

Article

A Dynamic Autocatalytic Network Model of Therapeutic Change

Kirthana Ganesh ^{*,†}  and Liane Gabora ^{*,†} 

Fipke Centre for Innovative Research, Department of Psychology, University of British Columbia, 3247 University Way, Kelowna, BC V1V 1V7, Canada

* Correspondence: kirthana.ganesh@ubc.ca (K.G.); liane.gabora@ubc.ca (L.G.)

† These authors contributed equally to this work.

Abstract: Psychotherapy involves the modification of a client's worldview to reduce distress and enhance well-being. We take a human dynamical systems approach to modeling this process, using Reflexively Autocatalytic foodset-derived (RAF) networks. RAFs have been used to model the self-organization of adaptive networks associated with the origin and early evolution of both biological life, as well as the evolution and development of the kind of cognitive structure necessary for cultural evolution. The RAF approach is applicable in these seemingly disparate cases because it provides a theoretical framework for formally describing under what conditions systems composed of elements that interact and 'catalyze' the formation of new elements collectively become integrated wholes. In our application, the elements are mental representations, and the whole is a conceptual network. The initial components—referred to as *foodset items*—are mental representations that are innate, or were acquired through social learning or individual learning (of *pre-existing* information). The new elements—referred to as *foodset-derived items*—are mental representations that result from creative thought (resulting in *new* information). In clinical psychology, a client's distress may be due to, or exacerbated by, one or more beliefs that diminish self-esteem. Such beliefs may be formed and sustained through distorted thinking, and the tendency to interpret ambiguous events as confirmation of these beliefs. We view psychotherapy as a creative collaborative process between therapist and client, in which the output is not an artwork or invention but a more well-adapted worldview and approach to life on the part of the client. In this paper, we model a hypothetical albeit representative example of the formation and dissolution of such beliefs over the course of a therapist–client interaction using RAF networks. We show how the therapist is able to elicit this worldview from the client and create a conceptualization of the client's concerns. We then formally demonstrate four distinct ways in which the therapist is able to facilitate change in the client's worldview: (1) challenging the client's negative interpretations of events, (2) providing direct evidence that runs contrary to and counteracts the client's distressing beliefs, (3) using self-disclosure to provide examples of strategies one can use to diffuse a negative conclusion, and (4) reinforcing the client's attempts to assimilate such strategies into their own ways of thinking. We then discuss the implications of such an approach to expanding our knowledge of the development of mental health concerns and the trajectory of the therapeutic change.

Keywords: autocatalytic network; creativity; conceptual network; psychotherapy; therapeutic change; uncertainty; worldview



Citation: Ganesh, K.; Gabora, L. A Dynamical Autocatalytic Network Model of Therapeutic Change. *Entropy* **2022**, *24*, 547. <https://doi.org/10.3390/e24040547>

Academic Editors: Franco Orsucci and Wolfgang Tschacher

Received: 1 February 2022

Accepted: 22 March 2022

Published: 13 April 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

While the efficacy of psychotherapy as a form of treatment has been clearly established [1], there is uncertainty about *why* it works [2]. Statistical approaches model the psychotherapeutic process using moderator and mediator variables [3,4], but this does not go far toward explaining each mind's unique, self-organizing network of associations, how this structure took shape, and how it responds to psychotherapy. We have only a hazy understanding of how specific elements of the psychotherapy process contribute

to therapeutic change [5]. This paper aims to take a step forward toward a more precise understanding of what happens in psychotherapy and what makes it effective using a dynamical systems framework to model the interconnected, self-organizing nature of an individual's worldview, and its dynamical change over time. In so doing, we aim to both strengthen the theoretical bases of psychotherapy, and sharpen our capacity to improve it in practice.

We view psychotherapy as a creative collaborative process between therapist and client, in which the output is not an artwork or invention, but an outlook and approach to life on the part of the client. A client's outlook and behavior flows from the web of knowledge and experience that collectively constitute a way of seeing the world and being in the world, i.e., a *worldview*. What makes network approaches to cognition particularly promising is that this web of knowledge and experience can be described as a network consisting of loosely connected clusters (i.e., the network has an intermediate degree of modularity) that can be characterized using tools from network science [6–12]. Thus, the network approach offers a novel way of understanding mental health concerns and their treatment [13]. Psychotherapy (whether it be behavioral, cognitive, etc.) attempts to 'destabilize' a distressing or pathological mental state and shift the individual toward a healthier mental state [13–15]. What distinguishes the 'autocatalytic' approach to cognition taken here from other network-based models of cognition is its capacity to generate new elements (such as mental representations or schemas) out of interactions between existing ones. This makes it useful in the context of describing psychotherapeutic change, though the approach has been applied more broadly to other forms of cognitive change as well [16–18]. Since therapy entails change in the structure and dynamics of this network, network science seems a natural place to start in modeling the therapeutic process.

We note that it is not simply the case that positive interpretations (i.e., narratives that make the individual feel good) are adaptive while negative interpretations (i.e., narratives that make the individual feel bad) are maladaptive. It is often necessary that negative situations be acknowledged as such to spur action and find solutions; however, one worldview may predispose the individual to overcoming challenges and finding opportunities, while another leads to unnecessary distress and feelings of helplessness. Thus, the therapist strives to help the client to 'unravel' their worldview just enough to 'reweave' it into one that is, for that client, adaptive.

One's society and culture provides stories, narratives, scripts, and schemas, as well as larger conceptual frameworks (such as science or religion) that offer prescriptions for integrating them into a worldview; however, since no one else is privy to an individual's entire repertoire of personal experiences and intimate observations, the worldview one weaves is ultimately unique. Much as, for any given set of dots there are multiple ways of connecting them, for any given set of experiences or mental representations, there may be many ways of integrating them into a worldview. However, this sentence is not to be interpreted as implying that mental representations are, indeed, anything like 'dots'; they clearly have a context-dependent inner structure. A worldview may selectively include, or exclude, certain experiences and positively (or negatively) valenced items, (or weight them more strongly). Some worldviews may be more adaptive than others, i.e., more conducive to thriving, personal growth, and the well-being of the individual and their social and environmental milieu.

We model the psychotherapeutic process using a certain kind of network referred to as a Reflexively Autocatalytic foodset-derived (RAF) network. Though the term 'autocatalytic networks' reflects their initial application to the origin of life [19,20], RAFs provide a general mathematical setting for studying networks that arose out of earlier work in graph theory [21]. The term *reflexively* is used in its mathematical sense, meaning that each part is related to the whole. The term *autocatalytic* will be defined more precisely shortly, but for now it refers to the fact that the whole can be reconstituted through interactions amongst its parts. The term *foodset* refers to the elements that are initially present, as opposed to those that are the products of interactions amongst them. As in other network science approaches,

the nodes of the network represent units of information such as words, concepts, memories, or mental representations of concrete or abstract knowledge, and connections between nodes (by way of free association, shared features, or co-occurrences) are represented as edges (For example, ‘chair’ and ‘wood’ are nodes, and the relationship between them, i.e., wood can be used to create a chair, is represented as an edge).¹

What differentiates RAFs from other approaches in network science is that the nodes are not just passive transmitters of activation; they actively galvanize, or ‘catalyze’ the synthesis of novel (‘foodset-derived’) nodes from existing ones (the ‘foodset’). The generalized RAF setting is conducive to the development of efficient (i.e., polynomial time), algorithms for questions that are computationally intractable (i.e., NP-hard. [30]). These features make RAFs uniquely suited to model how new structure grows out of earlier structure, i.e., generative network growth [30]. Such generativity may result in phase transitions to a network that is self-sustaining and self-organizing [31–33], as well as potentially able to evolve, i.e., exhibit cumulative, adaptive, open-ended change [34,35]. For this reason, RAFs have been used to model the origins of evolutionary processes, both biological—the origin of life (OOL) [36,37]—and cultural—the origin of culture (OOC), or more specifically, the kind of cognitive structure capable of generating cumulative, adaptive, open-ended innovation [17,38–41]. In a OOL context, RAFs were used to develop the hypothesis that life began as, not as a single self-replicating molecule, but as a set of molecules that, through catalyzed reactions, collectively reconstituted the whole [20]. Autocatalytic network theory has successfully demonstrated—mathematically or using simulations [36,42], and with real biochemical systems [37]—how self-maintaining structures that evolve and replicate can emerge from nonliving molecules. Because RAF nodes modify network structure, the RAF framework is consistent with the goal of understanding not just how networks are structured but also how they dynamically restructure themselves in response to internal and external pressures.

When autocatalytic models are applied in a cognitive context as they are here, they model not just network structure, but how the network reconfigures itself on the fly in response to changing needs and experiences. The observation that, similar to living organisms, cognitive networks are self-sustaining, self-organizing, and self-reproducing [43–46] suggests that cognitive networks constitute a second level of autocatalytic structure. By *cognitive network*, we refer to an individual’s web of concepts, language terms, and their associations, as well as knowledge and memories, and how they are structured. The self-sustaining nature of a cognitive network is evident in the tendency to reduce cognitive dissonance, resolve inconsistencies, and preserve existing schemas in the face of new information. Although the contents of a cognitive network change over time, it maintains integrity as a relatively coherent whole. Its spontaneously self-organizing nature is evident in the capacity to combine remote associates [47] (such as combining CHOCOLATE and BUNNY to invent CHOCOLATE BUNNY).² The cognitive autocatalytic network replicates in a piecemeal manner through social learning and story-telling. Psychotherapeutic change facilitates the piecemeal replication of adaptive perspectives and habits, as well as the reorganization of relationships between elements of the client’s worldview, and the RAF approach is well-suited to model this.

We begin with an introduction to the psychotherapeutic process. We then introduce RAF networks, and elaborate how they are used in this paper. Next, we present A RAF network model of therapeutic change facilitated by the therapist. The paper closes with implications of the model for fostering a concrete understanding of psychotherapeutic techniques, and suggestions for extending and testing it. A list of abbreviations, and a glossary of terms and their definitions can be found in the Appendix A.

¹ As explained elsewhere [16], the ‘autocatalytic’ approach taken here is consistent with distributed models of mental representations in memory [22], and with quantum models of their interactions [23–29].

² This proposed cognitive level of autocatalytic structure is not merely an extension of organismal needs; indeed, the biological and cognitive/cultural levels of endogenous control can be at odds (e.g., a scientist immersed in solving a problem may neglect offspring, or forget to eat).

2. The Therapeutic Process

Psychotherapy, or ‘talk therapy’ is rooted in formal Western medicine since the late 1800s and practices to alleviate human distress through conversation, known as the ‘moral cure’ has existed formally and informally for centuries [48]. Despite the fact that many effective forms of psychotherapy have been developed, there is uncertainty regarding the mechanisms of therapeutic change [2,5]. Different constructs across therapies show overlap, leading to difficulty with defining their roles and relative importance in the therapeutic process [49]. Thus, a paradigm shift towards dimensional, systemic, and interactional approaches to understanding mental illness and psychotherapy is warranted [50–52]. Multi-modal, multi-perspective research methods that enable us to capture the process of therapy in real-time are the future of psychotherapy research [53]. A complex systems approach can thus help re-conceptualize mental health concerns and treatment to more accurately represent the dynamic interactions involved [13], and help us understand therapeutic change [54].

The collaborative nature of psychotherapy is dyadic, and each member of the dyad influences the other through verbal, nonverbal, and physiological synchrony [55–58]. However, the client and therapist rarely have independently corroborative estimates of important process variables such as the therapeutic alliance [59], and any conceptualization of therapeutic change should make room for both perceptions. RAF networks can accommodate both perspectives within a single framework. The model presented in this paper focuses more on the change in the client’s mind, but the approach has the potential to be expanded to include the therapist, and thereby capture the bidirectional exchange more comprehensively.

Clients generally enter therapy to alleviate distress and increase well-being. Sometimes the decision is prompted by a specific problem, a troubling experience or belief, or something that is difficult to accept. A client may report symptoms of depression, such as sadness, hopelessness, and decreased motivation due to, for example, difficulties with interpersonal relationships. As such, the client’s approach towards such relationships, whether it is the thoughts, emotions, or behaviors involved, are currently insufficient/ineffective in helping them achieve their goals. Therapy may bring about modification of their cognitive network, by enabling them to find a new perspective on a problem or a troubling experience or belief, or come to terms with something they could not accept, thus integrating it into their worldview.

There are significant parallels between the creative process involved in, say, inventing something new, and the process of problem-solving in psychotherapy. Creativity flourishes in situations that involve a tension between uncertainty and constraints [60], or what has been referred to as *enabling constraints* [17,61]. We posit that the forging of a new, healthy, integrated conception of the world and one’s place in it is a creative process, and by cultivating a client-tailored therapeutic interaction, the therapist acts as the midwife of this process.³

3. Rationale for the Approach

Similar to other cognitive network approaches, RAF networks are hierarchical yet decentralized, and they can be analyzed with respect to density, connectedness, and size. They also draw upon the conventional cognitive science notion of *spreading activation* through nodes of a concept graph, and techniques such as shortest path distance and clustering analysis. However, the RAF approach differs from other network approaches used psychology and cognitive science in a number of respects:

³ Many describe the generative aspect of cognition as a ‘birthing’ of new ideas or attitudes, and this word directly captures the therapist’s role in facilitating this process.

3.1. 'Reactions' and 'Catalysis'

A RAF consists of not just nodes connected by edges, but also *reactions*, or interactions that trigger, or 'catalyze' new nodes. Thus, RAF nodes are not merely passive recipients of spreading activation; they actively redirect it. For example, seeing a Superman movie might spark a child who believes she is powerless to draw herself as a superhero. In this example, the 'reactants' are the mental representation of herself as powerless, as well as her drawing skills. The 'catalyst' is the Superman movie, and the 'product' is the drawing of herself as a superhero.

In chemistry, and in applications of RAFs to the origin of life, the term 'reaction' refers to an interaction between molecules. For consistency, in cognitive applications of RAFs, the term *reaction* is used to refer to an interaction between *mental representations* (MRs) in a cognitive network. It may involve *representational redescription* (RR): the re-coding of information in working memory by modifying, restructuring, elaborating, and/or performing mental operations upon it, or possibly in the absence of an external cue [62]. RR can also involve a shift of perspective, and it can result in a flash of insight, or a newly perceived application for an old idea.⁴ The issue of which concepts participate in a given reaction is discussed and mathematically modeled in [40].

RAF also has two kinds of edges: reaction edges and catalysis edges. *Reaction edges* are similar to the edges in conventional network science approaches. (They can be thought of as the 'anatomy' of the network). *Catalysis edges* are more dynamic. (They can be thought of as the 'physiology' of the network). MRs are *catalytic* because they not only participate in certain reactions, but also facilitate—or catalyze—other reactions.⁵ In chemistry, a catalyst speeds up a reaction that would otherwise occur very slowly if at all. By endowing cognitive network models with the capacity for catalysis we can model how one idea or environmental stimulus, triggers a mental operation (such as concept combination, or RR) that would otherwise occur very slowly or not at all. For example, realization of a novel or creative outcome (such as the drawing of a superhero version of oneself) may not have occurred without the galvanizing or 'catalyzing' impact of the experience of watching a Superman movie.

As in chemistry, the cognitive equivalent of a 'catalyzed reaction' may trigger another reaction, and so forth, resulting in a *reaction sequence*. In cognitive models, this reaction sequence is a stream of thought, which may ultimately have arisen from a problem, question, or cognitive dissonance. For example, the ultimate source of the cognitive reaction sequence culminating in the creation of a superhero character may be the desire never to feel powerless.

The rationale for treating mental representations (MRs) as catalysts comes, in part, from the literature on concepts, which provides extensive evidence that when concepts act as contexts for each other, their meanings change [66,67]. For example, an ISLAND has the property 'surrounded by water', but (hopefully) not a KITCHEN ISLAND. KITCHEN momentarily reconfigures the cognitive network, altering the perceived meaning of ISLAND. Such alterations in meaning are often nontrivial, and defy classical logic [68]; however, quantum models of concept interactions provided a means of formalizing the process by which a context (such as the goal of creating a spot to cut food) spontaneously bridges remote associates (such as KITCHEN and ISLAND) [23–25]. Although cognitive RAF models are influenced by how context is modeled in these quantum models of concepts, it is not committed to any formal approach to modeling context. Context is considered to be anything in the external environment, or anything from long-term memory that influences how a MR is instantiated in working memory. The extent to which one MR modifies the meaning of another is referred to as its *reactivity*.

⁴ Creative insights often arise subconsciously from just beyond the bounds of working memory [63].

⁵ The use of the word 'catalyze' in a cognitive context extends beyond autocatalytic models of cognition [64,65], though these other approaches are purely descriptive.

In sum, the RAF approach incorporates not just cognitive change due to adjustments in association strengths, but also cognitive change due to the prompting or ‘catalysis’ of new nodes. The resulting network is dynamic both in terms of structure (e.g., new nodes can be generated), and information flow (e.g., newly generated nodes can result in new information pathways).

3.2. Foodset versus Foodset-Derived

Another key feature of RAF models is the distinction between *foodset* items, which came into existence *outside* the network in question, and *foodset-derived* items, which come about through ‘catalyzed reactions’ *within* the network in question. An individual’s ‘mental foodset’, or simply, *foodset* includes memories of direct experiences, i.e., that came about by way of the senses, including any knowledge that has come about through individual learning (of pre-existing information) by way of direct experience in the world, or through social learning processes such as imitation or classroom learning. The foodset may also include innate responses, such as the fear of heights and corresponding inclination to back away from a precipice. Together, these innate responses, direct experiences, and socially transmitted knowledge constitute the *raw materials* from which the individual’s cognitive network is built. Thus, the worldview is “grounded in perception” because it grows out from this foodset.

Much as bricks and bags of mortar do not constitute a house, the foodset does not constitute a mental model of the world, a worldview. The set of *foodset-derived* items consists of mental contents that were generated by that individual from scratch (and constitute new information) using foodset elements, or perhaps other foodset-derived elements, as ingredients. The generation of foodset-derived items occurs by way of mental operations such as problem solving, insight, deduction, induction, and abduction. Since the elements of the worldview described by foodset-derived items are not grounded in perception, they can be viewed as ‘useful fictions’. Thus, if the therapist responds to powerlessness in a certain way, and the client learns and (later) copies that response, that way of responding is an item in the client’s foodset; however, if the therapist acts as a midwife for the client’s expression of emotions associated with powerlessness during a therapy session, the artwork is a foodset-derived item⁶. The approach thereby distinguishes between conceptual shifts originating within the mind of a given individual, and those that originated by others, and were learnt or assimilated by that individual. What foodset items all have in common is that they are raw materials the individual has at his/her disposal to work with in the generation of new MRs, and this generation of new MRs is a key component of to the conceptual change that occurs during psychotherapy.

In cognitive networks, the distinction between foodset and foodset-derived provides a natural means of grounding abstract concepts in direct experiences; foodset-derived elements emerge through ‘reactions,’ that can be traced back to foodset items. This enables us to identify the necessary precursor ideas for the emergence of new understandings, and the mental operations a given individual carried out to generate a particular idea. This capacity to model the reconfiguration of a cognitive network makes RAFs ideal for the study of change that occurs in psychotherapy.

3.3. Generational Cognitive/Cultural Change

Because of the distinction between foodset and foodset-derived MRs it is possible to tag new insights with their point of origin (i.e., keep track of whose mind did each idea arose in), and track cumulative change step by step within and across individuals. We posit that a mind can be described in terms of nested and overlapping RAFs, and these RAFs are what evolve through culture. Thus, each human lifetime constitutes a small segment of our collective cultural evolutionary lineage (see [34,38,39]). Each generation builds

⁶ This distinction is not as black and white as portrayed here, but for simplicity, we do not address that subtlety for now.

on the accomplishments of the previous generation, such that items that were foodset-derived for one generation become elements of the foodset for the next, and this kind of cumulative cultural evolution has also been modeled, both computationally [69–72], and mathematically using RAFs [17,34,38–40]. For example, an early hominid invented the first tool by realizing that repeatedly striking one stone with another can produce a stone that is sharp, and the mental script of how to make this tool is described as a foodset-derived item in that individual’s mind. This mental script was shared with peers, who in turn transmitted it to others, and in their minds it was a foodset item. As a more psychological example, ‘flattery makes friends’ constituted a foodset-derived item in the mind of the first person to have this thought. He or she may have shared this notion with others, and in their mind it is a foodset item, but one of them may build on it by realizing that imitation can be flattering, and therefore a route to friendship, in which case this new version is a foodset-derived idea (i.e., ‘imitation is the sincerest form of flattery’). Thus, our worldviews consist largely of information that has already been preprocessed into scripts, schemas, stories, and narratives by previous generations, and such ‘chunks’ constrain the shape of one’s worldview.

3.4. Potential to Scale Up

In this initial application of RAF networks to the therapeutic process, the examples used are fairly simple; however, a significant strength of the approach is that RAFs can scale up. The RAF approach can be used to analyze and detect phase transitions in extremely complex networks (such as the phase transition from no-RAF to RAF in Kauffman’s [20] binary polymer model) that have proven intractable using other analytic approaches [37,73].

4. Reflexively Autocatalytic Foodset-Derived Networks (RAFTs)

Let us now define the term *Reflexively Autocatalytic and foodset-derived network* (RAFT) more precisely [30–32,35,74]. The term *reflexive* is used in its mathematical sense to mean that each component is related (directly or indirectly) to the whole. As mentioned in Section 1, the term *autocatalytic* refers to the fact that the whole can be reconstituted through interactions amongst its components. A network qualifies as a RAFT network if it meets the following two criteria:

- (1) It is *reflexively autocatalytic*: each reaction $r \in \mathcal{R}'$ is catalyzed by at least one element type that is either produced by \mathcal{R}' or is present in the foodset F . This is sometimes referred to as *closure*.
- (2) It is *F-generated*: all reactants in \mathcal{R}' can be generated from the foodset F by using a series of reactions only from \mathcal{R}' itself.

Thus, a RAFT is a non-empty subset $\mathcal{R}' \subseteq \mathcal{R}$ of reactions that meets these two criteria: it is reflexively autocatalytic, and F-generated.

The term *catalytic reaction system* refers to a network consisting of components that can catalyze the generation of other components, and a catalytic reaction system can consist of one or more RAFTs. The largest RAFT, which subsumes all other RAFTs, is referred to as the *maxRAFT*. All other RAFTs are referred to as *subRAFTs*. A RAFT that cannot be broken down into smaller RAFTs is referred to as an irreducible RAFT, or *irrRAFT*. It is not necessarily the case that a catalytic reaction system contains a RAFT, but if it does contain one or more RAFTs, it has a unique *maxRAFT*. To put this more formally, if the network contains a RAFT, then the collection of all its RAFTs forms a partially ordered set (i.e., a poset) under set inclusion, with the *maxRAFT* as its unique maximal element. RAFTs can evolve, as demonstrated both mathematically and in simulation studies, through selective proliferation and drift acting on possible subRAFTs of the *maxRAFT* [32,75].

The catalytic reaction system is a tuple $\mathcal{Q} = (X, \mathcal{R}, C, F)$ consisting of a set X of types, a set \mathcal{R} of reactions, a catalysis set C indicating which molecule types catalyze which reactions, and a subset F of X referred to as the foodset. A subset \mathcal{R}' of the full reaction set \mathcal{R} of a catalytic reaction system \mathcal{Q} forms a RAFT if it is both *collectively autocatalytic* (by the

first criterion, because each of its reactions is catalyzed by some component in the system), and *self-sustaining* (because of the *F*-generated criterion).

RAFTs can enlarge and combine. The union of any two (or more) subRAFTs forms a RAFT (which explains why there is a unique maximal RAFT). These two subRAFTs may be disjoint, or they may have some reactions in common. A subRAFT \mathcal{R}' can also expand by combining with a 'co-RAFT', where a *co-RAFT* is any nonempty set of reactions that is not A RAFT but, when combined with \mathcal{R}' , forms A RAFT. RAFT expansion can also be extrinsically driven. For example, it can be due to social learning of a new story or skill, i.e., a change in the foodset. External stimuli may even trigger a 'reaction'; for example, the instruction to 'think creatively' may 'catalyze' the generation of new ideas. In a therapeutic context, this could take the form of a question or suggestion, such as to try seeing a particular interpersonal situation from the other person's perspective.

RAFTs emerge in a system of interacting components when their complexity passes a critical threshold [20,33]. In applications of RAFT networks to model the origin of life, the components are polymers: molecules made up of repeated units called monomers. In applications of RAFT networks to model cognitive networks, the components are MRs. The RAFT framework provides a means of analyzing the emergence of complex networks, identifying how phase transitions might occur, and at what parameter values. The phase transition from no RAFT to A RAFT has been analyzed (mathematically and through simulations), and applied to biochemical [31–33,36,42], cognitive [38–40], and ecological [76], systems.

During childhood, the individual assimilates experiences, stories, narratives, scripts, and schemas, and gradually weaves them into a network of understandings, and this process has been analyzed using the RAFT framework [16]. Eventually these pieces of knowledge condense into a maxRAFT, which grows and changes through childhood and beyond. Once the maxRAFT encompasses the majority of these fragments of knowledge they are mutually accessible. At this point, the child no longer requires a cue or reminder in the environment to access something from memory because the maxRAFT provides a route from any one idea to any other. The maxRAFT enables the individual to make plans and predictions, generate metaphors, and adapt old techniques or ideas to new circumstances; however, while an integrated maxRAFT network helps the individual think creatively and effectively negotiate the environment, it may be conducive to distorted thinking, and other biases that are *emotionally* dysfunctional, and result in mental health concerns.

An individual's worldviews could be said to be self-contained in that there exists a maxRAFT—meta-RAFT of sorts—that encompasses the majority of the individual's subRAFTs. We have modeled not just how this maxRAFT forms over the course of child development [16], but how the capacity for such a maxRAFT evolved over the course of human history [34,38–40]. The worldviews of different individuals are interconnected in that and subRAFTs of one individual are mirrored in subRAFTs of another, and indeed, RAFT structure can 'flow' and extend across individuals [17,18].

Cognitive RAFTs

When RAFTs are used to model cognition, all MRs in a given individual i are denoted X_i , and a specific MR $x = x_i$ in X_i is denoted by writing $x \in X_i$. MRs are either *foodset MRs*, or *foodset-derived MRs*. The *foodset* of individual i , denoted F_i , encompasses MRs that are either innately present, or that are the result of direct experience in the world, whether it be by way of social learning, or by way of natural or artificial stimuli. Thus, F_i has multiple components:

- \mathbb{S}_i denotes the set of MRs arising through direct experience that have been encoded in individual i 's memory. It includes:
 - MRs obtained through social learning from the communication of an MR x_j by another individual j , denoted $\mathbb{S}_i[x_j]$.
 - MRs obtained through individual learning, denoted $\mathbb{S}_i[\ell]$.
- Any *innate knowledge* with which individual i is born, denoted I_i .

F_i includes information obtained through social interaction with *someone else* who acquired this knowledge as a result of their own creative or analytical thought processes. (For example, if individual i learns from individual j that it is ok to say no, this is an instance of social learning, and “it’s ok to say no” becomes a member of F_i . In contrast, if individual i realizes on their own that it is ok to say no, then “it’s ok to say no” is not a member of F_i .) F_i includes everything in individual i ’s long-term memory that did not result from individual i engaging in RR. F_i also includes pre-existing information obtained by i through individual learning (which, as stated earlier, involves learning from the environment by non-social means), so long as this information retains the form in which it was originally perceived (and does not undergo redescription or restructuring through abstract thought). The crucial distinction between foodset and non-foodset items is not whether another person was involved, nor whether the MR was originally obtained through abstract thought (by *someone*), but whether the abstract thought process originated in the mind of the individual i in question.

Foodset-derived elements are denoted $\neg F_i$. Thus, $\neg F_i$ refers to mental contents that are *not* part of F_i (i.e. $\neg F_i$ consists of all the products $b \in B$ of all reactions $r \in R_i$). In particular, $\neg F_i$ includes the products of any reactions derived from F_i and encoded in individual i ’s memory. Its contents come about through mental operations *by the individual in question* on the foodset; in other words, foodset-derived items are the direct product of RR. Thus, $\neg F_i$ includes everything in long-term memory that *was* the result of one’s own thought processes. $\neg F_i$ may include a MR in which social learning played a role, so long as the most recent modification to this MR was a catalytic event (i.e., it involved RR).

A single instance of RR in individual i is referred to as a *reaction*, and denoted $r \in R_i$. RR is often applied recursively, such that the output of one thought serves as the input to the next. The set of reactions that can be catalyzed by a given MR x in individual i is denoted $C_i[x]$. The entire set of MRs either *undergoing* or *resulting from* r is denoted A or B , respectively, and a member of the set of MRs undergoing or resulting from reaction r is denoted $a \in A$ or $b \in B$. Thus, for example, if a client has the idea of expressing her grief at the passing of her father by painting a scene in which the clouds evoke her deceased father, the concepts FATHER and CLOUD are reactants in A , and the resulting concept FATHER-CLOUD is a product in B . This conceptual shift, treated as a ‘reaction’, is ‘catalyzed’ by the client’s desire to process the death of her father. It is in this way that the RAF approach tags novelty (in this case, the painting) with its point of origin (by showing in which mind in a cultural lineage it was a foodset-derived item).

The set of *all* possible reactions in individual i is denoted R_i . The mental contents of the mind, including all MRs and all RR events, is denoted $X_i \oplus R_i$. Recall that the set of all MRs in individual i , including both the food set and the food set-derived items, is denoted X_i . R_i and C_i are not prescribed up front; because C_i includes reminders and associations on the basis of one or more shared features, different kinds of interactions are possible between any given pair of MRs. Nonetheless, it makes sense mathematically to refer to R_i and C_i as sets.

5. Model

We now apply RAF theory to the modification of a client’s worldview in psychotherapy. To make this more concrete, we explain our model using a hypothetical interaction between a fictional yet representative therapist named Thera, and a client, named Clive. We show how the therapist elicits adaptive change in a dysfunctional belief in the client’s worldview.

5.1. Intake Form and Thera’s Emerging Mental Model of Clive

Thera learns from an intake form that Clive is a young man with no strong friendships, who has experienced debilitating social anxiety for years. His decision to start therapy was prompted by a recent incident in which his wife called him a ‘moron’ during an intense disagreement. This, in conjunction with several other earlier incidents, have forced him to conclude that he is ‘stupid’.

A portion of Thera’s mental model of her client after reading this report is shown in the first panel (panel (a)) of Figure 1. For the relationship between Thera and Clive to develop, they need to establish some form of psychological contact [77]. Thera welcomes Clive to the room and sits down. As she introduces herself, her body posture is relaxed. She provides several forms of non-verbal encouragement, such as smiling, nodding, supportive interjections, and eye contact. This makes him feel like someone worthy of the attention of another, which lifts his self-confidence, and allows him to speak more comfortably.

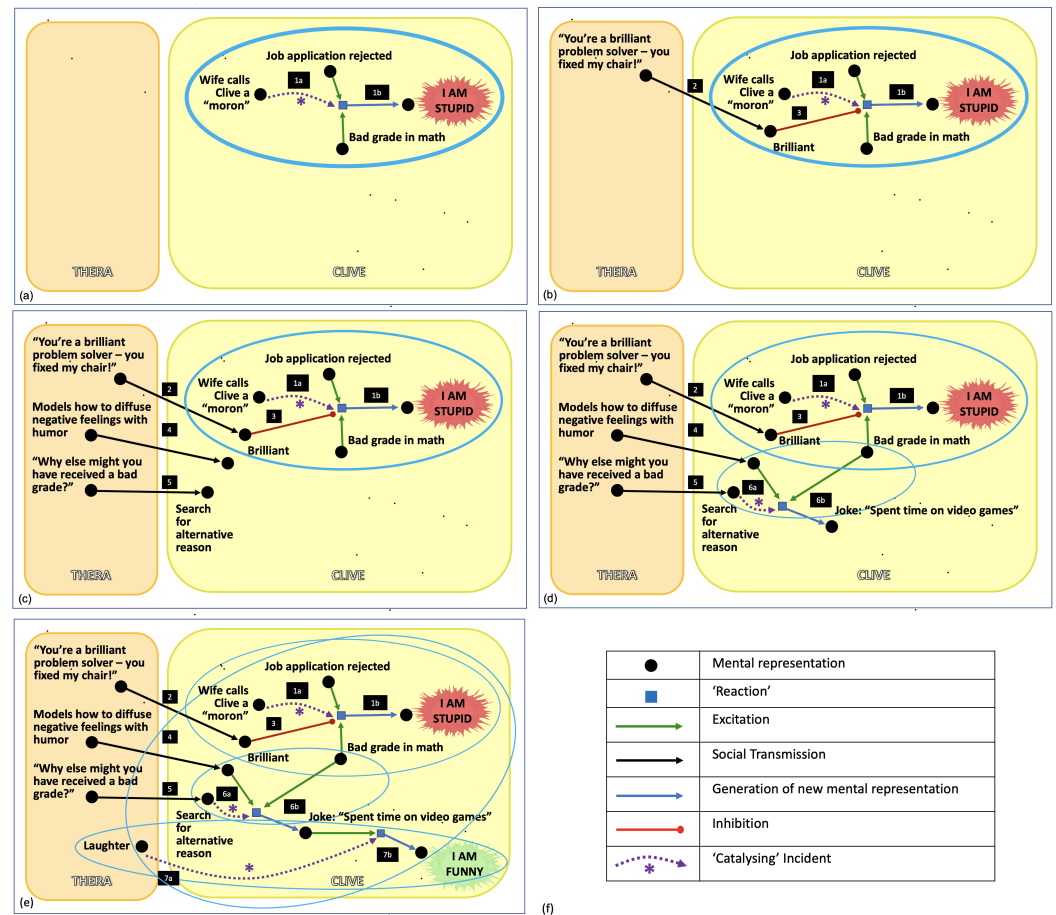


Figure 1. (a) RAF model of how client’s worldview is altered over the course of a psychotherapy session. Initially, Thera’s conception of Clive consists solely of what she read on his intake form. Following a ‘catalysing incident’ in which his wife called him a “moron”, he been interpreting other events as confirmation of the distressing belief, ‘I am stupid’. Collectively, these elements constitute a stable RAF, as indicated by the thick blue line forming an oval around them. The thickness of this line indicates that the RAF has a large impact on Clive’s thinking. (b) Thera praises Clive’s brilliant problem solving ability, which generates a new foodset item, the notion that he is ‘brilliant’. Since this is inconsistent with the belief ‘I am stupid’, it reduces the impact of that RAF, as indicated by the fact that the width of the line forming a blue oval is now thinner. (c) Two more foodset items are socially transmitted from Thera to Clive. (d) Making use of what Thera modeled for Clive about diffusing negative feelings using humor, he makes a joke. The joke is catalyzed by Thera’s prompt to explore alternate explanations for why he received a bad grade. The joke depletes negative feelings associated with the bad grade, such that it is less able to serve as a ‘reactant’ to support the belief that he is unintelligent, as illustrated by the further dissolution of the oval representing that RAF. His joke constitutes a second RAF. (e) Thera’s laughter at Clive’s joke catalyzes a new belief, ‘I am funny’, which enhances his self-esteem, and forms a third RAF. These first three RAFs, which are irr-RAFTs because they cannot be reduced further, interact with one another, and together form a maxRAF, which encompasses them all. (f) A key describing the symbols used in the various panels.

As psychotherapy proceeds, Thera starts to unearth information about Clive’s worldview using therapeutic techniques such as reflections, clarifications, and open-ended questions. Clive shares other significant life experiences that affected his self-esteem, such as getting a D- in a high school math class, and not getting a job after working hard to prepare for the interview. By asking questions such as, “Why do you think you received a poor grade?”, and “How did that make you feel?”, Thera is able to gather information regarding Clive’s interpretations of these events. From this, using her preexisting knowledge about social anxiety and mental health, she extrapolates Clive’s concerns, and builds a mental model of him in her mind.

While Clive’s belief that he is unintelligent is based in real-life adverse experiences—modeled here as ‘foodset items’—it involves extrapolation, and possibly distortion. It appears that the triggering incident in which his wife called him a moron served as a ‘catalyzing incident’ that initiated a tendency toward confirmation bias, such that he reinterprets other past and present events as confirmation of the belief, ‘I am stupid’. This, in turn, is damaging his self-esteem. The confirmation bias has thus exacerbated his preexisting concerns; his anxiety in social situations is now more severe due to his belief that he is ‘stupid’, and he has become isolated and lonely.

We can view what is happening here from the perspective of Clive’s worldview as a whole. The self-organizing, self-mending nature of a worldview can create a system of internal feedback combined with external influences that can sustain and amplify an existing (and in this case, negative) narrative [78]. Clive has come to interpret interactions with others as consistent with his negative self-image, and his mistrust and withdrawal have created a positive feedback system that reinforces his belief about social interactions.

To model this using the RAF approach and thereby understand it in more precise terms, the client, Clive, is denoted C , the poor grade is denoted G_C , and not getting the position he applied for is denoted P_C . These memories (and likely others) serve as the raw materials, or reactants, for Clive’s confirmation bias. First we note that they are part of his foodset, as follows:

$$G_C, P_C \in F_C. \quad (1)$$

G_C and P_C become reinterpreted as evidence for the belief “I am stupid”, denoted $b1_C$. The catalyzing event that initiates this, i.e., the fight with his wife where she called Clive a ‘moron’, is denoted by m_C . Thus, this process is described as follows:

$$G_C + P_C \xrightarrow{m_C} b1_C \in \neg F_C, \quad \neg F_C \mapsto \neg F_C \cup \{b1_C\}. \quad (2)$$

The catalysis of G_C and P_C by m_C is **Step 1a** in Panel (a) of Figure 1. The resulting formation of $b1_C$ is **Step 1b** in Panel (a) of Figure 1. The \neg sign indicates that $b1_C$ is not part of the foodset, i.e., it is a foodset-derived item. The portion of Equation (2) after the comma simply tells us that the set of foodset-derived items has expanded to include $b1_C$.

Note that this little cognitive network now satisfies the conditions for A RAF; (1) all reactions (in this case, there is just $r1$) can proceed, because the needed catalyst is present, and (2) the needed reactants (MRs of events that could be interpreted as confirmation of his lack of intelligence) are also present. The greater the extent to which ambiguous experiences are interpreted as evidence for the foregone conclusion that he is stupid, the greater to which the MR ‘I am stupid’ constitutes a stable attractor state. (For discussion of attractor states in psychology, see [14,79]). This attractor causes Clive emotional distress, and has an adverse impact on his quality of life.

We now show how Thera extrapolates from what Clive says to build a mental model of him in her mind. Thera interprets Clive’s statement as implying something more general about how he views the world and his place in the world: specifically, that his self-concept is increasingly dominated by the belief that he is (as his wife put it), “a moron”. We use the subscript T to refer to Thera. Her foodset, denoted F_T , includes knowledge of psychopathology and the treatment options available, as well as her growing knowledge of

Clive, C_T . The formation of this knowledge from Clive's discussion of himself, denoted C_C , is described as follows:

$$F_T \mapsto F_T \cup \{C_T\}, C_T \in \mathbb{S}_T[C_C]. \quad (3)$$

This equation tells us that her foodset now 'maps to' a foodset that includes knowledge of Clive, and the part after the comma tells us that knowledge of Clive was socially transmitted from Clive himself.

As Thera's understanding of Clive's concerns increases, she feels more emotionally connected to him, which in turn impacts the quality of her responses to him. Clive feels increasingly heard, and begins trusting her. Her reflections, statements, and questions may more readily facilitate adaptive change in Clive that may not have occurred otherwise. Thera's interactions create perturbations in Clive's cognitive network that disrupt the 'I am stupid' attractor state described above.

Elaborating on the 'Clive and Thera' example, we now show how the RAF model brings to light four distinct ways in which a therapist such as Thera facilitates therapeutic changes in the worldview of a client such as Clive.

5.2. Providing Counter-Evidence to Distressing Belief

During the session, Thera's chair malfunctions. Clive is able to fix the problem by adjusting the various knobs and gears on the chair. Thera says, "You're a brilliant problem solver—you fixed my chair!" This provides counter-evidence to the belief that he is unintelligent. Her observation, which lifts his mood, temporarily decreases his distress, and enhances his self-concept, is socially transmitted to him, resulting in the new MR: 'Someone thinks I'm brilliant'. Transmission of the information that he is brilliant, denoted B_C , is described by Equation (4), as follows:

$$F_C \mapsto F_C \cup \{B_C\}, \text{ where } B_C \in \mathbb{S}_C(B_T) \quad (4)$$

Thera's observation becomes part of Clive's 'mental foodset' because it did not come into existence within Clive's mind; it was 'born' in Thera's mind, and socially transmitted from Thera to Clive. His self-concept now contains both the constellation of experiences and negative beliefs about his intelligence, described as A RAF, consisting of multiple mutually consistent memories that support his belief that he is 'stupid', as well as a new experience that is inconsistent with this RAF. This experience of being described as 'brilliant' by Thera therefore has an *inhibitory* role on that RAF; it weakens the strength of that reaction, thereby diminishing the proclivity to interpret ambiguous events as confirmation of the belief, "I am stupid". This is depicted in Panel (b) of Figure 1, where **Step 2** refers to the social transmission of words of praise from Thera to Clive, and **Step 3** refers to its inhibitory role on the existing RAF.

5.3. Modeling Adaptive Mindset through Self-Disclosure

Thera models how to consciously resist the tendency to interpret ambiguous evidence in a negative manner through the use of self-disclosure regarding how she manages self-critical thoughts in her own life. She also spontaneously models adaptive responses in her interactions with him. She bumps her elbow on the desk, and then laughs at her clumsiness. The laughter enables her to re-frame the thought 'I am clumsy' into something innocuous. Thera's social transmission of this to him is **Step 4** in Panel (c) of Figure 1. Where the socially transmitted laughter is denoted L_C , this process is described as follows:

$$F_C \mapsto F_C \cup \{L_C\}, \text{ where } L_C \in \mathbb{S}_C(L_T) \quad (5)$$

The possibility of responding to one's inadequacies with laughter and/or self-deprecating humor is a new and striking concept for Clive. As such, this experience forms a foodset element in Clive's mind. His decision to use this strategy himself (i.e., to incorporate it as a reactant in subsequent steps) depends on a number of variables, such

as the degree to which he trusts and respects the therapist, his motivation to change, and so forth. (Elaboration of these variables is beyond the scope of this discussion).

5.4. Catalyzing Alternative Explanations

Thera not only transmits *existing* knowledge that Clive can import, wholesale, into his worldview, she also prompts the independent formation of *new* information, specifically tailored to his personality, that are conducive to adaptive perspectives and behaviors, and help him resolve or come to terms with the issues he faces. This could take the form of asking Clive questions that prompt him to reconsider existing beliefs, or by challenging Clive's beliefs directly. For example, Thera might ask him, "Might there be other reasons that you didn't do so well on that math test?" This question gently challenges Clive's forgone interpretation of the event. It prompts him to explore reasons other than that he is unintelligent. This is depicted in panel (c) of Figure 1, as **Step 5**.

Clive responds, laughing, "Well yeah, I spent a lot of time playing video games". This alternative interpretation does not play into the notion that Clive is unintelligent, and it shows that he has assimilated her proclivity to diffuse a negative conclusion with humor. It generates a *new* interpretation. This is depicted in Panel d of Figure 1, where **Step 6a** shows the catalyzing event (i.e., the search for an alternate explanation), and **Step 6b** shows the product of this 'reaction', Clive's joke.

We describe this in terms of RAFs as follows. Clive's memory of getting a bad grade on the test, denoted G_C , undergoes change, so it serves as a reactant that transforms through RR to the product, the joke about video games, denoted V_C . This new interpretation is provoked, or 'catalyzed' by Thera's question, denoted q_T . We describe this as follows:



This conceptual shift transforms one or more element(s) of the foodset F_C into a new foodset-derived MR, V_C , i.e., a member of $\neg F_C$. Clive's self-concept now contains a new MR—the joke he made—represented as a new node in his cognitive network.

5.5. Catalyzing a New Belief That Is Adaptive

When Clive dilutes the potency of this once-distressing memory with humor, Thera laughs. This laughter, denoted as l_T catalyzes a new belief in Clive's mind, 'I am funny', denoted as $b2_C$. This new belief buoys his self-esteem, and reduces his distress. The joke, denoted as J_C serves as a reactant. Thus, the reaction is described as follows:



This is depicted in Panel (e) of Figure 1, where **Step 7a** refers to the catalyzing event, i.e., Thera's laughter, and **Step 7b** refers to the resulting formation of the new belief, 'I am funny'.

Using these four distinct methods, Thera simultaneously reduces the strength of a distressing belief that was damaging his self-concept, and facilitate the creation of a new belief that enhances Clive's self-concept, as illustrated by the relative thickness of the RAFs in Figure 1.

We note there are now hierarchical levels of RAF structure, composed of interacting RAFs that collectively form a maxRAF (panel (e) of Figure 1) (The nodes in the therapist's mind associated with Steps 2, 4, and 5 in Figure 1 can either be included in this maxRAF or not, since from the RAF point of view they are merely copies of the corresponding elements in the client's worldview). We note also that RAF structure extends across the two individuals, providing a means of formally describing the dyadic relationship between therapist and client, and the emergence of a therapeutic alliance between them.

6. Discussion and Conclusions

We presented a RAF model of how a therapist fosters self-esteem and well-being in the client. The model illustrated four distinct ways by which a therapist accomplishes this: (1) providing direct examples/evidence contrary to a client's distressing belief about themselves, (2) challenging the client's existing interpretations of events, (3) using self-disclosure, provide examples of strategies for diffusing the potency of a negative belief, (4) reinforcing the client's attempts to assimilate such strategies in their own ways of thinking.

As discussed in Sections 3 and 4, RAF networks have been used to model the origins of evolutionary processes, biological (the origin of life) as well as cultural (the origin of cumulative innovation). We think this is not coincidental; indeed, elsewhere, we showed that both the evolution of early life and cultural evolution are instantiations of a primitive form of evolution—i.e., cumulative, adaptive, open-ended change—referred to as Self-Other Reorganization (SOR) [34,80,81]. Instead of replication using a self-assembly code, SOR entails internal self-organizing and self-maintaining processes within entities, as well as interaction between entities. The argument for SOR bolsters the argument that they share a deep structure, and thus strengthens the rationale for applying RAFs in both domains. In any case, the RAF approach to modeling therapeutic change is consistent with the theory that humans possess two levels of complex, adaptive, self-organizing structure: an organismal level, and a psychological level [43,45,82]. Psychological research tends to be data rich and theory poor [83], and psychological theorizing remains fragmented [84]. Psychotherapy research relies on momentary snapshots of the perceptions of client and therapist; it is vague about the nature of psychotherapeutic change, i.e., what happens at the level of mental representations and their interrelations and interactions, and how this kind of micro-level change alters the global structure of the client's worldview [5]. We take a first step towards such a global understanding in this paper. We posit that psychotherapeutic processes affect people not just at the individual level but at the society level, by providing a means to the creative transformation and cultural evolution of human worldviews.

Traditional methods for studying psychotherapeutic change have limitations [54] that the complex systems approach is well positioned to overcome [85–88], by enabling psychotherapy to be modeled and understood more precisely, using tools that embed it in a larger framework that includes other systems and disciplines. The above model of the therapeutic process provides a framework for empirical data collection and analysis. A next step is to incorporate into such a model specific factors that affect therapeutic outcomes (such as the degree of trust in the therapist). The impact of the therapeutic alliance between therapist and client on the therapy outcome is well-known [54,89]. It would be interesting to analyze psychotherapy sessions to track cognitive change, and the emergence of a therapeutic alliance, and its impact on this change. Our model accommodates the perspectives of both the therapist and client. While we have chosen to emphasize the client in this interaction, the RAF approach can also model potential changes in the therapist's worldview. The RAF approach could also be used to investigate a number of other issues related to psychopathology and treatment, such as the development of mental illness, the trajectory of various mental health concerns, and whether there are differences in the conceptual frameworks of individuals experiencing depression and those with anxiety. It could also be used to model the impact of different types of psychotherapy on conceptual network structure, and the impact of this structure on mental health and well-being. One promising possibility is to study whether individual differences in reliance on foodset versus foodset-derived information sources (i.e., the propensity to think things through for oneself) culminate in different kinds of conceptual networks, which might differentially affect therapeutic progress. In addition, using RAF networks to precisely model the psychotherapeutic process could be informative in the design and execution of computerized psychotherapies [90,91], or as an aid to the human psychotherapist for keeping track of, and visually depicting, specific interactions in the psychotherapy process and their outcome. We are a long way from this, but in keeping with the adage “a picture is worth a thousand words”, the RAF framework for psychotherapy could form the basis

for a software program that enables the therapist to visualize and identify change in RAF structure as it occurs over the course of psychotherapy, and to visualize desired possible future states of their clients' worldviews.

The RAF approach offers an established mathematical framework for integrating research on creative cognition, semantic networks, and the kinds of structures that exhibit cumulative, adaptive, open-ended change, i.e., that evolve, with a similarly dynamic process of psychotherapy. Though still in its infancy, it has the potential to provide a new way of understanding how the therapeutic alliance works, one that embeds psychotherapy research in the formal study of self-organizing structures and their role in evolutionary processes.

Author Contributions: Conceptualization, K.G. and L.G. writing, K.G. and L.G. All authors have read and agreed to the published version of the manuscript.

Funding: L.G. acknowledges funding from Grant 62R06523 from the Natural Sciences and Engineering Research Council of Canada. Funding was also provided by private donors, Susan and Jacques LeBlanc, for which are very grateful.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the writing of the manuscript, or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

CRS	Catalytic reaction system
RAF	Reflexively Autocatalytic foodset-derived network
MR	Mental representation
RR	Representational redescription
SOC	Self-organized criticality

Appendix A. Glossary of Terms

Abstract thought: the processing of internally sourced mental contents.

Autocatalytic: the whole can be reconstituted through interactions amongst its parts.

Catalyst: facilitates a transition that would otherwise be highly unlikely to occur. Here, the role of catalyst is played by a problem, desire, or need, or a realization or external stimuli that trigger a thought that would be highly unlikely to occur otherwise. For example, if a stranger on the street reminds you of a deceased relative, and this triggers a memory of being with that person, the stranger (or more precisely, your mental representation of the stranger plays the role of a catalyst.)

Catalytic reaction system: a network of interrelated parts, such as a conceptual network.

Closed RAF: A RAF that is stable unless the foodset changes or the reactions they take part in changes. (Formally, a closed RAF is A RAF that contains every reaction in the network that has each of its reactants and at least one catalyst present either in the foodset or as a product of some reaction in the RAF.) The maxRAF is always closed. The *closure* of any subRAF will contain the original subRAF, and be larger (unless the original subRAF was already closed).

Conceptual network: a web of shared properties, contexts, associations, and relationships of logic, causation, and so forth, that bind them together.

Co-RAF: a nonempty set of reactions that is not A RAF on its own, but that forms A RAF when combined with an existing set of reactions.

Catalytic reaction system: a network of components, such as a network of catalytic molecules, or a conceptual network.

Foodset, F : the elements that are initially present, as opposed to those that are the products of interactions amongst them.

foodset-derived (sometimes called F-generated, or foodset-derived), $\neg F$: an element that can be generated from the foodset F through a series of reactions in \mathcal{R}' itself. That is, an element of the network that is *not* part of the foodset. The term ‘foodset-derived’ is more often used in the cognitive application of RAFs.

Individual learning: obtaining pre-existing information from the environment by non-social means through direct perception.

IrrRAF: A RAF that is irreducible, i.e., cannot be broken down into smaller RAFs.

MaxRAF: the largest RAF in the network. It includes all other RAFs.

Mental representation (MR). Items in declarative or procedural memory composed of one or more concepts or percepts, and which came about through individual learning, social learning, or abstract thought. The set of all mental representations in individual i is denoted X_i . (As mentioned in the text, we emphasize that although we use the terms ‘mental representation’, we are sympathetic with the view that what we call mental representations do not ‘represent’, but act as contextually elicited bridges between the mind and the world.)

Reflexive: each part is related to the whole.

Reflexively autocatalytic: each reaction $r \in \mathcal{R}'$ is catalyzed by at least one element type that is either produced by \mathcal{R}' or is present in the foodset F .

Phase transition: rapid transition from one state to another.

Reactant: a mental representation that participates in a given ‘reaction’, i.e., an event that alters the structure of the conceptual network.

Reaction: a change of state or interaction between existing elements that results in a new element. The set of all possible reactions in individual i is denoted R_i .

Representational redescription (RR): conceptual restructuring that causes a mental representation to change. In the RAF framework this is modeled as a reaction.

Self-organized criticality (SOC): a phenomenon wherein, through simple local interactions, complex systems tend to find a critical state poised at the cusp of a transition between order and chaos, from which a single small perturbation occasionally exerts a disproportionately large effect.

SubRAF: A RAF that is not the maxRAF. It is a component of the maxRAF.

Transient RAF: a subRAF that is not closed. A transient RAF may add additional reactions until it becomes closed.

References

1. Dragioti, E.; Karathanos, V.; Gerdie, B.; Evangelou, E. Does psychotherapy work? An umbrella of meta-analyses of randomized control trials. *Acta Psychiatr. Scand.* **2017**, *136*, 236–246. [[CrossRef](#)] [[PubMed](#)]
2. Moldovan, R.; Pinetea, S. Mechanisms of change in psychotherapy: Methodological and statistical considerations. *Cogn. Brain Behav.* **2015**, *19*, 299–311.
3. Carey, T.A.; Huddy, V.; Griffiths, R. To mix or not to mix? A meta-method approach to rethinking evaluation practices for improved effectiveness and efficiency of psychological therapies illustrated with the application of perceptual control theory. *Front. Psychol.* **2019**, *10*, 1445. [[CrossRef](#)] [[PubMed](#)]
4. Kazdin, A.E. Understanding how and why psychotherapy leads to change. *Psychother. Res.* **2009**, *19*, 418–428. [[CrossRef](#)] [[PubMed](#)]
5. Carey, T.A.; Griffiths, R.; Dixon, J.E.; Hines, S. Identifying functional mechanisms in psychotherapy: A scoping systematic review. *Front. Psychiatry* **2020**, *11*, 291. [[CrossRef](#)] [[PubMed](#)]
6. Bassett, D.S.; Bullmore, E.D. Small-world brain networks. *Neuroscientist* **2006**, *12*, 512–523. [[CrossRef](#)] [[PubMed](#)]
7. Betzel, R.F.; Bassett, D.S. Generative models for network neuroscience: Prospects and promise. *J. R. Soc. Interface* **2017**, *14*, 20170623. [[CrossRef](#)] [[PubMed](#)]
8. Karuza, E.A.; Thompson-Schill, S.L.; Bassett, D.S. Local patterns to global architectures: Influences of network topology on human learning. *Trends Cogn. Sci.* **2016**, *20*, 629–640. [[CrossRef](#)]
9. Siew, C.S.; Wulff, D.U.; Beckage, N.M.; Kenett, Y.N. Cognitive network science: A review of research on cognition through the lens of network representations, processes, and dynamics. *Complexity* **2019**, *2019*, 2108423. [[CrossRef](#)]
10. Baronchelli, A.; Ferrer-i-Cancho, R.; Pastor-Satorras, R.; Chater, N.; Christiansen, M.H. Networks in cognitive science. *Trends Cogn. Sci.* **2013**, *17*, 348–360. [[CrossRef](#)]
11. Borge-Holthoefer, J.; Arenas, A. Semantic networks: Structure and dynamics. *Entropy* **2010**, *12*, 1264–1302. [[CrossRef](#)]

12. Kumar, A.; Steyvers, M.; Balota, D.A. A critical review of network-based and distributional approaches to semantic memory structure and processes. *Top. Cogn. Sci.* 2021, *in press*. [[CrossRef](#)] [[PubMed](#)]
13. Hofmann, S.G.; Curtiss, J.; McNally, R.J. A Complex Network Perspective on Clinical Science. *Perspect. Psychol. Sci.* **2016**, *11*, 597–605. [[CrossRef](#)] [[PubMed](#)]
14. Orsucci, F. *Mind Force: On Human Attractions*; World Scientific Publishing: Singapore, 2009.
15. Tschacher, W.; Haken, H. *The Process of Psychotherapy*; Springer Nature: Cham, Switzerland, 2019.
16. Gabora, L.; Beckage, N.; Steel, M. Modeling cognitive development with reflexively autocatalytic networks. *Top. Cogn. Sci.* 2022, *in press*.
17. Gabora, L.; Steel, M. From uncertainty to insight: An autocatalytic framework. In *Uncertainty: A Catalyst for Creativity, Learning and Development*; Beghetto, R., Jaeger, G., Eds.; Springer: Berlin/Heidelberg, Germany, 2022.
18. Ganesh, K.; Gabora, L. Modeling Discontinuous Cultural Evolution: The Impact of Cross-domain Transfer. *Front. Psychol.* **2022**, *13*, 786072. [[CrossRef](#)]
19. Farmer, J.D.; Kauffman, S.A.; Packard, N.H. Autocatalytic replication of polymers. *Phys. D Nonlinear Phenom.* **1986**, *22*, 50–67. [[CrossRef](#)]
20. Kauffman, S.A. *The Origins of Order*; Oxford University Press: Oxford, UK, 1993.
21. Erdős, P.; Rényi, A. On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci.* **1960**, *5*, 17–61.
22. Gabora, L. Revue of the ‘neurds’: Characterizing creative thought in terms of the structure and dynamics of human memory. *Creat. Res. J.* **2010**, *22*, 1–13. [[CrossRef](#)]
23. Aerts, D.; Aerts, S.; Gabora, L. Experimental evidence for quantum structure in cognition. In Proceedings of the International Symposium on Quantum Interaction, Saarbrücken, Germany, 25–27 March 2009; Bruza, P., Lawless, W., van Rijbergen, K., Sofge, D., Eds.; Springer: Berlin/Heidelberg, Germany, 2009; pp. 59–70.
24. Aerts, D.; Gabora, L.; Sozzo, S. Concepts and their dynamics: A quantum theoretical model. *Top. Cogn. Sci.* **2013**, *5*, 737–772. [[CrossRef](#)]
25. Aerts, D.; Broekaert, J.; Gabora, L.; Sozzo, S. Generalizing prototype theory: A formal quantum framework. *Front. Psychol.* **2016**, *7*, 418. [[CrossRef](#)]
26. Buskell, A.; Enquist, M.; Jansson, F. A systems approach to cultural evolution. *Palgrave Commun.* **2019**, *5*, 131. [[CrossRef](#)]
27. Gabora, L.; Aerts, D. Contextualizing concepts using a mathematical generalization of the quantum formalism. *J. Exp. Theor. Artif. Intell.* **2002**, *14*, 327–358. [[CrossRef](#)]
28. Gabora, L.; Aerts, D. Evolution as context-driven actualisation of potential: Toward an interdisciplinary theory of change of state. *Interdiscip. Sci. Rev.* **2005**, *30*, 69–88. [[CrossRef](#)]
29. Wang, Z.; Busemeyer, J.R. Interference effects of categorization on decision making. *Cognition* **2016**, *150*, 133–149. [[CrossRef](#)] [[PubMed](#)]
30. Steel, M.; Hordijk, W.; Xavier, J.C. Autocatalytic networks in biology: Structural theory and algorithms. *J. R. Soc. Interface* **2019**, *16*, 20180808. [[CrossRef](#)] [[PubMed](#)]
31. Hordijk, W.; Steel, M. Detecting autocatalytic, self-sustaining sets in chemical reaction systems. *J. Theor. Biol.* **2004**, *227*, 451–461. [[CrossRef](#)] [[PubMed](#)]
32. Hordijk, W.; Steel, M. Chasing the tail: The emergence of autocatalytic networks. *Biosystems* **2016**, *152*, 1–10. [[CrossRef](#)]
33. Mossel, E.; Steel, M. Random biochemical networks and the probability of self-sustaining autocatalysis. *J. Theor. Biol.* **2005**, *233*, 327–336. [[CrossRef](#)]
34. Gabora, L.; Steel, M. An evolutionary process without variation and selection. *J. R. Soc. Interface* **2021**, *18*, 20210334. [[CrossRef](#)]
35. Hordijk, W.; Steel, M. Autocatalytic sets and boundaries. *J. Syst. Chem.* **2015**, *6*, 1. [[CrossRef](#)]
36. Hordijk, W.; Hein, J.; Steel, M. Autocatalytic sets and the origin of life. *Entropy* **2010**, *12*, 1733–1742. [[CrossRef](#)]
37. Xavier, J.C.; Hordijk, W.; Kauffman, S.; Steel, M.; Martin, W.F. Autocatalytic chemical networks at the origin of metabolism. *Proc. R. Soc. London Ser. B Biol. Sci.* **2020**, *287*, 20192377. [[CrossRef](#)] [[PubMed](#)]
38. Gabora, L.; Steel, M. Autocatalytic networks in cognition and the origin of culture. *J. Theor. Biol.* **2017**, *431*, 87–95. [[CrossRef](#)] [[PubMed](#)]
39. Gabora, L.; Steel, M. Modeling a cognitive transition at the origin of cultural evolution using autocatalytic networks. *Cogn. Sci.* **2020**, *44*, e12878. [[CrossRef](#)] [[PubMed](#)]
40. Gabora, L.; Steel, M. A model of the transition to behavioral and cognitive modernity using reflexively autocatalytic networks. *Proc. R. Soc. Interface* **2020**, *17*, 20200545. [[CrossRef](#)] [[PubMed](#)]
41. Steel, M.; Xavier, J.C.; Huson, D.H. Autocatalytic networks in biology: Structural theory and algorithms. *J. R. Soc. Interface* **2020**, *17*, 20200488. [[CrossRef](#)] [[PubMed](#)]
42. Hordijk, W.; Kauffman, S.A.; Steel, M. Required levels of catalysis for emergence of autocatalytic sets in models of chemical reaction systems. *Int. J. Mol. Sci.* **2011**, *12*, 3085–3101. [[CrossRef](#)] [[PubMed](#)]
43. Barton, S. Chaos, self-organization, and psychology. *Am. Psychol.* **1994**, *49*, 5–14. [[CrossRef](#)] [[PubMed](#)]
44. Pribram, K.H. *Origins: Brain and Self-Organization*; Lawrence Erlbaum: Hillsdale, NJ, USA, 1994.
45. Maturana, H.; Varela, F. Autopoiesis and cognition: The realization of the living. In *Boston Studies in the Philosophy of Science*; Cohen, R.S., Wartofsky, M.W., Eds.; Reidel: Dordrecht, The Netherlands, 1973; Volume 42.
46. Varela, F.; Thompson, E.; Rosch, E. *The Embodied Mind*; MIT Press: Cambridge, MA, USA, 1991.

47. Mednick, S.A. The associative basis of the creative process. *Psychol. Rev.* **1962**, *69*, 220–232. [[CrossRef](#)]
48. Cautin, R.L. A century of psychotherapy, 1860–1960. In *History of Psychotherapy: Continuity and Change*; American Psychological Association: Washington, DC, USA, 2011; pp. 3–38. [[CrossRef](#)]
49. Finsrud, I.; Nissen-Lie, H.A.; Vrabel, K.; Høstmælingen, A.; Wampold, B.E.; Ulvenes, P.G. It's the therapist and the treatment: The structure of common therapeutic relationship factors. *Psychother. Res.* **2022**, *32*, 139–150. [[CrossRef](#)] [[PubMed](#)]
50. Lambert, M.J. Outcome in psychotherapy: The past and important advances. *Psychotherapy* **2013**, *50*, 42–51. [[CrossRef](#)] [[PubMed](#)]
51. Caspi, A.; Moffitt, T.E. All for one and one for all: Mental disorders in one dimension. *Am. J. Psychiatry* **2018**, *175*, 831–844. [[CrossRef](#)] [[PubMed](#)]
52. Bernhardt, I.S.; Nissen-Lie, H.A.; Rabu, M. The embodied listener: A dyadic case study of how therapist and patient reflect on the significance of the therapist's personal experience. *Psychother. Res.* **2020**, *31*, 682–692. [[CrossRef](#)] [[PubMed](#)]
53. Carey, T.A.; Tai, S.J.; Mansell, W.; Huddy, V.; Griffiths, R.; Marken, R.S. Improving professional psychological practice through an increased repertoire of research methodologies: Illustrated by the development of MOL. *Prof. Psychol. Res. Pract.* **2017**, *48*, 175–182. [[CrossRef](#)]
54. Hayes, A.M.; Andrews, L.A. A complex systems approach to the study of change in psychotherapy. *BMC Med.* **2020**, *18*, 197. [[CrossRef](#)] [[PubMed](#)]
55. Tschacher, W.; Meir, D. Physiological synchrony in psychotherapy sessions. *Psychother. Res.* **2019**, *30*, 558–573. [[CrossRef](#)] [[PubMed](#)]
56. Ramseyer, F.; Tschacher, W. Synchrony in dyadic psychotherapy sessions. In *Simultaneity*; Vrabel, S., Rössler, O., Marks-Tarlow, T., Eds.; World Scientific: Singapore, 2008.
57. Reich, C.M.; Berman, J.S.; Dale, R.; Levitt, H.M. Vocal synchrony in psychotherapy. *J. Soc. Clin. Psychol.* **2015**, *33*, 481–494. [[CrossRef](#)]
58. Ramseyer, F.; Tschacher, W. Nonverbal synchrony in psychotherapy. *J. Consult. Clin. Psychol.* **2011**, *79*, 284–295. [[CrossRef](#)]
59. Bachelor, A. Clients' and Therapists' Views of the Therapeutic Alliance: Similarities, Differences and Relationship to Therapy Outcome. *Clin. Psychol. Psychother.* **2011**, *20*, 118–135. [[CrossRef](#)]
60. Beghetto, R.A. Structured uncertainty: How creativity thrives under constraints and uncertainty. In *Creativity under Duress in Education? Resistive Theories, Practices, and Actions*; Mullen, C.A., Ed.; Springer International Publishing: Cham, Switzerland, 2019; pp. 27–40. [[CrossRef](#)]
61. Kauffman, S.A. *Humanity in a Creative Universe*; Oxford University Press: Oxford, UK, 2016.
62. Karmiloff-Smith, A. *Beyond Modularity: A Developmental Perspective on Cognitive Science*; MIT Press: Cambridge, MA, USA, 1992.
63. Bowers, K.S.; Farvolden, P.; Mermigis, L. Intuitive antecedents of insight. In *The Creative Cognition Approach*; Ward, S., Finke, R.A., Eds.; MIT Press: Cambridge, MA, USA, 1995; pp. 27–51.
64. Beghetto, R.; Jaeger, G. *Uncertainty: A Catalyst for Creativity, Learning and Development*; Springer: Berlin/Heidelberg, Germany, 2021.
65. Cabell, K.R.; Valsiner, J. *The Catalyzing Mind: Beyond Models of Causality, Annals of Theoretical Psychology*; Springer: Berlin/Heidelberg, 2016; Volume 11.
66. Barsalou, L.W. Context-independent and context-dependent information in concepts. *Mem. Cogn.* **1982**, *10*, 82–93. [[CrossRef](#)]
67. Hampton, J.A. Disjunction of natural concepts. *Mem. Cogn.* **1988**, *16*, 579–591. [[CrossRef](#)] [[PubMed](#)]
68. Osherson, D.N.; Smith, E.E. On the adequacy of prototype theory as a theory of concepts. *Cognition* **1981**, *9*, 35–58. [[CrossRef](#)]
69. Veloz, T.; Tëmkin, I.; Gabora, L. A conceptual network-based approach to inferring the cultural evolutionary history of the Baltic psalter. In Proceedings of the 34th Annual Meeting of the Cognitive Science Society, Sapporo, Japan, 1–4 August 2012; Miyake, N., Peebles, D., Cooper, R.P., Eds.; Cognitive Science Society: Austin, TX, USA, 2012; pp. 2487–2492.
70. Gabora, L.; Leijnen, S.; Veloz, T.; Lipo, C. A non-phylogenetic conceptual network architecture for organizing classes of material artifacts into cultural lineages. In Proceedings of the 33rd annual meeting of the Cognitive Science Society, Boston, MA, USA, 20–23 July 2011; Carlson, L., Holscher, C., Shipley, T.F., Eds.; Cognitive Science Society: Austin, TX, USA, 2011; pp. 2923–2928.
71. Gabora, L.; Tseng, S. The social benefits of balancing creativity and imitation: Evidence from an agent-based model. *Psychol. Aesthet. Creat. Arts* **2017**, *11*, 457–473. [[CrossRef](#)]
72. Gabora, L.; Smith, C. Two cognitive transitions underlying the capacity for cultural evolution. *J. Anthropol. Sci.* **2018**, *96*, 27–52. [[PubMed](#)]
73. Sousa, F.; Hordijk, W.; Steel, M.; Martin, W. Autocatalytic sets in *E. coli* metabolism. *J. Syst. Chem.* **2015**, *6*, 4. [[CrossRef](#)]
74. Steel, M. The emergence of a self-catalyzing structure in abstract origin-of-life models. *Appl. Math. Lett.* **2000**, *13*, 91–95. [[CrossRef](#)]
75. Vasas, V.; Fernando, C.; Santos, M.; Kauffman, S.; Szathmáry, E. Evolution before genes. *Biol. Direct* **2012**, *7*, 1. [[CrossRef](#)] [[PubMed](#)]
76. Cazzolla Gatti, R.; Fath, B.; Hordijk, W.; Kauffman, S.; Ulanowicz, R. Niche emergence as an autocatalytic process in the evolution of ecosystems. *J. Theor. Biol.* **2018**, *454*, 110–117. [[CrossRef](#)]
77. Rogers, C. The necessary and sufficient conditions of therapeutic personality change. *J. Couns. Psychol.* **1956**, *21*, 95–103. [[CrossRef](#)]
78. O'Connor, B.; Gabora, L. Applying complexity theory to a dynamical process model of the development of pathological belief systems. *Chaos Complex. Lett.* **2009**, *4*, 75–96.

79. Hiver, P. Attractor States. In *Motivational Dynamics in Language Learning*; Dornyei, Z., MacIntyre, P., Henry, A., Eds.; Multilingual Matters: Bristol, UK, 2015; pp. 20–28.
80. Gabora, L. From deep learning to deep reflection: Toward an appreciation of the integrated nature of cognition and a viable theoretical framework for cultural evolution. In Proceedings of the 2019 Annual Meeting of the Cognitive Science Society, Montreal, QC, Canada, 24–27 July 2019; Nadel, L., Stein, D., Eds.; Cognitive Science Society: Austin, TX, USA, 2019; pp. 1801–1807.
81. Gabora, L. Creativity: Linchpin in the quest for a viable theory of cultural evolution. *Curr. Opin. Behav. Sci.* **2019**, *27*, 77–83. [[CrossRef](#)]
82. Gabora, L. Ideas are not replicators but minds are. *Biol. Philos.* **2004**, *19*, 127–143. [[CrossRef](#)]
83. Fried, E.I. Lack of theory building and testing impedes progress in the factor and network literature. *Psychol. Inq.* **2020**, *31*, 271–288. [[CrossRef](#)]
84. Teo, T. *Outline of Theoretical Psychology: Critical Investigations*; Springer: Berlin/Heidelberg, Germany, 2018.
85. Orsucci, F. Introduction. In *Human Dynamics: A Complexity Science Open Handbook*; Orsucci, F., Ed.; Nova Science Publishers: New York, NY, USA, 2016; pp. 119–133.
86. de Felice, G.; Giuliani, A.; Halfon, S.; Andreassi, S.; Paoloni, G.; Orsucci, F. The misleading Dodo Bird verdict. How much of the outcome variance is explained by common and specific factors? *New Ideas Psychol.* **2019**, *54*, 50–55. [[CrossRef](#)]
87. de Felice, G.; Giuliani, A.; Gelo, O.C.G.; Mergenthaler, E.; De Smet, M.M.; Meganck, R.; Paoloni, G.; Andreassi, S.; Schiepek, G.K.; Scozzari, A.; et al. What Differentiates Poor- and Good-Outcome Psychotherapy? A Statistical-Mechanics-Inspired Approach to Psychotherapy Research, Part Two: Network Analyses. *Front. Psychol.* **2019**, *11*, 788. [[CrossRef](#)] [[PubMed](#)]
88. de Felice, G.; Orsucci, F.; Scozzari, A.; Gelo, O.C.G.; Serafini, G.; Andreassi, S.; Vegni, N.; Paoloni, G.; Lagetto, G.; Mergenthaler, E.; et al. What Differentiates Poor and Good Outcome Psychotherapy? A Statistical-Mechanics-Inspired Approach to Psychotherapy Research. *Systems* **2019**, *7*, 22. [[CrossRef](#)]
89. Fluckiger, C.; Del, A.C.; Wampold, B.E.; Horvath, A.O. The alliance in adult psychotherapy: A meta-analytic synthesis. *Psychotherapy* **2018**, *55*, 316–340. [[CrossRef](#)]
90. Liu, Z.; Qiao, D.; Xu, Y.; Zhao, W.; Yang, Y.; Wen, D.; Li, X.; Nie, X.; Dong, Y.; Tang, S.; et al. The efficacy of computerized cognitive behavioral therapy for depressive and anxiety symptoms in patients With COVID-19: Randomized controlled trial. *J. Med. Internet Res.* **2021**, *23*, e26883. [[CrossRef](#)]
91. Foroushani, P.; Schneider, J.; Assareh, N. Meta-review of the effectiveness of computerised CBT in treating depression. *BMC Psychiatry* **2011**, *11*, 131. [[CrossRef](#)]