) and 'anxiety' (associated with *entering* a dangerous situation) scales.

11.3. Septo-hippocampal system

A challenge test is required that is sensitive to septo-hippocampal system damage and anti-anxiety drugs, but *not* to amygdalar or periaqueductal gray lesions. Obvious tasks include spatial navigation, delayed matching to sample and behavior on a fixed interval schedule of reward-delayed matching to sample is preferable because it can be set up in an anxiety-free form (see McNaughton and Corr, 2004).

12. Testing problems

As a preliminary test of the 'fear or anxiety hypothesis', we first might conduct a literature search to ascertain whether psychometric measures of fear and anxiety are, indeed, statistically separable, perhaps correlating each with some general measure of neuroticism. Then we might search for empirical evidence to show that such measures of fear and anxiety have different predictive validities. Perkins and Corr (in preparation) pursued both strategies and found evidence for discriminable fear and anxiety factors. We might then turn to the functional neuroimaging literature, to use the prism afforded by RST to inform interpretation of data. Let us see how far RST takes us.

Canli et al. (2001) required female-only participants to watch positive and negative scenes. The results showed that brain reactivity to positive (relative to negative) pictures correlated with extraversion in a number of cortical and subcortical locations, including the amygdala; similarly, brain reactivity to negative (relative to positive) pictures correlated with neuroticism at several sites. However, the study lacked a baseline (neutral) condition, so the neural responses represented differences in activation patterns to positive and negative images — thus, it was not possible to distinguish increased activation to positive stimuli from decreased activation to negative stimuli.

In a second study, Canli et al. (2002) tested both males and females, and it also included a baseline (neutral) condition. The finding that amygdala activation to happy (versus neutral faces) faces correlated with extraversion was replicated. Importantly, this relation was specific to happy faces, because amygdala activation to other emotional facial expressions did not correlate significantly with this trait.

Reuter et al. (2004) tested males and females in an fMRI-emotion task, using slides that depicted facial reactions indicative of fear, disgust, sexual arousal and joy. They also took specific measures of sensitivity to reward and sensitivity to punishment, as measured by the Carver and White (1994) BIS/BAS scales. BIS scores were significantly correlated with brain activity in response to disgust in the anterior cingulate, (right) amygdala, and thalamus. Unexpectedly, there was also a significant correlation between BAS score and brain activity induced by disgust in the insula. Significant correlations between BIS scores and fear-induced brain activity were observed in the cingulate and thalamus — the BAS scale was not associated with fear-related brain activation. The BAS scale was associated with brain activation to erotic pictures, and this activation was observed in the hippocampus. Such data show the complexity of FFFS, BIS and BAS reactions to emotive stimuli.

13. Anxiety: amygdala and/or septo-hippocampal system?

Neuroimaging can be used to resolve theoretical questions, and inform personality research; but much of it to date has an assumed 'amygdalocentric' perspective. There are, however, problems with this perspective. First, the amygdala seems to do too much: it is responsible for cue-reinforcer conditioning with aversive and appetitive stimuli, and unconditioned responses, such as aggressive and sexual behavior, also depend on this structure. Thus, the amygdala seems to be involved in controlling the outputs of many, if not all, emotion systems: it is responsible for producing general emotional arousal (the evidence discussed above shows that it is also involved in mediating emotional reactions to happy faces).

In addition, anxiety is not about fear of pain solely, but also failure or loss of reward. There is evidence that frustrative nonreward (i.e., the non-appearance of an expected reward) is aversive. Certain forms of frustration, particularly those that are anticipatory, are reduced by anxiolytics. However, lesion of the amygdala does not affect frustration-mediated responses — so it seems that there is a problem to be solved: an obviously negative emotional state, that is sensitive to anxiety-reducing drugs, is not affected by destruction of the amygdala 'emotion center'. Therefore, this negative emotion must be generated somewhere else. Also, there are other behaviors related to anxiety that are not affected by lesions to the amygdala.

The Gray and McNaughton (2000) theory hypothesis that the processing of anxiety-related information is performed by the septum and hippocampus – the 'septo-hippocampal system' (SHS) – and this outputs to the amydala which then generates the emotional arousal component of anxiety. Does neuroimaging—which is now a dominant methodology in the psychophysiology of personality, have the potential to resolve this theoretical conflict?

Ploghaus et al. (2001) presented healthy volunteers with a painful heat stimulus under conditions of either low or high anxiety (high anxiety was induced by the presentation of a stimulus that threatened an even more painful stimulus). Activation indexed by fMRI was found to be increased in the hippocampal formation, along with correlated activity in a region of the insular cortex specialised for pain perception. This finding is consistent with reformulated RST.

Furmark et al. (2002) tested social phobic patients during a public speaking task, which generates an approach (BAS)–avoidance (FFFS) conflict. PET activity was measured in a control (no treatment) group and before and after treatment of two kinds: drug and cognitive–behavioral therapy. In those patients who responded to treatment (not all did), and independently of the treatment applied, improvement in symptoms were associated with a decrease in rCBF in the hippocampus and the amygdala. This finding is consistent with Gray and McNaughton's (2000) theory that social phobia involves dysfunction in both the hippocampus and the amygdala — in contrast, simple phobia, where avoidance is possible in the absence of an approach–avoidance conflict, the hippocampus is not involved (i.e., simple phobia is fear, not anxiety).

14. Conclusion

Psychophysiology is a field of scientific enquiry concerned specifically with the relationship between psychological and physiological phenomena. This psychological–physiological mapping has been prone to a frustrating array of problems and pitfalls, some methodological but others substantially theoretical. Navigation through this dense jungle of concepts and methods has been slow, especially in the arena of personality research. A route map is needed, however, rudimentary and incomplete. This has been the aim of Gray's RST, and before him, Eysenck's theories.

We are still at an immature stage of technological development, and the advent of a new technology often seems to set the research agenda. There is the obvious danger that the dazzle of new technology obscures the need for proper theory. As RST purports to demonstrate, theory brings a much greater understanding of the psychophysiological basis of individual differences, moving us away from brute correlations between traits and structures, or activation within structures, to providing answers to the why questions, supported by findings in such diverse areas as evolutionary theory, learning theory, ethoexperimental analysis and pharmacological data. For example, molecular genetic and functional neuroimaging studies are now revealing the neurological basis of RST systems, and how they relate to personality dimensions (e.g., Reuter et al., 2005).

It is easy to imagine that if Pavlov were to enter a modern psychophysiology laboratory today investigating personality, he may well be impressed by the sophistication of the equipment but surely he would feel fully at home with the use of theory and formal tools of analysis. He would also probably agree that the further development of the psychophysiology of personality will, no doubt, continue to be inspired by Kurt Lewin's (1951, p.169) famous phrase, "There is nothing so practical as a good theory".

References

Blanchard, R.J., Brain, P.F., Blanchard, D.C., Parmigiani, S. (Eds.), 1989. Ethoexperimental Approaches to the Study of Behavior. Kluwer Academic Publishers, Dordrecht.

Canli, T., Zhao, Z., Desmond, J.E., Kang, E., Gabrieli, J.D.E., 2001. An fMRI study of personality influences on brain reactivity to emotional stimuli. Behavioral Neuroscience 115, 33–42.

Canli, T., Sivers, H., Gotlib, I.H., Gabrieli, J.D.E., 2002. Amygdala activation to happy faces as a function of extraversion. Science 296, 2191.

Corr, P.J., 2004. Reinforcement sensitivity theory and personality. Neuroscience and Biobehavioral Reviews 28, 317–332.

Corr, P.J., McNaughton, N., submitted for publication. Implications and extensions of J. A. Gray's behavioral inhibition system: a prolegomenon to the neuroscience of personality.

Corr, P.J., Pickering, A.D., Gray, J.A., 1995. Personality and reinforcement in associative and instrumental learning. Personality and Individual Differences 19, 47–71.

Eysenck, H.J., 1944. Types of personality: a factorial study of 700 neurotics. Journal of Mental Science 90, 859–861.

Eysenck, H.J., 1947. Dimensions of personality. K. Paul, Trench Trubner, London.

Eysenck, H.J., 1957. The Dynamics of Anxiety and Hysteria. Praeger, New York.

Eysenck, H.J., 1960. The Structure of Human Personality, 2nd ed. Methuen, London.

Eysenck, H.J., 1967. Biological Basis of Personality. Thomas, Springfield, IL.

Eysenck, H.J., Levey, A., 1972. Conditioning, introversion–extraversion and the strength of the nervous system. In: Nebylitsyn, V.D., Gray, J.A. (Eds.), Biological Bases of Individual Behavior. Academic Press, New York.

Fowles, D.C., 1980. The three arousal model: implications of Gray's two factor learning theory for heart rate, electrodermal activity and psychopathy. Psychophysiology 17, 87–104.

Fowles, D.C., 2000. Biological variables in psychopathology: a psychobiological perspective, In: Adams, H.E., Sutker, P.B. (Eds.), Comprehensive Handbook of Psychopathology, 3rd edition. Plenum Press, New York, pp. 85–104.

Furmark, T., Tillfors, M., Marteinsdottir, I., Fischer, H., Pissiota, A., Langstrom, B., Fredrikson, M., 2002. Common changes in cerebral blood flow in patients with social phobia treated with citalopram or cognitive–behavioral therapy. Archives of General Psychiatry 59, 425–433.

Goldberg, L.R., 1981. Language and individual differences: the search for universals in personality lexicons. In: Wheeler, L. (Ed.), Review of Personality and Social Psychology, vol. 2. Sage, Beverly Hills, CA, pp. 141–165.

Gray, J.A., 1964. Pavlov's Typology. Pergamon Press, Oxford.

Gray, J.A., 1970. The psychophysiological basis of introversion–extraversion. Behavior Research and Therapy 8, 249–266.

Gray, J.A., 1972. Learning theory, the conceptual nervous system and personality. In: Nebylitsyn, V.D., Gray, J.A. (Eds.), The Biological Bases of Individual Behavior. Academic Press, New York, pp. 372–399.

Gray, J.A., 1975. Elements of a Two-Process Theory of Learning. Academic Press, London.

Gray, J.A., 1979. Pavlov. Fontana, London.

Gray, J.A., 1981. A critique of Eysenck's theory of personality. In: Eysenck, H.J. (Ed.), A Model for Personality. Springer, Berlin, pp. 246–276.

Gray, J.A., 1982. The Neuropsychology of Anxiety. Oxford University Press, Oxford.

Gray, J.A., 1983. Anxiety, personality and the brain. In: Gale, A., Edwards, J.A. (Eds.), Physiological Correlates of Human Behavior, vol. 3. Academic Press, London, pp. 31–43.

Gray, J.A., 1987. The Psychology of Fear and Stress. Cambridge University Press, London.

Gray, J.A., 1994. Three fundamental emotion systems. In: Ekman, P., Davidson, R.J. (Eds.), The Nature of Emotion: Fundamental Questions. Oxford University Press, Oxford.

Gray, J.A., McNaughton, N., 2000. The Neuropsychology of Anxiety: An Enquiry into the Functions of the Septo-Hippocampal System, 2nd ed. Oxford University Press, Oxford.

Hebb, D.O., 1955. Drives and the C.N.S. (Conceptual Nervous System). Psychological Review 62, 243–254.

Konorski, J., 1967. Integrative Activity of the Brain. Chicago University Press, Chicago.

Lacey, J.I., 1967. Somatic response patterning and stress: some revisions of activation theory. In: Appley, M.H., Trumbull, R. (Eds.), Psychological Stress. Appleton Century Crofts, New York, pp. 14–43.

Lewin, K., 1951. Field theory in social science: selected theoretical papers. In: Cartwright, D. (Ed.). Harper & Row, New York.

McNaughton, N., Corr, P.J., 2004. A two-dimensional neuropsychology of defense: fear/anxiety and defensive distance. Neuroscience and Biobehavioral Reviews 28, 285–305.

Miller, N.E., 1944. Experimental studies of conflict. In: Koch, S. (Ed.), Psychology: A Study of a Science–General Systematic Formulations, learning and Special Processes, vol. 2. McGraw-Hill, New York, pp. 196–292.

Mowrer, O.H., 1960. Learning Theory and Behavior. Wiley, New York.

Olds, J., Milner, P., 1954. Positive reinforcement produced by electrical stimulation of septal area and other regions of rat brain. Journal of Comparative Physiological Psychology 47, 419–427.

Pavlov, I.P., 1897. The Work of the Digestive Glands. (Translated by W. H. Thompson, 1902, and reprinted in 1982 by The Classics of Medicine Library, Birmingham).

Pavlov, I.P., 1927. Reflexes: An Investigation of the Physiological Activity of the Cerebral Cortex. Oxford University Press, Oxford. (Translated and edited by G.V. Anrep.).

Perkins, A.M., Corr, P.J., in preparation. Differentiating fear and anxiety: A preliminary exploration of Gray & McNaughton's (2000) theory of personality.

Ploghaus, A., Narain, C., Beckmann, C.F., Clare, S., Bantick, S., Wise, R., Matthews, P.M., Rawlins, J.N, Tracey, I., 2001. Exacerbation of pain by anxiety is associated with activity in a hippocampal network. Journal of Neuroscience 15; 21(24), 9896–9903.

Pickering, A.D., 1997. The conceptual nervous system and personality: from Pavlov to neural networks. European Psychologist 2, 139–163.

Pickering, A.D., Diaz, A., Gray, J.A., 1995. Personality and reinforcement: an exploration using a maze-learning task. Personality and Individual Differences 18, 541–558.

Popper, K.R., 1972/1999. The logic and evolution of scientific theory. In: Popper, K.R. (Ed.), All Life is Problem Solving. Routledge, London, pp. 3–22.

Reuter, M., Stark, R., Henning, J., Walter, B., Kirsch, P., Schienle, A., Vaitl, D., 2004. Personality and emotion: test of Gray's personality theory by means of an fMRI study. Behavioral Neuroscience 118, 462–469.

Reuter, M., Schmitz, A., Corr, P.J., Hennig, J., 2005. Molecular genetics support Gray's personality theory: the interaction of COMT and DRD2 polymorphisms predicts the behavioral approach system. International Journal of Neuropsychopharmacology 8, 1–12.

Slater, E., 1943. The neurotic constitution. Journal of Neurology and Psychiatry 6, 1–16.