

Narrative Fallacy and Other Limitations of Psychodynamic Case Formulation

Alexey Tolchinsky

Professional Psychology Program

The George Washington University

Author Note

Alexey Tolchinsky  <https://orcid.org/0009-0009-4721-515X>

Alexey Tolchinsky is at the Professional Psychology Program,
The George Washington University.

I have no conflicts of interest to disclose

I am grateful to Nancy McWilliams, Loring Ingraham, and Alex Smith for their valuable
comments.

Correspondence concerning this article should be addressed to

Alexey Tolchinsky, 1922 F. St., NW, Suite 100, Washington, DC, 20052.

Email: alexeyt@gwu.edu

©American Psychological Association, 2023. This paper is not the copy of record and may not
exactly replicate the authoritative document published in the APA journal. The final article is
available at: <https://doi.org/10.1037/pri0000231>

Abstract

This paper describes a phenomenon of narrative fallacy in psychodynamic psychotherapy and makes a case that it is likely related to therapists' working memory limitations. The impact of narrative fallacy on model complexity is examined with a focus on further improving the efficacy of clinical treatment. The specifics of the psychodynamic meaning-making process gravitating toward searching for causal or teleological explanations are reviewed. Tentative perspectives on these processes from the standpoint of active inference and neuropsychology are presented. This paper also examines the issues with verifiability in the psychodynamic hypothesis testing process with a focus on the Barnum/Forer effect. Preliminary proposals are formulated to help mitigate the impact of these issues on clinical practice.

Keywords: narrative fallacy; case formulation; active inference; verifiability, working memory

Clinical Impact Statement: This paper describes the limitations of psychodynamic psychotherapy related to the narrative style of conceptualizing clinical issues. Measures designed to address these challenges and potentially improve clinical efficacy for hard-to-treat conditions are proposed.

The efficacy of psychodynamic psychotherapy is established for some clinical conditions in rigorous meta-analyses (Shedler, 2010; Leichsenring et al. 2015); however, evidence is lacking for post-traumatic, obsessive-compulsive, bipolar, and schizophrenia spectrum disorders (Leichsenring et al. 2015). This paper suggests that one of the possible reasons for such lack of evidence is that our models of these conditions are insufficiently complex, which results in suboptimal accuracy in our understanding of these phenomena. Specifically, this manuscript focuses on narrativity in psychodynamic case formulations, which creates a limit on psychodynamic model complexity.

To clarify, this paper does not aim to argue the importance of accuracy or truth in a philosophical sense (Wallace, 1988), nor does it advocate for increased complexity for complexity's sake. Instead, it advocates for data-fitting complexity and accuracy as they relate to clinical efficacy. For example, models of psychopathology based on a deficit of neurotransmitter N appear to reduce dynamic phenomena at a macro level of the brain-mind to a molecular level, which is an excessive simplification (Sulis, 2021). This could be one of the possible reasons why the effect size of pharmacological treatment of depression is relatively small (0.31, Turner et al. 2008, as cited in Shedler, 2010).

Therefore, this paper highlights the importance of how we collect and process clinical information in any treatment, including psychodynamic psychotherapy, and how we conceptualize the patient's issues to minimize the probability of being significantly off-target. It is in this sense that sufficient model accuracy and complexity are essential.

The primary tool for conceptualizing the patient's distress in psychodynamic psychotherapy is case formulation. It is a complex process, which requires theoretical

knowledge, supervision, and experience (McWilliams, 2021). McWilliam's (1999) book "Psychoanalytic case formulation" is widely regarded as one of the best sources on the subject, and it is often used as a textbook by therapists in training.

This paper is an attempt to examine some of the features and limitations of the psychodynamic case formulation process from the standpoint of active inference (Parr et al., 2022) and neuropsychological models of working memory and language processing. Possible suggestions on what could be done to mitigate these issues will be discussed.

The list of the case formulation's features and limitations in this paper is incomplete and subjective, although contextualized in the relevant literature to the degree possible. Instead of presenting a comprehensive review of all aspects of case formulation, this paper attempts to focus on the specific issues related to what is described below as "narrative fallacy (Taleb, 2007)," which seems to create an upper limit on the psychodynamic model complexity.

The craft

A psychodynamic case formulation is, among other things, a story (e.g., Freud's remark, 1901, p. 160; Schafer, 1992). The therapist weaves various data points into a coherent narrative (e.g., Joseph, 1985, pp. 448-450; Mitchell & Black, 1996, p.184). One of the ways this narrative is used by both the patient and the therapist is to make sense of the patient's issues (McWilliams, 1999, p. 11). Sense-making implies, among other things, formulating a tentative explanation of the patient's presenting problem, or at least contextualizing it.

A case formulation typically contains a nuanced "psychological portrait" of the patient and a temporal sequence of psychological events in the patient's life (e.g., McWilliams, 2004, pp. 220-224). The narrative may also contain representations of other individuals or entities in the patient's life, such as their job. Mental representations of the patient's self and its various components

(wishes, fears, impulses, etc.) and the representation of other people, objects, and events are some of the building blocks in the story, the characters.

During the work in therapy, the therapist and the patient also formulate tentative hypotheses about the relationships between these characters. These relationships contribute to meaning-making, as we will see later in this paper.

In a psychodynamic case formulation, both the relationships and the mental representations are dynamic – they evolve with time and form complex clusters and configurations, such as being in a state of conflict (Mitchell & Black, p. 16).

While “painting” the psychological portrait of the patient with words, the therapist typically uses a “palette” of colors in psychological space, which may include continuous or discrete variables, such as “character style” (McWilliams, 2011), psychoanalytic level of functioning, diagnostic labels, and so forth.

In addition to the patient’s mental representations, psychodynamic constructs nearly always become the building blocks in the case formulation (e.g., Joseph, 1985), for example: “John has a conflict between a wish to assert himself with his boss and an anxiety to lose his job. He displaces his anger onto a safer target – his son.” Here, the therapist first inferred about a relationship between the two mental objects – John’s wish and his anxiety. In the second sentence, the assumed consequence of the first one was described. The first sentence presented a problem – an unresolved conflict. The second one was an inference on how John’s mind goes about trying to “solve” it. Together, a problem and a solution present a coherent and meaningful narrative, which can be easily digested by the reader familiar with psychodynamic terminology, or it can be translated into the language that the patient is familiar with.

The case formulation may contain a working explanation of the patient's presenting problem. Causality can be linear (e.g., trauma causing depression) or circular (e.g., depression exacerbating sleep disorder, which is exacerbating depression, thus creating a cycle.) Symptoms can be multiply determined (e.g., depression being caused by a loss of a job and genetic predisposition, chronic pain, and sleep disturbance).

When completed, a tentative formulation becomes a model, which can be used to guide the therapeutic interventions and the process of psychotherapy. The therapeutic interventions based on the current formulation are carefully timed and tactfully composed, hopefully leading to the patient's insight, and, subsequently, to the working through (Gabbard, 2005, p. 106) phase of psychodynamic psychotherapy.

Historical perspective on narrativity in case formulations

In the introduction to his book "The Feeling Brain," Solms (2018) discussed a comparison of case histories and stories and provided the following citation from Freud and Breuer's *Studies on Hysteria*:

It still strikes me myself as strange that the case histories I write should read like short stories and that, as one might say, they lack the serious stamp of science. I must console myself with the reflection that the nature of the subject is evidently responsible for this, rather than any preference of my own. (p. 160)

Narrativity in psychodynamic psychotherapy was elaborated by Roy Schafer (1992) and other authors, please refer to Mitchell and Black (1996) for a more complete historical account.

While Freud warned us about the "arrogance of the present" – the ease with which people can criticize statements made more than a century ago – must we continue to be content with

Freud's statement that the "nature of the subject is responsible for this?" In other words, do we agree that narrative is a sufficient description of the nature of the mind?

What happens when we create stories?

Structure

Case formulation stories are structured temporally (e.g., Mitchell and Black, p. 184). For example, they may begin with early childhood, or with an onset of symptoms, and they unfold in time, leading up to the intake interview. There is usually one line of the narrative. Even when the therapist weaves in multiple lines of the narrative, due to the serial nature of language processing (unlike the parallel processing by the CPU in a computer), only one of the threads is considered at any given moment in time; and for the duration of that period, we hear a linear, sequenced structure.

A reasonable observation here is that such is the very nature of "talk therapy." One of the possible interventions – the act of collaborative naming of the patient's distress in the therapeutic dyad, putting something raw and "unnamed" into concepts, sentences, and paragraphs - is, among other things, putting things in order (Mitchell & Black, 1996, pp 19-35). Each word is a semantic memory – an abstract representation and a category. Thus, naming the patient's vague distress as "shame" is a classification, a categorical choice. Using the sequential and syntactical order of words in a sentence is also ordering, as is sequencing sentences together into a paragraph.

We believe in psychodynamic therapy that such collaborative conceptualization of the unnamed, raw, and preconscious or unconscious material is a vital component in therapy. This is a complex topic and much has been said in support of "representing" the raw material in words (consider a nuanced discussion of Freud's, Sullivan's, Stern's, and Loewald's views on this topic in Mitchell & Black, 1996, pp. 187-188). Arguably, using Freud's metaphor on another topic,

working directly with the raw, unnamed material is akin to “fighting an enemy in absentia.” However, we do not sufficiently discuss the limitations of the audio-verbal method of collecting data about the patient’s distress and analyzing such data.

Stabilization

Another effect that case formulation stories have is stabilization. As an illustration, imagine turbulent white-water rapids. They flow in front of you. Imagine further, that you decided to take a photograph of the view with a camera. You may get different effects depending on your chosen camera’s shutter speed, but no matter how you take this picture, you will “stop time” in the frame.

Low-level neuronal processes resemble a waterfall, they flow (Parr et al., 2022). At the higher levels of the brain-mind hierarchy, we have both the flows, such as emotional cascades, and temporarily static objects, such as thoughts or images that we maintain in the working memory for some time; we also have static categories describing some of the brain-mind states, such as coma, wakefulness, REM sleep, dreaming, and so forth.

In psychodynamic psychotherapy we call our case formulations “dynamic,” and certainly “mind in a conflict” is a more dynamic approach than seeing the patient’s distress as a deficit in skills or information. However, when we consider the brain-mind at a temporal scale of seconds, the verbal case formulation might not be any more dynamic than a picture of a waterfall – it does a good job describing the rocks, and, perhaps, the speed of the flow, but it is not the flow, nor does it use the tools of modeling flows as described in the *Active inference perspective on emotional dynamics* section below.

Studying the quality of flows in the brain-mind with words or sentences is a limited method; a consideration to add other tools to the work might be justified, as described in the *What can be done* section below.

Retrospective view

A related issue to the stabilization is retrospection. In the process of the work in therapy, the patient may share what they are thinking and feeling nearly in the present. However, the case formulations are retrospective in nature (e.g., Mitchell & Black, p. 234). Typically, some factors in the patient's past are deemed to be formative in explaining his or her current presenting problem and these factors have been conveyed to the therapist retrospectively.

A photograph of a waterfall is always a picture of something in the past. For case formulation, there is a specific feature of retrospection – the story is being told about the events in the past by a storyteller, who already knows what happened after these events took place. Unlike a photograph of a waterfall, which is just averaging dynamic information, there is a level of distortion in such storytelling. Taleb (2007) calls it a retrospective distortion. Stories tend to differ significantly when the narrator describes the events as they are unfolding and when she tells the story of what happened in the past from the vantage point of the present (Taleb, 2007).

Coherence

Coherent narratives require less effort to comprehend than incoherent ones (Thorn and Page, 2008), which makes them preferable and, possibly, comforting, both to the therapists and the patients as compared to the disjointed pieces of data. For example, the patient may be temporarily comforted with a coherent story that seems to explain his distress, more so than he would be with a list of five possible causal factors with probabilities assigned to each one. Please refer to the *Neuropsychological perspective* section below for more details.

Meaning

Coherence is not sufficient to make a formulation meaningful for the patient. One of the possible down-up viewpoints on meaning expressed by Freeman (Stanford Complexity Group, 2013) is the difference between information and knowledge. To use Freeman's example, when a female fox exhales, she is putting information in the air about her existence. This information means two different things for a male fox and a hare. The hare's knowledge associated with this information has the meaning of "run," while the meaning for a male fox might be "approach." In a top-down view, the semantic meaning of a word designates a concrete object it represents. The word "apple" is thus meaningless to a baby, who never saw or touched an apple. Quantum entanglement is similarly meaningless to those who do not understand the basics of quantum mechanics.

While a comprehensive, philosophical definition of meaning is beyond the scope of this paper, perhaps, we can agree that a patient can reduce uncertainty, metabolize, and consolidate the data from working memory into long term memory (LTM) when she recognizes that the story "makes sense" to her. Conversely, coherent, but meaningless stories are more resource-intensive and not comforting to the patient.

Narrative fallacy

Therapists often "think in stories" when creating case formulations. Taleb (2007) described a related phenomenon to this observation, which he called "narrative fallacy." Here is a description in his own words: "The narrative fallacy addresses our limited ability to look at sequences of facts without weaving an explanation into them, or, equivalently, forcing a logical link, an arrow of relationship, upon them" (p. 118).

If we translate Taleb's statement into the language of active inference, then it will correspond to what Parr and colleagues summarized as follows: "Put simply, we should be uncertain (adopt a high entropy belief) when we have no information" (p. 28). Specifically, if we have no information about the relationship between A and B in the world, we should remain uncertain about their relatedness until we receive information that would allow us to update a prior belief to a hopefully more accurate posterior belief.

One might justifiably comment here that seeking an explanation of the unexplained in psychodynamic therapy is by design – the patient at least sometimes needs an explanation of his distress, among other things. Indeed, coming up with an explanation of the patient's presenting problem that is rooted in a data-driven understanding of the etiology of the patient's condition could be useful. For example, when a patient suffers from disturbing intrusive memories, which are concrete, timeless, and intense, a therapist, having collected enough data to suspect acute PTSD, can tentatively conceptualize the patient's symptom as a "flashback," which, if needed, allows her to explain how the flashbacks start, how they operate, and so forth.

An important nuance here is McWilliams's point (2011, p.10) that it is often futile to diagnose based on the manifest problem alone, without considering the context of the patient's personality structure.

However, the issue with narrative fallacy influencing our work is that at times psychodynamic therapists are prone to what can be metaphorically called a form of "hypersalience" – where we draw connections between the pieces of data without any reason other than our theory-driven beliefs.

This is what Taleb wrote about – drawing a connection from A to B should be based on substantive data. For example, it is justified to hypothesize post-traumatic symptoms based on

the patient's report of experiencing an acutely traumatizing event a year before the intake and feeling profoundly helpless during that event, the therapist observing what seems to be dissociations in session, and transference/countertransference dynamics supporting the therapist's impressions of trauma.

As an illustration of the point made in the previous two paragraphs, consider the following summary by Busch et al. (1999) about the psychodynamic components of panic disorder. This summary is based on the investigations of video-recorded psychoanalytic work with 15 patients:

The ego weaknesses, typical defenses, and preoedipal dynamics described above intensify the subsequent danger of oedipal longings in the panic patient. Thus, aggressive strivings of the positive oedipus complex and dependent and homosexual fantasies of the negative oedipus complex seem more dangerous, and castration anxiety, seen in an extreme sense of body vulnerability and death fears ("I'm having a heart attack"), further fuels the separation anxiety activated earlier. (p. 777)

It seems that the text above is loosely connected to the patients' panic and strongly connected to the authors' theoretical framework (Ego Psychology, see Mitchell & Black, 1996, pp-53-58 for review.) While this text is unlikely to be shared with any patient verbatim, a potential issue here is that we think this way about the patients and their challenges. Sometimes, we create fiction.

In this example from Busch et al. (1999), a rather large theoretical "castle" of an explanation was built. It seems inaccurate to call this castle a "working hypothesis," as it is not falsifiable. A more appropriate term here is "belief." Importantly, however, this formulation is coherent and meaningful (to an Ego psychologist).

Here lies a phenomenon from which some of the power of the case formulation stories is born. A patient might temporarily feel relief after hearing a story about their symptoms, even when the story is based on beliefs. For example, when faced with debilitating mental pain after the loss of a child, a parent may find some comfort in a spiritual explanation. A tragedy may be significantly more painful when it is perceived as utterly meaningless. Meaning is typically established by weaving the tragedy into the context of a coherent, logical story that the patient understands.

Taleb (2007) wrote the following about this phenomenon “In a famous argument, the logician W. V. Quine showed that there exist families of logically consistent interpretations and theories that can match a given series of facts. Such insight should warn us that mere absence of nonsense may not be sufficient to make something true” (p. 129-130).

Another feature of the narrative fallacy is that therapists do not usually think probabilistically when formulating cases. A therapist’s extended consciousness (Solms and Turnbull, 2018) resembles a storyteller more so than a statistician. (Kahneman, 2011; Taleb, 2007). A related issue is that causal narratives skew probability assessment – even for statisticians. You can find an example of that in an experiment by Kahneman and Tversky cited by Taleb (2007, pp. 134-135).

An additional component of Taleb’s (2007) claim has to do with the preference for explanations. He writes that we are prone to choose causal links when attempting to establish a relationship between objects. While we do generally prefer some explanation as a binding medium between disjointed pieces of data to none, we can add that different kinds of explanations (teleological, causal, etc.) can be preferred by different people (Liquin &

Lombrozo, 2018). The critical point is not the exact nature of an explanation, but the act of binding.

However, it may be useful to spend some time on the different kinds of explanations, as they differ in how they affect clinical practice. While the concept of causality is complex and has various definitions (see Bunge, 2017 for review), it seems that most definitions converge on the following technical part: when we say “A caused B” we are talking about an influence of A (the cause) on B (the effect).

What do we mean by “influence” in clinical practice? When we deal with a condition that has an acute onset, such as acute PTSD, we seem to be justified in stating that one of the key factors that influenced the formation of symptoms was the patient’s exposure to a traumatic event. However, for most of the conditions we work with, we do not have a clear picture of their etiology. In the absence of such a clear picture, we fill in the explanatory gaps with inference – we connect the patient’s current symptoms to some other factors.

With teleological explanations, we may attribute a purpose or a function to a clinical phenomenon. For example, we can infer that one of the functions of affect is to inform our mind of a specific issue requiring our attention in circumstances when we are in an unpredictable environment (Solms, 2019). This approach can indeed be useful since seeing depressive affect as an information source rather than something to avoid could be a fruitful therapeutic stance. However, depression may have many functions, and it is unclear how to assess the accuracy of our inference about one specific function of depression. How would we test such hypotheses?

Verifiability

A classification of the models that therapists create when building case formulations is the subject of debate. Some clinicians call it a “working hypothesis,” which implies using a

hypothetico-deductive method. A “working hypothesis” in this framework is tested against the data coming up in the case. When the incoming data support the formulation, it is maintained. Otherwise, a new working hypothesis must be formulated and tested again.

However, Freud (1909) stated the following: “For a psychoanalysis is not an impartial scientific investigation, but a therapeutic measure. Its essence is not to prove anything, but merely to alter something” (p. 104). McWilliams (2011) wrote: “Many concepts central to analytic thinking have not only not been systematically researched and validated; they are inherently so resistant to being operationalized and manipulated that it is difficult to imagine how they even could be empirically tested (see Fisher & Greenberg, 1985)” (p.4).

Solms (2018) wrote a paper “The Scientific Standing of Psychoanalysis,” in which he outlined the core theoretical claims of psychoanalysis from the scientific standpoint and presented the research data on the efficacy of psychodynamic psychotherapy for some conditions. Further, Solms’ current neuropsychanalytic approach to the clinical case formulation is based on a hypothetico-deductive method and he uses multiple perspectives to test his hypotheses (2023, personal communications).

When we pose a question of formal falsifiability of psychodynamic case formulations, we need to look at the level of specific statements. Popper (1949) suggested that a statement is falsifiable when it can be contradicted by an empirical test. This implies (a) the existence of a clear falsifier, which can be observed intersubjectively with the current technological methods (b) sufficient specificity of the statement (c) the proof of the statement being false must be positive; and the absence of data cannot falsify the statement; (d) confirmatory observations do not validate the theory.

From this standpoint, all the statements in the case of panic above (Busch et al., 1999) are not falsifiable. Moreover, they are inductive – based on generalizing specific clinical observations (of the patients in the study) onto universal statements about panic. In that, they resemble Popper’s famous example of an observation-based conclusion that “all swans are white” (p.26). To clarify, the statements by Busch and colleagues (1999) do not seem to be inferences in the Bayesian sense – there are no prior or posterior probabilities discussed in the paper. The modern applications of Popper’s falsifiability criterion do not prohibit inductive logic when it uses the application of Bayes’s theorem; what is essential is that a clear falsification criterion is defined, including a statistical/probabilistic one (see Mayo, 2022 for a review and discussion).

Therefore, we generally cannot claim formal falsifiability of psychoanalytic hypotheses in case formulations, but we should be, at least in theory, open to evaluating the accuracy of our initial formulations and changing them if the new data in the case contradicts them.

Putting aside the debate on the formal falsifiability of our clinical hypotheses, we have other obstacles on the path of hypothesis testing. Two examples of such obstacles include the Barnum/Forer effect and effort justification bias.

The verification of the case formulation’s fit for the incoming data is based on the therapist’s opinion, which she forms based on (a) the patient’s reaction to the therapist sharing with him the elements of the formulation (b) the patient’s response to interventions (c) the patient’s feedback about how they feel with respect to their presenting issues in the course of therapy. Therefore, the verification is subjective. While psychology is the science of the subjective, when we talk about a therapist’s subjective opinion being the basis for hypothesis testing, we are affected by the natural constraints of the therapist’s mind. To start with, let us

consider the possible influence of the Barnum/Forer effect on the therapist's perception of a good fit between the model and the data.

Forer (1949) gave 39 students a personality inventory. Then, a week later, he presented each of them with "results," – identical personality descriptions, which were based on vague, ambiguous, and general statements, mostly from an astrology book. He then asked the students to rank the accuracy of these statements on a scale from 0 (poor) to 5 (accurate). The mean accuracy was 4.3. Please refer to Dickson et al. (1985) for a literature review on this effect.

The Barnum/Forer effect influences the practice of psychotherapy. When the therapist hears the patient's positive reaction to a tentative case formulation, the therapist hears that something that she said resonated with the patient. As the literature on the Barnum/Forer effect demonstrates, such an experience of resonance is entirely possible when a generic text is used. The patient's experience of "resonance," therefore, cannot be a reliable measure of the hypothesis' fit for his specific issues.

Secondly, when the therapist invests mental energy in creating an initial formulation, she may be naturally biased to retain this formulation due to the effort justification bias or sunk cost fallacy (see Inzlicht et al., 2018 for review). While we do not have formal empirical proof that in clinical practice therapists infrequently change their preliminary hypotheses; the readers are invited to consider the influence of effort justification bias on our work.

When we consider the patient's response to intervention as a measure of hypothesis accuracy, then it is again subjective and usually based on such things as the presence of affect in the patient's response. This method seems to be error-prone. Consider iatrogenic intervention causing the patient to become angry – he is affective, but not better. Consider the patient becoming highly frustrated with the therapist due to insurance changes and the therapist's out-of-

network fee becoming unmanageable. Consider the patient with labile affect. On what basis can we attribute the patient's affect to the therapist's intervention preceding it and not to other factors, or a combination of factors? When the therapist is certain that the patient cried exactly because the therapist just said something, would such an overly confident inference not be a form of confusion between correlation and causation?

Finally, when the patient tells us that they are better, does it mean that our hypothesis is accurate? Conversely, does it mean that our hypothesis was wrong when they tell us they feel worse?

In medieval times the theory of malaria was that it was caused by evil spirits; people were advised to build houses high on the hill and to close the windows at night – this was a piece of effective advice based on a wrong theory. Something “working” does not prove that the underlying theory is accurate (Bernard & Goodyear, 2004). With respect to the patient reporting feeling worse, practicing therapists understand that in the course of treatment, the patient may feel temporarily worse before reaching a stable improvement, particularly in trauma work.

If we combine the lack of specificity and formal falsifiability in psychodynamic working hypotheses with a subjectively based hypothesis verification process prone to the Forer/Barnum effect and effort justification bias, then we can conclude that we do not have a rigorous independently verifiable framework of hypothesis testing in psychodynamic psychotherapy.

Neuropsychological perspective

While working memory is a limited resource (Miller, 1956), it is necessary in conscious speech processing, as part of the executive system (Thorn and Page, 2008). When working memory is strained, we operate in an effortful regime and tend to experience some discomfort

(Kahneman, 2011). This contributes to us naturally gravitating to regimes of functioning that are less taxing for the working memory.

When speech is perceived by the listener, complex auditory, phonological, semantic, and syntactic processing occurs, some unconsciously and some consciously. If we focus on the working memory contribution to some of these phenomena, then, two neuropsychological processes seem pertinent for the narrative fallacy discussion – within sentence binding and across sentence binding (Thorn and Page, 2008). Thorn and Page defined within sentence binding as “the temporary retention of single sentences in a relatively automatic manner, without requiring additional working memory resources” (p.63); and across sentence binding as the integration of larger chunks of prose (p.68).

The research in these areas summarized in Thorn and Page’s (2008) book suggests that coherent, meaningful narratives require less effort to process as compared to incoherent or meaningless ones due to the strain on the working memory of the latter.

As studies reviewed by Thorn and Page suggest, one of the specific, likely metabolically preferred, features of a coherent and meaningful narrative is that it lends itself to chunking (Miller, 1956). Another feature is that when processing such narratives, the listener can use stored knowledge in the LTM to help working memory processing, which is a less resourceful regime of processing.

Thorn and Page summarized the following for within sentence binding:

In line with the work by Caplan and Waters (e.g., Caplan & Waters, 1999, Waters & Caplan, 2004), we would argue that the syntactically and semantically guided chunking and the development of sequential redundancy involved in within-sentence binding arises automatically, at no extra cost to attentional resources. This may emerge through the

interaction between phonological STM and stored language knowledge in LTM, with subsequent storage of the chunked representations in the episodic buffer. (p. 81)

In contrast to the coherent and meaningful sentences, Thorn and Page commented that for unrelated sentence sequences within sentence binding may occur, but it would require a resource-demanding integrative process to bind them together in the episodic buffer. In reference to across sentence binding, Thorn and Page stated the following:

Relatively automatic binding, through the application of LTM knowledge, is also applied to the chunking of multiple sentences, provided they unite to form a coherent and meaningful story. Additional central executive resources only become important when binding is required between multiple unrelated propositions and/ or sentences (e.g., Jefferies et al., 2005). (p. 80)

To summarize, two neuropsychological processes likely contribute to human preference for meaningful and coherent stories over disjointed pieces of data – chunking and the use of LTM to augment working memory functioning. These processes allow us to avoid a regime of functioning when working memory is overloaded, which is effortful and metabolically costly. It seems that these neuropsychological phenomena are related to the preference for narrative styles of thinking described in the section *Narrative Fallacy* above.

Active inference perspective on causal and teleological explanations

From the standpoint of active inference, beliefs, inference, and causality are entirely appropriate modeling tools. One of the possible active-inference-based understandings of the perception-action cycle is that an organism infers about a cause that corresponds to some inner state/sensation (Maisto et al., 2021; Parr et al., 2022).

The difference between the active inference beliefs about the causes and the psychodynamic ones described in the case of panic above (Busch et al., 1999) seems to be that the former can be evaluated quantitatively and can be updated based on the data collected in the environment while the latter cannot be quantitatively evaluated or formally tested.

Regarding teleological explanations, active inference can account for an agent's teleological (purposeful) behavior (Parr et al., 2022, p. 53), but not for the functions or purpose of symptoms/psychopathology. Broadly, psychopathology under active inference is seen as a “failure in belief updating,” (Parr et al., 2022, p. 185), whereby a specific failure in updating specific beliefs results in certain symptoms.

Active inference perspective on emotional dynamics

One of the ways to describe emotional dynamics in the brain-mind is as a set of complex flows. Various methods are used in physics and mathematics to study such objects, including vector fields and vector calculus. Friston and colleagues (2021) use Helmholtz decomposition, which states that a vector field satisfying specific conditions can be decomposed into a curl-free vector field and a solenoidal vector field. More specifically, due to assuming non-equilibrium steady states (see Friston, 2019, p.10), they can use Helmholtz decomposition to express the dynamics of states (flow of states).

Within this framework, what happens when the patient puts his emotional dynamics into words? As stated earlier in this paper and in Tolchinsky (2023), the brain-mind contains flows, and at multiple levels of the hierarchy at the same time. Let us consider an acute fear reaction as an illustration. Following Panksepp's terminology (1998), a capitalized notation (e.g. FEAR) will be used hereafter to designate the primary emotional systems in Panksepp's taxonomy

(Panksepp & Biven, 2012), while a lowercase use (e.g. fear) will designate the vernacular meaning of emotions.

While the fear reaction may start from a categorical “decision” in the FEAR system (Panksepp & Biven, 2012), an acute fear cascade in the brain is a set of rapid flows. The patient is not directly aware of the brain’s inner processes but may become aware of the mental part of the emotional cascade – the affect (Solms, 2020, p. 13). This awareness happens at the level of core/elemental consciousness (Solms & Turnbull 2018, p. 247; Solms, 2019, p.1). Subsequently, the patient may label this affective state verbally as “fear,” which is a process at the higher level of consciousness, sometimes referred to as extended/reflexive consciousness (Solms & Turnbull 2018, p. 247). Here is a perspective on this process in a quote from the book on active inference (Parr et al., 2022):

There is an emerging bedfellow for *interoceptive inference*—namely, *emotional inference*. In this application of Active Inference, emotions are considered part of the generative model: they are just another construct or hypothesis that the brain employs to deploy precision in deep generative models. From the perspective of belief updating, this means anxiety is just a commitment to the Bayesian belief “I am anxious” that best explains the prevailing sensory and interoceptive queues. (p. 216)

Parr and colleagues further suggest that emotional inference leads to “domain general” (p. 217) belief updating, which is the integration of information from interoceptive and exteroceptive sensory streams, and, likely, a process at a higher level in a predictive hierarchy as compared to domain-specific beliefs.

Let us now look at the therapist attempting to name the patient’s emotional states during the work in a dyad. As a possible outlook informed by the active inference framework, the

therapist (a) starts by having generative models of various states of the universal, abstracted brain-mind and (b) adjusts these models based on the data collected from the work with the patient in therapy (including data shared by the patient, the therapist's observations, transference and countertransference impressions, etc.) (c) infers about what the patient is feeling based on these models and then (c) acts – verbally – by labeling a set of dynamic circumstances as, say, “fear.” This verbal label is then essentially a static value of a discrete variable representing a certain affective state of the patient.

In this process, the therapist is engaged in modeling the patient's mind together with the patient. The therapist's static label for the patient's affective state can be described as the therapist's internal Bayesian belief about the non-transparent, external phenomenon he is trying to model. While the therapist's extended consciousness is likely not a statistician, the process described above of “model-infer-act” is probabilistic, and we are not necessarily aware of that at the moment of naming. For example, the therapist could be wrong and mislabel a state of hyper-arousal from coffee as fear.

The problem in clinical practice is that when we name emotions (silently or out loud), we virtually never say that this is fear with a 69% probability. At best, we say that this could possibly be fear.

As part of the process in psychotherapy, we train our models to adjust the labels and increase precision.

We could describe another cascade, a sequence constructed from static labels, such as a shift from a child's neutral state of contentment to activation of her PANIC system (Panksepp & Biven, 2012) as a reaction to her caregiver leaving the room for an extended period of time, which is followed by the child moving on to a state of protest – activation of RAGE system

(Panksepp & Biven, 2012). Subsequently, when the protest does not result in the caregiver's return, the child's RAGE system becomes deactivated, and she moves on to a state of despair. Subsequently, with the return of a caregiver, the child may return to the state of contentment, and her PANIC system returns to the euthymic state.

This is a flow of states, which we have labeled in words and sequenced. This flow can possibly be described as solenoidal/cyclical. An example of a gradient flow could be an adult patient with severe depression who ended up committing suicide during the activation of PANIC and then RAGE systems.

The processes described in the previous two paragraphs are not examples of narrative fallacy, provided we stay data-driven, and test our inferences in the environment. Additionally, we have some empirical support for the process described above as a solenoidal flow (Panksepp & Biven, 2012).

Narrative fallacy happens when we apply an externally imposed narrative, such as a monoamine theory of depression, or the discussion of a state of panic from the standpoint of oedipal dynamics. We then sequence static, abstract labels from the theory and create relational links between the patient's mental objects while being detached from the patient's mental environment. Instead, we draw connections based on the theory alone.

What can be done?

The formulations we create stem from the underlying psychodynamic theories. Therefore, a starting point in updating the process of case formulation could be updating the theoretical models. Should we have a revision of our theoretical models, it would likely lead to an update in their clinical applications as well. The current paper only outlines preliminary thoughts about the

aspects of the case formulation process that can be considered in anticipation of possible theoretical revisions.

First, we can consider a gradual shift from narrative-based formulations to a multi-dimensional structure that resembles McWilliams' recent method (2021, pp. 44-67). Each dimension can describe a separate variable relevant to the patient's distress. Dimensions can be measured as continuous variables, such as the level of the patient's generalized arousal, or as discrete ones, such as a set of neuroaffective systems (Panksepp & Biven, 2012) that the therapist thinks are dysregulated for the specific patient. In line with the active inference framework, we can then consider the probabilistic representations of the variables. The model's parameters can be adjusted during the work in therapy, as new data is accumulated in the case. While this model may not be something that a therapist can hold in working memory, it can certainly be used with the therapist's conscious mind augmented by writing things down, or by software containing the representation of the model.

This is the key point of this paper – we should not be constrained in designing clinical models by their need to fit the limitations of our working memory.

Second, in addition to the audio-verbal exchange between the patient and the therapist, we can consider the possibility of using additional tools in the clinical work that would allow us to collect data on the flows that do not lend themselves easily to verbal descriptions, such as EEG dynamics (which is already done in neurofeedback treatments). In addition, when we study the flows of states, the use of tools would allow us to capture the dynamics that are faster than we can detect with the naked eye. The tools may include video cameras or AI-based software to document and analyze the patients' flows nearly in real-time.

Third, we can consider formalizing the process of verification of our initial working hypothesis such that it does not consist merely of the therapist's opinion and is less prone to the Barnum/Forer effect and effort justification bias.

As an example of the second point, Beatrice Beebe uses video cameras to study the flows of sounds, emotional expressions, and gestures in the infant-mother dyads (see American Enterprise Institute, 2018 for an example). She then watches the videos in slow motion with the mother. This process allows her to investigate the dynamics that are, perhaps, preconscious, and too fast for the extended consciousness level of the mind. As one of the components of her work, together with the mother she can assess the level of attunement in the dyad, to see if the mother is overreacting to the baby, under-reacting, or being on the same "wavelength." This insight is an example of "slowing down" – processing the fast flows of states at the higher, slower level of the predictive hierarchy, but with the use of data about the faster flows that were carefully collected with the help of video cameras.

The changes proposed in this paper may be considered by some psychodynamic theorists unwelcomed due to a perceived threat of "distorting the transference dynamics;" however, not only Beatrice Beebe but also ISTDP clinicians routinely use video cameras in their psychodynamic work. While neurologists certainly use narratives in their formulations, they also collect data with EEG, MRI, and, increasingly, chaos-theory-informed software tools (see Tolchinsky, 2023 for review). The use of tools in psychodynamic psychotherapy warrants a discussion and an evaluation of possible gains and losses.

References

- American Enterprise Institute (2018, October 19) Decoding the nonverbal language of babies | LIVE STREAM. [Video]. YouTube.
<https://www.youtube.com/watch?v=0yVU2lmlMdk&t=15m30s>
(No DOI available)
- Bernard, J., & Goodyear, R.(2004). *Fundamentals of clinical supervision*. Allyn & Bacon. (No DOI available.)
- Bunge, M. (2017). *Causality and modern science*. Routledge.
<http://dx.doi.org/10.4324/9781315081656>
- Busch, F., Milrod, B., Rudden, M., Shapiro, T., Singer, M., Aronson, A., & Roiphe, J. (1999). Oedipal dynamics in panic disorder. *Journal of the American Psychoanalytic Association*, 47(3), 773-790. <http://dx.doi.org/10.1177/00030651990470030401>
- Dickson, D., & Kelly, I. (1985). The 'Barnum Effect' in personality assessment: A review of the literature. *Psychological Reports*, 57(2), 367-382.
<http://dx.doi.org/10.2466/pr0.1985.57.2.367>
- Forer, B. (1949). The fallacy of personal validation: a classroom demonstration of gullibility. *The Journal of Abnormal and Social Psychology*, 44(1), 118.
<http://dx.doi.org/10.1037/h0059240>
- Freud, S. (1901). *The psychopathology of everyday life*. *Standard Edition*, (Vol. 6). London: Hogarth. (No DOI available.)
- Freud, S. (1909). *Analysis of a Phobia in a Five-Year-Old Boy*. *The Standard Edition*, (Vol. 10). 1-150. (No DOI available.)
- Friston, K. (2013). Life as we know it. *Journal of the Royal Society Interface*, 10(86),

20130475. <http://dx.doi.org/10.1098/rsif.2013.0475>

Friston, K. (2019). A free energy principle for a particular physics. *arXiv preprint* arXiv:1906.10184. (No DOI available.)

Friston, K., Da Costa, L., & Parr, T. (2021). Some interesting observations on the free energy principle. *Entropy*, 23(8), 1076. <http://dx.doi.org/10.3390/e23081076>

Gabbard, G. (2005). *Psychodynamic psychiatry in clinical practice*. (4th ed.). American Psychiatric Pub. (No DOI available.)

Inzlicht, M., Shenhav, A., & Olivola, C. Y. (2018). The effort paradox: Effort is both costly and valued. *Trends in cognitive sciences*, 22(4), 337-349. <http://dx.doi.org/10.1016/j.tics.2018.01.007>

Joseph, B. (1985). Transference: The total situation. *International Journal of Psychoanalysis*, 66(4), 447-454. (No DOI available.)

Leichsenring, F., Leweke, F., Klein, S., & Steinert, C. (2015). The empirical status of psychodynamic psychotherapy-an update: Bambi's alive and kicking. *Psychotherapy and psychosomatics*, 84(3), 129-148. <http://dx.doi.org/10.1159/000376584>

Liquin, E., & Lombrozo, T. (2018). Structure-function fit underlies the evaluation of teleological explanations. *Cognitive Psychology*, 107, 22-43. <http://dx.doi.org/10.1016/j.cogpsych.2018.09.001>

Kahneman, D. (2011). *Thinking, fast and slow*. Macmillan. (No DOI available.)

Maisto, D., Barca, L., Van den Bergh, O., & Pezzulo, G. (2021). Perception and misperception of bodily symptoms from an active inference perspective: Modelling the case of panic disorder. *Psychological Review*, 128(4), 690. <http://dx.doi.org/10.1037/rev0000290>

Mayo, D. (2022). The statistics wars and intellectual conflicts of interest. *Conservation Biology*, 36(1). <http://dx.doi.org/10.1111/cobi.13861>

McWilliams, N. (1999). *Psychoanalytic case formulation*. Guilford Press. (No DOI available.)

McWilliams, N. (2004). *Psychoanalytic psychotherapy: A practitioner's guide*. Guilford Press. (No DOI available.)

McWilliams, N. (2011). *Psychoanalytic diagnosis: Understanding personality structure in the clinical process*. Guilford Press. (No DOI available.)

McWilliams, N. (2021). *Psychoanalytic supervision*. Guilford Publications. (No DOI available.)

Miller, G. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological review*, 63(2), 81. <http://dx.doi.org/10.1037/h0043158>

Mitchell, S., & Black, M. (1996). *Freud and beyond. A history of modern psychoanalytic thought*. Basic Books. (No DOI available.)

Panksepp, J. (1998). *Affective neuroscience: The foundations of human and animal emotions*. Oxford University Press. (No DOI available.)

Panksepp, J., & Biven, L. (2012). *The archaeology of mind: Neural origins of human emotion*. WW Norton & Company. (No DOI available.)

Parr, T., Pezzulo, G., & Friston, K. J. (2022). *Active inference: The free energy principle in mind, brain, and behavior*. MIT Press. <http://dx.doi.org/10.7551/mitpress/12441.001.0001>

Popper, K. (2005). *The logic of scientific discovery*. Routledge. (No DOI available)

Schafer, R. (1992). *Retelling a life: Narration and dialogue in psychoanalysis*. Basic Books. (No DOI available)

- Shedler, J. (2010). The efficacy of psychodynamic psychotherapy. *American psychologist*, 65(2), 98. <http://dx.doi.org/10.1037/a0018378>
- Solms, M. (2018a). The scientific standing of psychoanalysis. *BJPsych International*, 15(1), 5-8. <http://dx.doi.org/10.1192/bji.2017.4>
- Solms, M. (2018b). *The feeling brain: Selected papers on neuropsychoanalysis*. Routledge. <http://dx.doi.org/10.4324/9780429481758>
- Solms, M. (2019). The hard problem of consciousness and the free energy principle. *Frontiers in Psychology*, 2714. <http://dx.doi.org/10.3389/fpsyg.2018.02714>
- Solms, M., & Turnbull, O. (2018). *The brain and the inner world: An introduction to the neuroscience of subjective experience*. Routledge. <http://dx.doi.org/10.4324/9780429481239>
- Stanford Complexity Group (2013, March 30) *The Stanford Complexity Group Presents: Walter J. Freeman.wmv*. [Video]. YouTube. <https://www.youtube.com/watch?v=mrSzPOUdJIA>
(No DOI available)
- Sulis, W. (2021). Contextuality in neurobehavioural and collective intelligence systems. *Quantum Reports*, 3(4), 592-614. <http://dx.doi.org/10.3390/quantum3040038>
- Taleb, N. (2007). *The black swan: The impact of the highly improbable*. Random House. (No DOI available.)
- Thorn, A., & Page, M. (2008). *Interactions between short-term and long-term memory in the verbal domain*. Psychology Press. <http://dx.doi.org/10.4324/9780203938966>
- Tolchinsky, A. (2023). A case for chaos theory inclusion in neuropsychoanalytic modeling. *Neuropsychoanalysis*, 25(1) <https://doi.org/10.1080/15294145.2023.2191983>
- Wallace, E. (1988). What is “truth?” Some philosophical contributions to psychiatric issues. *The American Journal of Psychiatry*, 145(2), 137-147. (No DOI available.)