STEPS TOWARD AN EVOLUTIONARY-BASED THEORY OF PSYCHOTHERAPY - I

Michael T. McGuire, Alfonso Troisi

Summary

This paper asks the question: what is required to develop an evolutionary-based theory of psychotherapy which can generate falsifiable hypotheses? Points discussed include: evolved constraints on treatment outcome, brain attributes which may affect treatment, differences between illness and mental disorders, explanations (hypotheses) which defy falsification, adaptive and maladaptive behavior, intervening events, what should be measured, and how psychotherapy works.

It is concluded that it is possible to develop an evolutionary-based theory of psychotherapy which can generate falsifiable hypotheses but to do so will require significant changes in how therapy is conceptualized and what is measured.

Key words: Psychotherapy – Evolutionary Biology – Neuroscience – Illness – Mental Disorders – Behavior Measurement

Declaration of interest: none

[1] Michael T. McGuire, M.D. is Professor of Psychiatry and Biobehavioral Science – Emeritus, University of California at Los Angeles, USA.
[2] Alfonso Troisi, M.D. is a Research Psychiatrist at the Department of Neurosciences, and Professor of Psychosomatics, Medical School, University of Tor Vergata, Rome, Italy.

Corresponding author: Michael McGuire, M.D., P.O. Box 1646, Cottonwood, CA 92022 USA. Email: mmcguire@evalflex.com

Introduction

This and a companion paper (McGuire and Troisi 2006) are 'thought papers' which address the question: what is required to develop an evolutionary-based theory of psychotherapy which can generate falsifiable hypotheses? Psychotherapy is defined as a therapist and patient talking together with the aim of changing functionally defined behavior. This is a narrow definition which differs from the more frequently voiced aims of therapy.

Adopting a narrow definition of therapy has advantages and disadvantages. Narrowing reduces the number of possible hypotheses, the behaviors that might be the targets of therapy, and the number of possible outcome measures. In our view, narrowing is essential because psychotherapy involves far more than changing functionally defined behaviors — therapists provide emotional support, advise, educate, attempt to minimize undesirable signs and symptoms, and so forth. Changing functional defined behaviors – hereafter "functional behavior(s)" — is not incompatible with these other activities. Sign and symptom reduction may accompany or even facilitate behavior change. But the reverse is not necessarily true as, for example, when therapy results in patients feeling better but their compromised investment behavior in kin remains unchanged. Further, there are multiple sources of compromised functional behaviors. Selection effects on traits is one possibility. Personal experiences of unpleasant anxiety and/or depression is another possibility. And there are compromised functional behaviors which are associated with brain diseases such as Huntington's disease and dementia. Each of these invites a different type of intervention and intervention objective.

Evolutionary concepts focus on functional systems that have been selected to increase inclusive fitness. Obvious examples of these systems include survival, reproduction, kin investment, and reciprocal exchanges among nonkin. Often one or more of these systems and/or the behaviors associated with them are compromised due to selection effects. Yet, functional systems and behaviors are not cast in concrete. They are 'ecologically sensitive' – that is they are responsive to experience and one may seek a mate or invest in kin in a variety of ways — hence variation in their phenotypes is expected. And at times there are uncertainties — as mentioned, compromised systems or behaviors may be a consequence of adverse experiences, dementia, addiction, etc. These may differ or be similar to systems or behaviors compromised by selection. So some sorting out of selection effects and other contributing fac-

Received January 2006, accepted March 2006

tors is required – this issue is addressed below.

There is an important difference between rational interventions – which we define as developing an intervention that predicts a specific change – and empirical interventions – doing something because it worked in the past. A familiar example of the former type is prescribing a drug such as an antibiotic for patients whose symptoms are thought to be due to a bacterial infection responsive to that specific drug. Many of the reports of successful psychotherapy are of the empirical type – they work but the 'why' and 'how' of their workings remain a mystery. And they work only sometimes and they seemingly work better with some therapists than others – points which reduce their predictive potential. Here we are concerned with rational interventions.

In this paper our focus is on issues which we believe require discussion in advance of proposing falsifiable hypotheses. We have adopted a 'hard-line' approach towards existing theories and findings. This means that hypotheses which have not been subject to research that could falsify them, however suggestive and/or consistent they might be with evolutionary concepts, or which seem unlikely to lead to such research, are not discussed in detail. Thus much of the literature dealing with mental disorders and evolutionary biology is not cited. We emphasize here, and we will do so again, that when we cite examples of hypotheses that are unlikely to be falsified, our intent is not to disparage or dismiss them. In every branch of science, whether during a period of development or at a more mature phase, theorizing and hypothesis generation are critical for identifying questions, fostering research, and interpreting data. We and others are doing just that.

We begin with a discussion of possible evolved constraints on treatment outcome.

Evolved constraints on treatment outcome

An informative place to start a discussion of possible evolved constraints on psychotherapy is to review points developed by Nesse (2005) in a recent paper entitled *Evolutionary Psychology and Mental Health* which provides a concise summary of evolutionary biology's contributions to psychiatry.

Nesse points out that at present there are no evolutionary-based treatments for mental disorders. Rather, he argues that evolutionary biology's main contribution to psychiatry is that of providing a perspective and offering a theoretical framework for integrating findings from the many disciplines (e.g., psychiatry, psychology, neurosciences, pharmacology) which study and treat persons labeled as suffering from mental disorders or those who are not so labeled but seek treatment (Nesse 2005). We agree with his view.

He then goes on to discuss evolutionary processes which might contribute to disorders as they are defined in psychiatry's official taxonomic system (Diagnostic and Statistical Manual of Mental Disorders [1994], or DSM-IV) and/or result in traits which could constrain effective therapy. Although processes and traits could apply to both patients and therapists, with a few exceptions only their possible relevance to patients is discussed.

Co-evolution. "Rapid co-evolution arises from competition within species for more extreme traits" especially those that lead to winning social competition (Nesse 2005, p. 905; Alexander 1975) — for example, patients often deceive therapists and attempt to manipulate and control the therapeutic process – behaviors which can easily undermine effective therapy.

Tradeoffs. "Design trade-offs make the perfection of any trait impossible" (Nesse 2005, p. 905; Nesse and Williams 1994) — possible traits such as emotional instability, excessive dependence, and excessive deception may be 'imperfect' in the sense that they interfere with efforts to learn, revise tactics that have failed to achieve desired goals, and compromise efforts to achieve therapeutic change.

Constraints. "Systems shaped by natural selection are subject to special constraints, especially path dependence" (Nesse 2005, p. 906) — traits such as slow information processing, an inability to develop insight, and a lack of imagination, which are characteristic of many patients who enter therapy, may be consequences of path dependence.

Defenses. In the ideal sense, defenses are protective (Nesse 2005, p. 906). Yet often patients' defenses such as the fear of adopting new tactics, arranging an introduction to a possible mate, changing jobs, and skepticism about therapy slow or halt the pace of treatment and limit what therapy might accomplish.

Of course not all of the selection processes and/or traits apply to all patients (see also Fabrega 1997, 2002). Yet it is likely that for many patients some do and it is also likely that often they are contributing factors to why many patients seek treatment.

Hand-in-hand with the points above, there is a rapidly growing list of findings from the neurosciences which have clear implications for psychotherapy. Only selected findings are discussed here (see McGuire, in press, for a detailed review). For convenience we refer to these findings as "brain quirks" but they may also be viewed as possible constraints.

Memory. Studies consistently show that long-term memories are inaccurately reconstituted. This may be due to the fact that the brain's cells and chemicals that are responsible for storing memories differ from those that are responsible for subsequent recall (Lee et al. 2004). But other factors such as mood, lack of sleep, familiarity with what is recalled, and the frequency of recall are known to affect memory (Frankland et al. 2004, Bower 2003, Hairston and Knight 2004, Huber et al. 2004, Fenn et al. 2003, Walker et al. 2003, Miller 2004, Anderson et al. 2004). Studies of short-term memory capacities raise related issues. For example, research subjects who are briefly shown multiple objects on a screen can recall between 1 and 8 objects (Vogel and Machizawa 2004). Within subjects, short-term recall capacities remain remarkably consistent across research methodologies which suggests that the capacity is a trait, not a state. These findings are not surprising — experienced therapists know that there are significant individual differences in the accuracy and detail of what patients can recall. Nonetheless, to the degree that accurate and detailed recall is essential for effective therapy, differences in patients' memory capacities may foretell that some patients are unlikely to benefit from treatment.

A related issue involves the understanding and interpretation of memories. Interpretation may be primarily phenomenological (essentially, viewing things from within), which is characteristic of patients, or hermeneutic (essentially, viewing things from outside), which is characteristic of therapists (Ricoeur 2005). These very different ways of interpreting what patients report are also consistent with what therapists experience: patients and therapists differ, and frequently, in their understanding of patients' experiences, symptoms, states, motives, etc. Further, patients often experience difficulty in understanding and utilizing therapists' hermeneutically-based comments. To the degree that understanding what therapists say is critical for therapy, patients who lack this capacity may benefit minimally from treatment.

The tangled web of interpretation. If two adults view the same stimuli, such as a movie, the *initial* processing of the stimuli as assessed by functional MRI (fMRI) reveals a surprisingly high degree of cross-viewer similarity. Similarity increases when there is an unanticipated event such as a shooting (Pessoa 2004). (Recent studies demonstrate that neuronal location and activation in specific areas of the brain closely mirrors the degree of fMRI activation for the same brain area [Mukamel et al. 2005]). However, when the same two adults *interpret* the stimuli, again using fMRI activation as the basis for comparison, the similarity observed during the initial period of stimuli processing disappears (an outcome which is likely to apply to both patients and therapists) (Pessoa 2004). There is nothing new about the fact that people interpret the same stimuli differently or that these differences are likely to have multiple sources such as age and sex differences, different prior experiences, different beliefs, etc. (e.g., Haidt 2001) – pro- and anti-politically correct interpretations by same-age, same-sex, same-culture individuals observing the same event illustrate as much. Yet there are clear implications for therapy. For example, discussions about patient's lives as well as the events of therapy will likely be interpreted differently by patients and therapists – an implication that adds to the complexity of efforts by patients and therapists to understand each other.

Facial expressions and personality. If a happy face is shown to normal individuals there is minimal left and right amygdale activation as measured by fMRI activation (Canli et al. 2002). If fearful faces are shown, there is moderate activation. (Similar fMRI findings are reported for responses to strongly positive and negative words [Hamann and Mao 2002]). Extroverts respond differently however. When shown pictures of happy faces, strong left amygdale activation is observed (activation of the left amygdala is thought to be involved in the development of positive feelings towards others) while there is minimal amygdale activation in response to fearful faces (Canli et al. 2002, for related studies see Adams et al. 2003, Whalen et al. 2004). Several implications emerge from these findings. One is that patients' social environments will influence responses to and the interpretation of stimuli. Another is that therapists' efforts to accurately understand patients' social environments through patients' reports will be influenced and possibly distorted by in-built brain processing quirks. And a third is: if in-built brain quirks are traits, therapeutic options may be limited.

The brain's negative bias. Studies repeatedly show the brain has a negative bias – 'the glass half empty' view of the world (Marano 2003). A negative bias may help avoid difficulties – for example, it's usually wise to look before you leap. But it may also interfere with patients experimenting with new ways of interacting socially, adopting new tactics to achieve goals, or bringing a positive attitude to therapy.

Although brain quirks are likely products of evolution, some of their features are not – for example, the interpretation of stimuli is in part experience- and culture-influenced. Further, quirks don't apply to all patients in the same way. For each quirk individual differences are present and often strikingly so – recall the wide range of objects which individuals can recall (1 to 8) in short-term memory studies. In addition, the causes of differences vary. Nesse lists 12 factors such as gene-environment interactions which may contribute to these differences (Nesse 2005, see also Raser and O'Shea 2005 for a relevant discussion of "noise" associated with gene expression). Such differences are very real – for example, recovered depressed patients and healthy control subjects respond differently to criticism but not to praise (Hooley et al. 2005). Nesse's list of factors as well as numerous findings such as those of Hooley and colleagues invite the following prediction: the probability that identical factors contribute to similar individual differences among any two patients, even should they be diagnosed with the same DSM disorder or have the same compromised functional behavior, is literally zero.

For the purposes of this paper the combined list of constraints, traits, quirks, and individual differences suggests two main points. First, they may help explain why therapy sometimes fails although, in certain instances, they might do the opposite. For example, intermediate short-term memory capacities may facilitate therapy by avoiding an excessive focus on detail. Or the brain's negative bias may constrain some patients from acting on impulse until therapy can address such tendencies. Second, they lead to the question: how might a theory of psychotherapy incorporate the items on the list? Before trying to answer this question, other issues deserve consideration.

Other issues

Illness and disorder. There is a distinction between illness and disorder. Illness is the subjective state of individuals which changes very little over time – feelings of loneliness were probably similar in ancient Egypt and among the Incas to the way they are experienced today in Los Angeles, Cairo, and Tokyo. Disorders are another matter. They are the way cultures, individuals, or a profession such as psychiatry conceptualize and categorize illness. In psychiatry, what is considered a mental disorder changes over time – compare DSM-I with DSM-IV. While such changes are expected with any evolving taxonomic system there are different reasons for their changes. For example, changes in biological taxonomies are based on new information (e.g., findings in molecular genetics) whereas changes in psychiatric taxonomy are the products of arbitrary

decisions by experts. For psychiatry, such changes introduce an element of uncertainty regarding diagnosis and treatment. Further, there is a long list of difficulties such as shame, fear, and emotional instability which are observed in association with many DSM diagnoses. These not only add to diagnostic uncertainty and the complexity of therapy but they also introduce obstacles to identifying the causes of illnesses – for example, the symptoms of fever, nausea, and weight loss tell us only that some one is ill, not the cause of the illness.

Further, in psychiatry, the tendency to categorize many signs, symptoms, and/or behaviors as disorders is excessive — if one adds up the DSM-IV listed percentages of children who supposedly suffer from DSM-IV defined disorders the total is close to 64% of the child population of the United States (American Psychiatric Association 1994, McGuire and Troisi 1998).

Excessive categorizing can be given an informative perspective by looking at the way we view other species. Consider dogs. All types exist. And both within- and across-species their behaviors and temperaments differ, often significantly. Rarely are these differences considered "abnormal" or assigned to some disorder-like category even though they often have genetic bases (Lindblad-Toh et al. 2005). Moreover such differences are often difficult to change. In contrast, with humans we are much more specific and judgmental with regard to what is considered "normal" and what we categorize as a mental disorder. Although this specificity may have its origins in patients' reports of their illnesses, it also has cultural roots as multiple cross-cultural and historical studies have shown.

Our point is not that mental disorders are cultural myths. It's other points we wish to stress. First, categorizing patients' illnesses as disorders can serve as an obstacle to the development of falsifiable hypotheses. This is particularly so if disorder categories change, if their definitions are imprecise, or when therapists differ in their diagnoses as they often do.

Second, from the perspective of an evolutionary-based theory of psychotherapy, traits, states, quirks, and constraints are likely to be more effectively studied by adopting an individual difference perspective. This approach would facilitate identifying if and when these attributes are 'intervention sensitive' within and across patients. For example, therapy is unlikely to change slow information processors into fast processors or vice versa. Or, for many patients, emotional instability may be a rock-solid trait and unresponsive to any type of social intervention. Yet for other patients it may be responsive to treatment. These points lead to the question: what should be measured in therapy? We address this question in a companion paper (McGuire and Troisi 2006) although it is perhaps clear at this point that change in functional behaviors will take center stage if what is treated and measured is to be consistent with evolutionary concepts.

Third, many studies designed to assess the efficacy of therapy are difficult to interpret. This is often obvious when therapists treat patients with drugs that are postulated to alleviate specific disorders. In literally all instances only a percentage of those treated meet improvement criteria. Then the question arises: did those patients who didn't improve suffer from the same disorder as those who did? Similar interpretative difficulties apply to psychotherapy interventions. Sometimes patients' attributes change, sometimes not, and sometimes they get worse. Did they suffer from different disorders? Was classifying patients with the same disorder relevant to the research? Was it different therapist techniques that explain the outcome? Etc.

In our view these points add up to the strong possibility that attempts to take account of possible evolved constraints, traits and states, brain quirks, diagnostic categories, and so forth introduces a complexity that is impossible to manage intellectually let alone research successfully. A mind-set change is needed.

Explanations that defy falsification. Psychiatry has a truly enviable record of generating hypotheses and explanations that defy falsification and which at times mislead intuition (e.g., Hobson 2005) – here the term "explanation" refers to statements that range from those that are highly consistent with existing data and theory to those that are far more speculative. The "epigenetic inheritance" hypothesis – "the transmission to offspring of parental phenotypic responses to environmental challenges" – is perhaps an example (Harper 2005, p. 340). Or, consider the social navigation hypothesis of depression (Watson & Andrews 2002). It postulates that depression evolved to perform two functions: depression induces cognitive changes that facilitates the solution of social problems and the somatic and psychological costs of depression may induce social partners to help or make concessions. We have no particular difficulty with this hypothesis. It is intuitively attractive. But questions arise. Does the hypothesis have therapeutic utility? Most important for our purposes, is it falsifiable? Clear answers to these questions seem unlikely because what might be measured is unclear or methodologically unmanageable, or both.

Much the same can be said about studies which focus on postulated states such as countertransference where, for example, physician responses such as anger and feelings of impotence to treating patients classified as suffering from Narcissistic Personality Disorder are taken as indications of countertransference (Betan et al. 2005). The explanation could be correct. But physician responses might also be indications of frustration associated with their inability to initiate change both while treating patients and in other activities. For example, similar responses occur when one has failed to fix an automobile, a bicycle, or an electric train and in these instances countertransference is an unlikely cause. Again, falsification is the issue.

And, on a related issue, therapeutic outcome in the treatment of illnesses has little to do with the falsification of etiological or pathogenetic theories. Indeed, positive outcomes be misleading — findings that placebo administration is at times an effective form of treatment suggest this conclusion. Further, the often observed tendency to assume that positive treatment outcomes confirm etiological theories is frequently enhanced by the development of complex intervention algorithms which posit multiple possible yet often contradictory pathways for interventions for supposedly the same disorder. Seldom do the authors of these schemes recognize that what is being postulated is multiple etiologies (e.g., see Davidson et al. 2005).

Adaptive and maladaptive behavior. While there

are many explanations of maladaptive features of DSM classified disorders, they largely defy falsification (see Dubrovsky 2002 for a discussion of related points). These are interesting and often intuitively attractive explanations which may be informative regarding the causes and consequences of behavior (e.g., Surbey 1987, Voland and Voland 1989, Hagen 2001, McGuire and Troisi 1998). Nonetheless they are seldom free from the above mentioned cultural influences on disorder classification – the seeming unending controversy about whether or not certain forms of ADHD qualify as a mental disorder or reflect the cultural labeling of evolved human individual variation is perhaps an example — see Gilbert (this volume) for a related discussion about certain forms of depression. From the perspective taken here, it is not at all axiomatic that a symptom, a sign, or even a cluster of symptoms and signs have implications beyond saying that they are manifestations of within-species variation.

Supposed maladaptive behavior hypotheses might play an important role in a theory of therapy provided there was evidence that these behaviors predict reduced survival rates and/or reproductive success and/or kin investment. For the majority of such behaviors as well as DSM classified disorders there are few studies demonstrating clear relationships – an exception is found in reports of reduced reproductive rates among persons labeled as schizophrenic (Howard 2005, McGrath et al. 1999). And, at times, interpretations seem contradictory – for example, one has to be alive and to have reproduced to qualify as having Post-partum Depression (e.g., Hagen 2002) and individuals labeled as suffering from Antisocial Personality Disorder often leave a string of children in towns from New York to San Francisco.

Efforts to integrate evolutionary hypotheses and explanations with psychiatry's current diagnostic system (DSM) may be much like trying to stick square pegs into round holes. More important, such efforts may overlook the fact that that there are alternative ways of thinking about behaviors which are labeled as mental disorders but which are very understandable from an evolutionary perspective. For example, consider the concept of evolved alternative functional behaviors (often referred to as "evolved alternative strategies") to achieve the same biological goal: "two or more discrete behavioral variants among adults of one sex and one population when these variants serve the same functional ends" (Troisi 2005, p.159). Such variants are observed in other species (e.g., Brockmann 2001). Thus, for example, behaviors which are classified as Antisocial Personality Disorder, Insecure Attachment Disorder, and certain features of depression may be more informatively characterized as evolved or learned alternative strategies which serve the same functional ends as do supposed 'normal' behaviors (Troisi 2005, Belsky 1999, Gilbert this volume). Should studies show that survival rates and reproductive success are not compromised with these behaviors, all the more reason to view them as evolved alternative strategies. Of course psychiatrists might still classify them as disorders and judges might still send persons labeled as antisocial personalities to prison. But here our concern is with a theory of therapy and falsifiable hypotheses that are consistent with evolutionary arguments and such issues as: if there are functionally successful variants is therapy likely to alter them?

Intervening event explanations. Empirically people often don't respond with symptoms of depression immediately following a significant personal loss – there is a delay. A similar delay is often observed with the onset of anxiety – it may develop well after the event that supposedly triggeres it. The same delay is seen in response to adverse experiences. Intervening events in the brain are thought to explain such things as event-response delays, imprecise recall, the failure to revise tactics that have not worked, etc. Typical examples of intervening events include repression, self-deception, and intelligence. In psychiatry, the attraction of intervening events appears to be largely the legacy of psychoanalytic theory where, for example, repressed events and/or unrecognized mechanisms are postulated to affect current behavior — see Gilbert (this volume) for a historical discussion of many of these points. The idea is attractive and we have no doubt that intervening events occur. Nor do we doubt that they affect behavior, cognition, emotion, therapy outcome, etc. – could it be otherwise? Yet they may also exist only theoretically and have nothing to do with structures or processes or behavior. More important for this paper, is there a methodology that permits falsification of intervening event explanations and hypotheses? We are skeptical. A century of hypothesis development and research has done little to establish that these events, at least when they are defined in psychological terms, can be falsified convincingly. Perhaps this is why research which measures physiological and behavioral change in response to drug administration is so popular currently – physiological change can be though of as an intervening event.

That many postulates about intervening events may be wrong seems inevitable. The subject of tactics to achieve goals provides a convenient example. The default paradigm for tactics is that individuals adopt and act on them to achieve goals. If a goal is achieved the same tactic is likely to be used again. If it is not achieved, one supposedly recognizes this and makes adjustments or devises new tactics – in effect, a type of trial-and-error decision-making process. Many patients are thought to be unable to recognize that their tactics (as distinct from "strategizing" which is the act of identifying, valuing, and prioritizing the acquisition of an object) fail to achieve their desired goals and/or to make adjustments when their tactics are ineffective – hence one possible source of repetitive behavior. But recent research raises questions about the relevance of the default model and the hypothesis that there are in-built self-correcting processes among even normal subjects. For example, Johanasson et al. (2005) failed to detect mismatches between intention and outcome in simple decision tasks although study subjects were quick to offer introspective derived reasons for the mismatches.

To be clear, let us draw attention to what we are not saying. We are not suggesting that therapists should stop thinking about intervening events. This won't happen – trying to explain uncertainties is a human trait. Nor are we suggesting that discussions about intervening events should be excluded from therapy – they may be helpful in assisting patients to develop useful concepts about why they feel and behave as they do. Our

point is that these events have questionable value for a theory of therapy or the development of falsifiable hypotheses.

What should be measured to determine if treatment results in change? The list of possible measures is long and often ambiguous. For example are symptoms or behavior more critical to measure? Is subjective (phenomenological) improvement a better measure than therapists' hermeneutic assessments? What about cognitive change? And how is it best to deal with symptoms and signs that often change on their own – that is, without the benefit of treatment? Etc. Not unexpectedly, therapists frequently disagree over which measures are important (Bower 2005).

Before addressing the measurement question it is important to revisit functional behaviors. As Feierman (this volume) points out, behaviors are defined by their function. We agree. However, such a definition raises obvious problems for the task being addressed here. One is the vexing problem of variance. There is, for example, a great deal of variance in the behavior associated with finding a mate. Likewise with reciprocal exchanges and especially so when the 'currency' of exchanges differ – for example, A cares for B's children for a day and B repays the favor by obtaining an invitation for A to join the exclusive local garden club. That all of these possible favor-repayment variations can be sorted out prior to or even after therapy seems unlikely. Allowing for this variance leads in one direction: namely, a critical requirement for viewing a functional behavior as compromised is that it consistently fails to achieve its objective. For example, one might successfully introduce one's self to a potential mate via a letter, email, a telephone call, or a face-to-face introduction arranged by a friend. But if one only sends emails and if there is no response, one's objective has not been achieved.

Measures of course differ depending on the hypotheses that are being tested. As informative as a discussion of the hypothesis-measure relationship might be there are far too many points to cover to do it justice with but one exception: paper-and-pencil tests designed to assess symptom states, personality, etc. We mention these types of assessments – we refer to them as "indirect measurement tools" – because they are used frequently to assess intervention effectiveness: an assessment is administered before therapy → therapy is provided → an assessment is administered following therapy → findings from the first and second assessments are compared → conclusions are drawn. Are the assessment tools valid? Can the conclusions be believed? And how might such questions be answered from an evolutionary perspective? For example, consider the evolutionary explanations of self-deception and deception (Alexander 1975; Trivers 1976, 2000). Might test responses be influenced by deception and self-deception? Might most deceptions go undetected which over 60 years of research suggests is the case (Lock 2004)? And what of the possible influence of brain quirks?

The preceding questions notwithstanding, indirect measurement tools have merit in that they attempt to introduce systematic evaluation and/or intervention outcome measures to the study of illness. But they often have their origins in concepts that are difficult to map to or integrate with evolutionary concepts. For example, in our view, there is little in tools such as the SCL-90, the MMPI, or the Beck Depression Scale that are either consistent with or refute evolutionary concepts.

Therapist strategies and tactics. Like patients, therapists differ in their therapeutic strategies and tactics. Training may minimize such differences among therapists, yet differences remain. What seems to happen is that therapists develop 'styles' which 'fit' with some patients but not with others and their styles in part reflect their views about what needs to be done to alleviate patients' illnesses. Although there are commonalities in what therapists do – they listen, they focus on patient's concerns and symptoms, etc. – differences in techniques are probably greater and more consequential than is usually acknowledged. Such differences have obvious implications for intervention testing because they mean that interventions will be difficult to standardize.

How psychotherapy works. As a final issue, do we know how psychotherapy works? This is an important question because it may influence how a theory of therapy develops. For example, is psychotherapy primarily a learning experience – if so, then should learning be measured? Is the relationship between the therapist and patient as critical as usually claimed – if so, should the relationship be measured? Compared to states, are evolved traits more or less likely to respond to certain types of therapeutic interventions? These are questions for which there are few solid answers.

Discussion

Several themes have guided our thinking.

The distinctions between illness, disorder, and functional behavior is one. Psychotherapy treats illnesses or functionally defined behaviors, not disorders although therapists often discuss what they are doing as if it's disorders that are being treated. The same point applies to pharmacological interventions – it's illnesses that are treated. Or, as Gilbert (this volume) puts it: psychotherapy is about treating and changing patients' "phenotypic profiles". Our point here is not that people who are labeled as having a disorder are not suffering. They usually are. Nor are we suggesting that therapists not discuss disorder labels with patients – often patients are aware of how they are labeled. What we are suggesting is that making clear distinctions between illness, disorder labeling, and functional behavior are essential steps in developing an internally consistent evolutionary theory of psychotherapy and falsifiable hypotheses.

Individual difference has been another guide. Any two persons with the same formal disorder classification, even with similar sign-symptom profiles, are unlikely to have the same, perhaps even similar, evolved constraints, brain quirks, past experiences, motivations to achieve goals, etc. For example, there are usually clearly discernable differences in the severity of literally every symptom, sign, behavior, etc. among patients who are classified as suffering from the same disorder. Further, it is not at all clear how such differences influ-

ence therapy. Their affect may be positive or negative. They may facilitate or impede treatment. In short, individual differences are an inescapable fact no matter what one's views about psychotherapy. There is nothing surprising about this point. Yet it does lead to all kinds of concerns with respect to hypothesis testing — if patients and therapists differ, and all do, how is it possible to systematically study intervention effects?

A related theme concerns maladaptive characterizations of behavior. These are often influenced by psychiatry's current disorder classification system (DSM). How such characterizations might be incorporated into a theory of therapy, if at all, is not clear. For example, do brief and extended periods of jealousy differ? Are brief periods of jealousy adaptive and extended periods maladaptive? Where is the cutoff point?

Another theme is that measures that are commonly used to assess therapy outcome may have little to do with evolutionary concepts or even easily map to them. Sign and symptom change seem the least ambiguous of the measures that are used although clearly in most instances reports about symptom change are subjective and symptoms and signs may change independently of therapy. Indirect assessment tools seem more questionable. And as we have asked: how relevant are many of the commonly used measures in an evolutionary biological context?

We are faced with an incredible and in our view a unmanageable complexity in the ways behavior is explained. In part this is because human nature is complex, in part because our capacity to develop complex explanations – for example explanations of intervening events — far exceeds our capacity and/or willingness to try to test them. For each disorder and therapy there are literally hundreds of articles and books (e.g., Gilbert 1989, Sloman and Gilbert 2000) which document the highly complex nature of illnesses (e.g., Gilbert 1999) as they play out across patients — and there is no end in sight.

The net effect is that if one's aim is to develop an evolutionary-based theory of therapy and falsifiable hypotheses it may not be possible to do so within the prevailing diagnostic and conceptual framework of psychiatry.

So what is it that we recommend? A companion paper (McGuire and Troisi 2006b) offers an answer to this question. The answer has four parts: disregard much of the thinking that prevails currently in psychiatry; develop a simple evolutionary theory of therapy; generate truly falsifiable hypotheses that are consistent with evolutionary concepts; test the hypotheses using measures that can be reliably assessed.

Final points. We do not anticipate that the points developed in this paper will bring uniform agreement – disagreement seems more likely. For example, in reviewing an early draft of this paper, J.O. Beahrs (personal communication) argued strongly that, because "psychological realities" vary with patients' social contexts, traditional scientific concepts of cause and effect, which are based on context-free objectivity, are not appropriate for the study of psychotherapy. Elsewhere he writes: "Different causal rules must therefore apply" (Beahrs 2005, p. 817). He may be correct. What is in part at issue here are assumptions. Clearly, our assumptions are mechanistic: the same causal rules apply today as yesterday and context can be a causal factor. Beahrs' assumptions are organismic: essentially, systems evolve and as they do causal rules may change. These are two very different ways of explaining behavior (Pepper 1942).

Further, we have not fully addressed the complexity of many of the issues we have discussed. But if the points developed in this paper initiate a dialogue about a theory of therapy one of our aims will be accomplished. Finally, we say again that the citations we have used in our discussion are positive, not negative examples of clinicians and researches developing ideas about how to explain and study human nature.

This is a thought piece.

Acknowledgements

Comments on earlier drafts provided by John Beahrs and Paul Gilbert and especially those of Jay Feierman were very helpful in the development of this paper.

References

Adams Jr, RB, Gordon HL, Baird AA, et al. (2003). Effects of gaze on amygdala sensitivity to anger and fear faces. *Science* 300, 1536.

Alexander RD (1975) The search for a general theory of behavior. *Behavioral Science* 20, 77-100.

American Psychiatric Association (1994). *Diagnostic and Statistical Manual of Mental Disorders*, 4th Edition. American Psychiatric Association. Washington DC.

Anderson MC, Ochsner KN, Kuhl B, et al. (2004) Neural systems underlying the suppression of unwanted memories. *Science* 303, 232-235.

Betan B (2005) Countertransference phenomena and personality pathology in clinical practice: an empirical investigation. *American Journal of Psychiatry* 162, 890-898.

Belsky J (1999) Modern evolutionary theory and patterns of attachment. In J Cassidy and PR Shaver (eds) *Handbook of Attachment: Theory, Research, and Clinical Application*, p. 141-161. Guilford Press, New York.

Brockmann HJ (2001). The evolution of alternative strategies and tactics. *Advancements in the Study of Behavior* 30, 1-51.

Bower B (2003). Forgetting to remember. *ScienceNews* 164, 293.

Bower B (2005). Questions on the couch. *ScienceNews* 168, 299-301.

Canli T, Sivers H, Whitfield SL, et al. (2002). Amygdala response to happy faces as a function of extraversion. *Science* 296, 2191.

Davidson J, Bernik M, Connor KM, et al. (2005). A new treatment algorithm for posttraumatic stress disorder. *Psychiatric Annals* 35, 887-898.

Dubrovsky B (2002). Evolutionary psychiatry. Adaptationist and nonadaptationist conceptualizations. *Progress in Neuro-Psychopharmacology & Biological Psychiatry* 26, 1-19.

Fabrega H Jr (1997). *Evolution of Sickness and Healing*. University of California Press, Berkeley.

Fabrega H Jr. (2002). *Origins of Psychopathology*. Rutgers University Press, New Brunswick, New Jersey.

Fenn KM, Nusbaum HC & Margoliash D (2003). Consolidation during sleep of perceptual learning of spoken language. *Nature* 425, 614-616.

Frankland PW, Bontempi B, Talton LE, et al. (2004). The involvement of the anterior cingulated cortex in remote contextual fear memory. *Science* 304, 881-883.

Gilbert P (1999). *Overcoming Depression*. Oxford University

Press, New York.
Gilbert P (1989). *Human Nature and Suffering*. Lawrence Erlbaum, Hove, UK.
Hagen EH (2002). Depression as bargaining: the case of postpartum depression. *Evolution & Human Behavior* 23, 323-336.
Haidt J (2001). The emotional dog and its rational tail: a social intuitionist approach to moral judgment. *Psychological Reviews* 108, 814-834.
Hamann S & Mao H (2002). Positive and negative emotional verbal stimuli elicit activity in the left amygdala. *Neuroreport* 13, 15-19.
Hairsto IS & Knight RT (2004). Sleep on it. *Nature* 430, 27-28.
Harper LV (2005). Epigenetic inheritance and the intergenerational transfer of experience. *Psychological Bulletin* 131, 340-360.
Hobson JA (2005). Sleep is of the brain, by the brain and for the brain. *Nature* 437, 1254-1256.
Hooley JM, Gruber SA, Scott LA, et al. (2005). Activation in dorsolateral prefrontal cortex in response to maternal criticism and praise in recovered depressed and healthy control participants. *Biological Psychiatry* 57, 809-812.
Howard LM (2005). Fertility and pregnancy in women with psychotic disorders. *European Journal Obstetrics, Gynecology & Reproductive Biology* 119, 3-10.
Huber R, Ghilardi MF, Massimini M, et al. (2004). Local sleep and learning. *Nature* 430, 78-81.
Johansson P, Hall L, Sikstrom S, et al. (2005). Failure to detect mismatches between intention and outcome in a simple decision task. *Science* 310, 116-119.
Lee JLC, Everitt BJ & Thomas KL (2004). Independent cellular processes for hippocampal memory consolidation and reconsolidation. *Science* 304, 839-843.
Lindblad-Toh K, Wade CM, Mikkelsen TS, et al. (2005). Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature* 438, 803-819.
Lock C (2004). Deception detection. *ScienceNews* 166, 72-73.
McGrath JJ, Hearle J, Jenner L et al. (1999) The fertility and fecundity of patients with psychoses. *Acta Psychiatrica Scandinavica* 99, 441-446.
McGuire MT & Troisi A (1998). *Darwinian Psychiatry*. Oxford University Press, New York.
McGuire MT and Troisi A (2006). Towards an evolutionary-based theory of psychotherapy - II. (This volume).
Marano H (2003). Our brain's negative bias. *Psychology Today*. June 20.
Miller G (2004). Learning to forget. *Science* 304, 34-36.
Mukamel R, Gelbard H, Arieli A, et al. (2005). Coupling between neuronal firing, field potentials, and fMRI in human auditory cortex. *Science* 309, 951-954.
Nelissen K, Luppino G, Vanduffel W, et al. (2005). Observing others: multiple action representation in the frontal lobe. *Science* 310, 332-336.
Nesse RM (2005). Evolutionary psychology and mental health. In D Buss (ed) *Handbook of Evolutionary Psychology*, p. 903-927. Wiley, Hoboken, New Jersey.
Nesse RM & Williams GC (1994). *Why We Get Sick*. Times Books, New York.
Pepper S (1942). *World Hypotheses*. University of California Press, Berkeley.
Pessoa L (2004). Seeing the world the same way. *Science* 303, 1617-1618.
Raser JM & O'Shea EK (2005). Noise in gene expression: origins, consequences, and control. *Science* 309, 2010-2013.
Ricoeur P (2005). *Memory, History, Forgetting*. University of Chicago Press, Chicago.
Sloman L & Gilbert P (eds) (2000). Subordination and Defeat. Lawrence Erlbaum, Mahway, New Jersey.
Surbey MK (1987). Anorexia nervosa, amenorrhea, and adaptation. *Ethology & Sociobiology* 8, 12S-18S.
Trivers R (1976). Forward. In R Dawkins, *The Selfish Gene*. Oxford University Press, New York.
Trivers R (2000). The elements of a scientific theory of self-deception. *Annals of the New York Academy of Sciences* 907, 114-131.
Troisi A (2005). The concept of alternative strategies and its relevance to psychiatry and clinical psychology. *Neuroscience and Biobehavioral Reviews* 29, 159-168.
Vogel EK and Machizawa MG (2004). Neural activity predicts individual differences in visual working memory capacity. *Nature* 428, 748-750.
Voland E & Voland R (1989). Evolutionary biology and psychiatry: the case for anorexia nervosa. *Ethology and Sociobiology* 10, 223-240.
Walker MP, Brakefield T, Hobson JA, et al. (2003). Dissociable states of human memory consolidation and reconsolidation. *Nature* 425, 616-620
Watson PJ & Andrews PW (2002). Toward a revised evolutionary adaptationist analysis of depression: the social navigation hypothesis. *Journal Affective Disorders* 72, 1-14.
Whalen PJ, Kagan J, Cook RG, et al. (2004). Human amygdala responsivity to masked fearful eye whites. *Science* 306, 2061.