The computations of a traumatized mind: a latent cause model of Post-Traumatic Stress Disorder

Francesco Rigoli, City University of London

Abstract

In cognitive psychology, a recent perspective based on the notion of Latent Cause (LC) has offered new insight on how learning and memory work. Here I explore the implications of this novel perspective to understand Post-traumatic Stress Disorder (PTSD). The proposal is that, because of a propensity to interpret events as manifestations of multiple LCs (a propensity facilitated by experiencing traumas in childhood), PTSD patients would form a LC associated with the trauma which would be responsible for typical symptoms of the illness (specifically, intrusive symptoms and associated fear). Moreover, after the trauma, some patients would develop a second LC, now associated with presence of trauma-related cues combined with absence of danger. Development of the latter LC would interfere with extinction and explain why, for some patients, exposure to trauma-related cues (even when supported by interventions such as exposure protocols) fails to provide much improvement. This proposal has potential clinical implications, raising the possibility that some patients might benefit from exposure to mildly painful aspects of the trauma in conjunction with trauma-related cues.

Keywords: trauma; PTSD; post-traumatic stress disorder; latent cause; structure learning; Bayesian; exposure therapy

Introduction

Months after being involved in a severe car accident, Philip continued having recurrent nightmares about the accident, experienced distressing images and memories about it, and was very anxious every time he drove the car again. Moreover, at times, he had flashbacks where he re-experienced the moments when, right after the crash, he was entrapped in the car with an excruciating pain and panicking about his two daughters, who were also there (likely, the two girls eventually reported only minor injuries). This brief description is typical of Post-Traumatic Stress Disorder (PTSD), a mental illness characterised by two broad sets of symptoms: (i) pervasive negative thoughts and affect (with associated physiological repercussions such as hyperarousal and recklessness) and (ii) intrusive symptoms such as images, thoughts, nightmares, and flashbacks about the traumatic event, accompanied by intense fear.[1] Intrusive symptoms tend to arise preferentially following exposure to trauma-related cues. Because of the distress accompanying the symptoms, patients often avoid trauma-related cues in order to prevent the symptoms to arise.

What are the psychological mechanisms responsible for the development and maintenance of PTSD? Classical conditioning theories provide some of the most insightful approaches to this question.[2] These rely on the simple idea that, when a highly aversive experience (an unconditioned stimulus (US)) occurs in a certain situation (a conditioned stimulus (CS)), an association is established between the CS and the US. Thus, subsequent exposures to the CS will elicit an expectation of the US and in turn a strong fear response. This mechanism would underpin PTSD, where the traumatic event and trauma-related cues would play the role of US and CSs, respectively. This perspective prescribes repeated exposure to the CS as treatment.[3-7] This is based on the reasoning that, if the CS is presented time after time, its association with the US should progressively dissolve. However, spontaneous behaviour of PTSD patients would often hinder extinction insofar as patients tend to avoid trauma-related cues (given the distress elicited by these cues), thus preventing any experience of these cues in the absence of the US. Exposure therapy would help patients to resist avoiding trauma-related cues, hence enabling extinction to occur. Classical conditioning theory offers an elegant explanation of why exposure to trauma-related cues (the CS) is likely to trigger intense fear, which is a key aspect of

PTSD. Furthermore, its recommendation to adopt exposure as treatment has proved to be successful.[3-7] However, albeit being one among the most well-established treatments of PTSD, exposure therapy fails to be resolutive for some patients: following repeated re-experience of trauma-related cues in safe conditions, some patients report only limited improvement, while another group of PTSD sufferers scarcely manifest any progress at all.[3-7] Understanding why exposure therapy is of limited effectiveness with some patients appears to be paramount. With this regard, possible factors have been highlighted by research:[3-7] for example, certain psychiatric medications (e.g., benzodiazepines) appear to interfere with the efficacy of exposure therapy, and dropouts are not uncommon (e.g., due to stigma, logistical barriers, perceived need for treatment, or scarce toleration of exposure). Yet, these factors do not fully explain why exposure therapy is of limited effectiveness with some patients. Insight on this can be gained by assessing recent advancements in learning and memory literature, and by examining how these advancements might inform our understanding of PTSD. Based on this logic, here I consider a recent new perspective on learning and memory and assess its implications for understanding PTSD. This new perspective is grounded upon the notion of *Latent Cause* (LC), arguing that, based on experience, the brain infers the existence of a set of LCs, each associated with a cluster of probabilistic events.[8-15] This perspective has shed new light on apparently puzzling phenomena in the domain of learning and memory (as will be examined below). Here I explore the possibility of interpreting traumatic memories in PTSD as manifestation of a LC learnt during the traumatic event. After a brief overview of previous research on LCs, the paper assesses whether this line of enquiry can provide any new insight to understand PTSD.

Latent cause theory

The concept of LC has been thoroughly investigated in the context of classical conditioning paradigms.[8-15] In these paradigms, during the acquisition phase, a CS is repeatedly paired with a US with some probability. When this phase is completed, an animal typically exhibits a fear response when presented with the CS alone. This is sometimes followed by an extinction phase in which the CS is presented many times without any US. This phase is characterised by a progressive diminution

of the fear response. Traditional interpretations of the mental processes engaged during classical conditioning (e.g., the Rescorla-Wagner rule) posit that every new experience affects the strength of the association between the CS and the US, in turn determining the vigour of the ensuing fear response.[16] Thus, the acquisition phase would strengthen the CS-US association, leading to a progressively more vigorous fear response, whereas extinction would loosen the CS-US association and diminish the response vigour. However, a puzzling observation for standard interpretations of classical conditioning is the evidence that, after being extinguished, the fear response often re-emerges spontaneously.[17-19] For instance, although immediately after extinction no fear response is displayed, this response might reappear later when an animal is placed again in the experimental chamber. Spontaneous fear reinstatement is hard to reconcile with traditional models of classical conditioning, because these predict that any CS-US link should be vanished after extinction.

To explain spontaneous fear reinstatement (as well as other puzzling aspects of aversive behaviour), it has been argued that learning in the brain relies on inferring the LCs underlying sensory experiences.[11,12] A LC is a discrete category, distinct from alternative categories, assumed to be the cause of a set of sensory inputs. When, expressed by an array of sensory inputs, a new event occurs, the brain would perform two main forms of inference (typically modelled in Bayesian terms, where novel evidence is integrated with prior beliefs).[12,20] First, it would infer what the underlying LC is, with options being that the event belongs either to a familiar LC (i.e., one already experienced in the past) or to a new LC (i.e., one never experienced before). This inference depends on judging to what extent the event (specifically, the accompanying array of sensory inputs) resembles the prototypical expression of each LC; if the resemblance is poor, a new LC is invoked. Establishing whether a new LC should be postulated would not depend only on how well the event fits with available LCs, but also on prior beliefs about whether events tend to cluster to few LCs or to spread among many LCs. This prior belief is captured by a *spreading parameter*, with higher values of this parameter capturing a tendency to invoke new LCs for new events.[12,20] Experiencing a new event is accompanied by a second form of inference: if the event is judged to belong to a familiar LC, then the brain would need to update its beliefs about that LC (formally, to update its parameters), whereas, if the event is judged

to belong to a new LC, then the brain would have to estimate the characteristics (i.e., parameters) of this LC anew.

Let us apply this logic to interpret classical conditioning.[11,12] During acquisition, a new LC (a conditioning LC) would be formed, with both the US and the CS combined reflecting its sensory expression. At the start of extinction, despite exposure to the CS, absence of the US would be interpreted by an animal not as an expression of the conditioning LC, but of a newly encountered LC (a CS-alone LC). In other words, after extinction, two LCs would now coexist in memory, one (the conditioning LC) associated with a co-occurrence of both the US and CS, and the other (the CS-alone LC) associated with occurrence of the CS but with absence of the US. Thus, the conditioning LC would be totally unaffected by extinction. The consequence of this would be that, after some time, presenting the CS might activate either LC, thus explaining the spontaneous reinstatement of the fear response (occurring when the conditioning LC is retrieved from memory).

In addition to explaining spontaneous fear reinstatement, this perspective offers a possible solution for establishing a durable extinction.[13] To ensure such durable extinction, it is necessary to devise a strategy to access the conditioning LC in such a way that this can be modified, preventing the formation of a CS-alone LC. This strategy consists in a gradual, rather than abrupt, extinction.[13] For example, a gradual extinction is obtained by presenting a progressively milder US following the CS, the logic being that, now that both CS and US co-occur (albeit the latter in a milder form), the brain will attribute the event to the conditioning LC, rather than envisaging a new LC. At the same time, it is important that, during extinction, a progressively milder US is delivered, in order to elicit new learning and modify the beliefs about the conditioning LC. If carried out with balance (the intensity of the US should be neither too weak, to avoid creation of the CS-alone LC, nor too strong, to ensure new learning), the prediction is that gradual extinction should ultimately erase the expectation (encoded in the conditioning LC) that the CS is associated with the US. This prediction has received initial empirical support by data showing that a gradual extinction can prevent spontaneous fear reinstatement.[13]

Overall, the LC framework offers a compelling new perspective about the mental processes involved during classical conditioning. Given that in PTSD a traumatic experience plays a role similar to the role played by a US in classical conditioning, applying the LC framework to understanding PTSD appears as a promising endeavour. Thus, the next section proposes a Latent Cause Theory of PTSD (LCT-PTSD).

A Latent Cause Theory of PTSD

Let us follow the vicissitudes of three fictional characters: Philip, Tom, and Jack. Despite all three experience the same severe car accident, Philip alone develops PTSD after the accident. Let as apply the logic followed by the LCT-PTSD to explain why Philp, but not Tom nor Jack, ends up suffering from this mental illness. Within the LCT-PTSD, two time points can be regarded as critical; the time of the car accident, and the few months after the accident has occurred (fig. 1). Let us start focusing on the time of the car accident. Like any other extremely negative event (e.g., war, natural disasters, rape, etc.), the car accident urges the brain to establish whether this event maps to a previous or to a new LC. Some people might invoke a new LC, while others might assign the event to an old LC (fig. 1). Regarding our three characters, let us suppose that Philip and Jack associate the car accident to a new LC, while Tom does not. The LCT-PTSD argues that: (i) a prerequisite for PTSD to develop is that the brain assigns an extremely negative event to a new LC (as in Philip's case); (ii) assigning the event to a new LC is not sufficient for PTSD to emerge (as in Jack's case); (iii) people who assign the event to an old LC will not develop PTSD (as in Tom's case).

Below, we will examine why, according to the LCT-PTSD, assigning the event to a new LC is necessary but not sufficient for PTSD to ensue. But first let us ask another question: why would some people (e.g., Tom) assign the same extremely negative event to an old LC while other people (Jack and Philip) assign it to a new LC? Two factors might be at play. A first factor might be to what extent the extremely negative event fits with available latent LCs. Paradoxically, people who are more accustomed with similar negative events might be less predisposed to invoke a new LC. For example, imagine that Tom is a medical doctor often working in the ambulance, and hence familiar with scenes of severe car accidents (albeit involving other people). This prior experience might induce Tom to

assign the occurrence of the own car accident to a previous LC associated with being in car accident settings. Conversely, imagine that Jack and Philip have no previous experience of car accidents at all. This lack of experience will induce them to invoke a new LC associated with their own car accident. The LCT-PTSD proposes that a second factor affecting whether an extremely negative event is mapped to a new LC or not is the spreading parameter (as introduced above), capturing a propensity to associate new LCs to new events.[10] In our example, imagine that Tom has a low spreading parameter, Jack has an intermediate parameter, and Philip has a high parameter. Individuals characterised by higher spreading parameter would be more likely to assign the event to a new LC. This raises the question of where the spreading parameter originates. The LCT- PTSD argues that, in conjunction with predisposing genetic factors, past experience concerning negative events plays a key role in shaping this parameter. Making negative experiences that are very diverse would teach the brain that the world's structure encompasses several LCs, each associated with only few events. Conversely, less variegated negative experiences would promote the belief that fewer LCs exists, with numerous events clustering around each LC. This idea implicates that experiencing highly negative and unusual events in childhood (e.g., physical or sexual abuse, severe illnesses, or death of intimate figures) might promote a higher spreading parameter, in turn predisposing towards interpreting highly negative events occurring in adulthood as belonging to a new LC. This fits with empirical evidence indicating that traumas occurred in childhood predispose adults to develop PTSD.[21,22] Within the LCT-PTSD this predisposition is reflected in a high spreading parameter.

Let us now analyse why, according to the LCT-PTSD, PTSD is not developed by people who assign the extremely negative event to an old LC (fig. 1). Consider Tom as an example. Tom connects the own car accident to the old LC associated with being in car accident settings. This new experience prompts Tom to update his beliefs about the old LC, which now will appear as more negative. However, the level of negativity for this LC will not be extreme; this is because the LC is associated with the experience of the own car accident, arguably imbued with extreme negative value, but also with other situations, which are appraised as not so extreme. The consequence of an increased negative value attributed to the old LC might be a mild increase in anxiety when, right after the own

car accident has occurred, the LC is activated (e.g., when driving or when working in the ambulance again). However, every time this LC is retrieved but nothing as bad happens, beliefs are updated in a way that the LC will appear as progressively less negative. In a relatively short amount of time, any anxiety that can be traced back to the own car accident will be vanished.

Let us now consider what happens to Jack and Philip, who both attribute the car accident to a new traumatic LC. Being such LC new, it encompasses only memory traces associated with the traumatic event. Thus, for Jack and Philip, presentation of trauma-related cues will activate the traumatic LC and retrieve trauma-related memories. Intrusive symptoms (e.g., images, thoughts, flashbacks) can be interpreted as the product of retrieving such trauma-related memories elicited by the activation of the traumatic LC. According to the LCT-PTSD, intrusive symptoms are initially to be expected for everyone who forms a new traumatic LC after an extremely negative event. In our example, these symptoms are initially experienced by both Jack and Philip (although, as we shall see below, they will become chronic only for Philip). Note that something different occurs to Tom, who incorporates the car accident event within an old LC and thus with previous information. Within the old LC, memories of the car accident episode are integrated with previous memories. In this way, exposure to cues related with the own car accident will evoke general memories about car accidents (not only those specific to the own accident), thus preventing intrusive symptoms to arise. This explains why, after the car accident, Tom never experiences intrusive symptoms.

So far, the destiny of Jack and Philip has been the same. At which time point do their pathways diverge? According to the LCT-PTSD, the second critical period for the development of PTSD, which separates Philip from Jack, is the few months after the negative event (fig. 1). During this period, in different occasions both Jack and Philip are exposed to trauma-related cues. This exposure will trigger memories of the event in the form of intrusive symptoms and associated fear. However, at a more careful look, for Jack and Philip very different cognitive processes might be at play during exposure to these cues. Let us have a better look at these processes, starting with Jack. For him, exposure to cues will activate the traumatic LC. Lack of any negative outcome will prompt a belief updating about this LC, which will be progressively associated with less negative value. In turn, this will reduce the

intensity of fear when exposed to cues the next time. If repeated, this new learning about the traumatic LC will progressively diminish both intrusive symptoms and the associated fear, eventually leading to a recovery from the distress initially caused by the negative event. This process is not far from the effect of exposure as traditionally described by classical conditioning models.[2]

Let us consider what happens to Philip. Like Jack, Philip would experience intense fear and intrusive symptoms when exposed to trauma-related cues. However, when nothing bad happens following the cues, Philip's brain would come up with a rather different interpretation compared to Jack's brain: Philip's brain would interpret presence of cues combined with lack of any negative event as a manifestation of yet another new LC (an interfering LC).[11,12] Thus, according to the LCT-PTSD, two separate LC would now coexist in Philip's brain, one (the traumatic LC) linked with the negative event combined with associated cues, and the interfering LC linked with the cues combined with lack of any negative event. The problem of postulating another LC would be that exposure to cues in the absence of any negative event would prevent any new learning regarding the traumatic LC. Thus, a vicious circle is proposed to arise for Philip: extinction would not work because lack of any negative event in conjunction with presence of cues would not elicit any learning about the traumatic LC: the traumatic LC would remain intact, impervious to any new learning. The consequence would be the persistence of intrusive symptoms and of intense fear when the traumatic LC is engaged following trauma-related cues, resulting in full-fledged PTSD.

One last key question to understand how the LCT-PTSD explains PTSD is why Philip's path diverges from Jack's. In other words, why would some, but not other, people who have developed a traumatic LC postulate yet another LC when exposed to cues in the absence of any negative event? A possible critical determinant is once again the spreading parameter, which determines a propensity to interpret events as manifestations of new LCs (above, we have already discussed which factors are important in shaping this parameter).[10] While Jack has an intermediate spreading parameter, Philip has a high one, explaining why, more often than Jack, he tends to postulate new LCs when facing new events.

In summary, the LCT-PTSD applies latent cause theory to interpret PTSD. It proposes that formation of a traumatic LC is a necessary, but not sufficient, condition for PTSD to develop. Activation of the traumatic LC after exposure to trauma-related cues is proposed to elicit intrusive symptoms and intense fear. In addition to the formation of the traumatic LC, a second necessary condition for the insurgence of full-fledged PTSD is proposed to be the formation of an interfering LC associated with trauma-related cues combined with absence of the negative event. This interfering LC is proposed to block new learning about the traumatic LC, implying that the capability of the latter to trigger intrusive symptoms and fear remains unabated. The next section explores the implication of this line of reasoning for treating PTSD.

Implications for treatment

Let us introduce a fourth fictional character: Marc. Marc has a story analogous to Jack's: he has experienced a severe car accident, he has interpreted the event by forming a traumatic LC, and (like Jack but unlike Philip) he activates the traumatic LC when exposed to trauma-related cues. Thus, like Jack, Marc can potentially modify the traumatic LC in light of new experience. However, while Jack realises this potential by means of repeated exposure to trauma-related cues, Marc staunchly avoids these cues. This is common among people who have faced traumas, because trauma-related cues elicit symptoms that are highly discomforting.[2] Thus, to avoid this discomfort, cues are often avoided. However, in the long run avoidance prevents new learning about the traumatic LC, thus maintaining symptoms. As a result, while after few months since the trauma Jack spontaneously recovers, for Marc symptoms persist up to a time when he receives a PTSD diagnosis.

Marc is a perfect candidate for exposure therapy, which is among the most effective and popular treatments of PTSD.[3-7] Among several versions developed, Prolonged Exposure (PO) therapy is one of the most powerful exposure protocols.[3,5,23] PO therapy includes both imaginal and in vivo exposure.[24] During the former, usually occurring in sessions with the therapist, the patient is invited to recollect aspects of the traumatic episode and to discuss any ensuing thought and emotion. These

sessions are recorded so that the patient can listen to them again for further elaboration. In vivo exposure, usually performed by the patient alone in-between sessions, consists in approaching people, places, and objects associated with the trauma. The logic behind PO therapy, and behind any exposure protocol in general, is that re-experiencing trauma-related cues as safe will persuade the patient that these cues are not dangerous, in such a way that eventually these cues will stop eliciting symptoms.

PO therapy is ideal for people like Marc, who tend to avoid trauma-related cues. PO therapy can guide Marc to approach these cues. According to LCT-PTSD, by presenting Marc with trauma-related cues in the absence of any danger, PO therapy can help Marc updating beliefs about the traumatic LC, in such a way that, with time, the cues will not be associated with danger anymore. Hence, for people like Marc, the outcome of the therapy is likely to be the remission of symptoms.

While PO therapy is highly effective with patients like Marc, it might struggle to help patients like Philip, who do not update the traumatic LC when exposed to trauma-related cues in the absence of danger. For these patients, targeting and modifying the traumatic LC appears to be particularly tricky. The LCT-PTSD might help understanding why some patients (such as Philip) fail to benefit from PO therapy in a substantial way: by focusing on exposure to harmless trauma-related cues, standard PO techniques might fail to modify the traumatic LC for these patients. Moreover, the LCT-PTSD can also suggest possible solutions. Following animal studies adopting gradual extinction,[13] presenting trauma-related cues combined with some mildly negative aspects associated with the trauma might facilitate update of the traumatic LC. For example, during imaginal exposure, the therapist might encourage the patient not to emphasise only innocuous cues, but also painful aspects linked with the trauma. Of course, how this can be translated into a concrete protocol requires careful considerations of complex practical issues, but the basic idea is clear: exposure to negative aspects of the trauma, and not only to innocuous trauma-related cues, might be beneficial for some patients.

Before any protocol can be established where patients are exposed to negative traumatic aspects, research needs to clarify fundamental empirical questions. An obvious question is whether, and to what extent, exposure to negative traumatic aspects is beneficial. With this regard, the LCT-PTSD

predicts that the benefits concern a particular subgroup of patients, namely those for whom standard exposure protocols fail (i.e., patients developing an interfering LC); this prediction needs to be tested empirically. If this prediction is confirmed, it will be valuable for research to identify which patients' characteristics predict success of protocols implementing exposure to negative traumatic aspects over and above standard protocols. Another important question is how distressing it is for patients to evoke negative traumatic aspects during sessions. Understanding this is crucial because it can inform the decision of whether to recommend exposure to negative traumatic aspects from the start or only after standard protocols fail (if distress is substantial, then exposure to negative traumatic aspects might be recommended only after failure of standard exposure protocols). Moreover, this aspect needs to be assessed in light of evidence that some patients dropout because they are distressed by exposure therapy;[3-7] recalling negative traumatic aspects might further increase the chance of dropouts.

In summary, the LCT-PTSD can inspire clinicians to improve methods for treating PTSD, especially contributing to refine exposure techniques. The LCT-PTSD suggests that, for some patients, exposure to mildly negative aspects of the trauma might be necessary to modify the traumatic LC and promote healing.

Discussion

Latent cause theory has recently offered tremendous insight about how memory and learning work.[8-15] By presenting the LCT-PTSD, here I explore whether the notion of LC can contribute to understanding PTSD. The proposal is that two key moments are critical in the development of the illness: the time of the trauma and the few months after. According to the LCT-PTSD, at both time points patients invoke a new LC, in the first case a traumatic LC and in the second case an interfering LC (the latter would be formed only by patients more resistant to exposure treatment). Activation of the former LC would account for intrusive symptoms and for the associated fear, while existence of the latter LC would interfere with any modification of traumatic memories during presentation of trauma-related cues, explaining why exposure fails to help some patients. The proposal has potential

clinical implications, raising the possibility that some patients might benefit from exposure to mildly painful aspects of the trauma in conjunction with trauma-related cues.

The LCT-PTSD makes several empirical predictions, such as that some PTSD patients invoke an interfering LC when exposed to trauma-related cues in the absence of danger, or that these patients can benefit from exposure to mildly painful aspects of the trauma combined with trauma-related cues. One key prediction of the LCT-PTSD is that PTSD patients are characterised by high spreading parameter. This idea fits with evidence that PTSD patients are predisposed towards dissociation, namely towards creating compartmentalised memory traces like in the case of the trauma (Nijenhuis & den Boer, 2009; Nijenhuis et al., 2010).[25,26] However, a recent study conducted by Norbury and colleagues[27] has failed to find an enhanced spreading parameter in PTSD patients. Still, a possible reason for this observation might be the following confound. A common approach to assess the spreading parameter (also adopted in Norbury and colleagues[27]) relies on classical conditioning paradigms, where one trial can be broken down in an initial phase and in an outcome phase.[10,12] During the initial phase, CS+ or CS- stimuli (the former predicting the US, the latter predicting absence of the US) are presented, whereas during the outcome phase a US may or may not be delivered. The assumption is that, at each phase, the brain infers the LC at play during the trial, and that the intensity (or probability) of the fear response (manifested during the initial phase) reflects this inference. Specifically, the intensity of the fear response is typically assumed to reflect beliefs about the *probability* of potential LCs.[10,12,27] For example, if one believes that a US-related LC is more likely than a safe LC, then fear will be more intense. However, it is reasonable to assume that the fear intensity depends not only on the probability of LCs, but also on their *value*. In other words, keeping the probability of the US-related LC constant, fear will increase as the value of the US becomes more negative (i.e., as punishment intensifies). When modelling fear responses to assess inference about LCs, including the role of value is fundamental. This role was not considered by Norbury and colleagues,[27] an aspect that might have affected substantially the analyses, especially given evidence that PTSD patients have enhanced sensitivity to punishment.[28] Research remains to be conducted

where the link between PTSD and the spreading parameter is investigated while controlling for the value of any US.

As another example of how value might confound inference about LCs, take fear generalization. Like in other anxiety disorders,[29,30] PTSD patients exhibit greater fear generalization (e.g., they express more fear when presented with CS-).[31-33] At first glance, this might hint towards a lower, rather than higher, spreading parameter in PTSD:[27] fear generalization might derive from PTSD patients attributing both CS+ and CS- to the same LC, rather than envisaging two separate LCs as control would do. However, an alternative explanation is that enhanced fear generalization arises because PTSD patients have enhanced sensitivity to punishment (i.e., because they attach a more negative value to the US). This fits with evidence showing that in the normal population fear generalization increases when the punishment intensity increases,[34] and with evidence of an increased punishment sensitivity in PTSD.[28] The case of fear generalization highlights how considering the role of US value is fundamental for an accurate assessment of the spreading parameter.

This point suggests that, although not explicitly considered by the LCT-PTSD, sensitivity to punishment might contribute to PTSD (note that, however, the LCT-PTSD is not incompatible with any role of punishment sensitivity). Evidence indicates that, when punishment is delivered, the reaction to this stimulus varies greatly across individuals and throughout the life of each individual.[35] This variability depends both on genetic factors and on life experience (also, people might have different sensitivity for different types of stimuli).[35] On this basis, higher punishment sensitivity might predispose towards PTSD (more generally, it might predispose towards any anxiety or depressive disorder):[35] in our example above, Tom might be protected from PTSD also because he is less sensitive to punishment. Another factor not considered by the LCT-PTSD (although, again, not incompatible either) might be linked with the prediction error experienced when a negative event occurs.[16] People exposed to similar events in the past (e.g., Tom) might be less surprised, and thus manifest less distress during the negative event. Finally, one third element contributing to PTSD might be an enhanced fear generalization, manifested in a tendency to activate the traumatic LC also in safe conditions.[31-33] Although considering punishment sensitivity, surprise, and fear generalization

might be important to fully understand PTSD, yet these factors fail to explain a crucial aspect for which the LCT-PTSD offers an explanation. This aspect is the observation that some patients benefit poorly from exposure to trauma-related cues presented in safe conditions.[3-7] If punishment sensitivity, surprise, and fear generalization were the only factors at play, then exposure to trauma-related cues without punishment should help virtually all patients; LCT-PTSD offers a possible explanation of why this is not the case.

By offering a formal description of how LCs are represented in the brain, the LC framework extends previous associative learning research about the role of context in classical conditioning.[36] LC theory not only systematises this previous research, but it also offers new insight. For instance, from previous research it remains unclear whether, once a contextual representation is formed, this can be modified with new learning or whether any new learning necessarily requires the formation of new contexts.[13,36] This point is particularly critical in PTSD, where it raises the question: can traumatic memories be modified, or can they only be inhibited? While the answer to this question remains unclear from previous literature, the LCT-PTSD theory offers a clear prediction: if the right conditions apply (i.e., if the traumatic LC is evoked), traumatic memories can indeed be modified.[13]

Some scholars have suggested that, in PTSD, trauma-related memories are qualitatively different from everyday memories. Traced back to Janet's seminal work on dissociation,[37,38] this idea has inspired, among other accounts, Dual Representation Theory.[39,40] This posits the coexistence of two separate memory systems: a Verbally-Accessible System where memories are encoded in verbal form, are easy to retrieve consciously, and are integrated with one another in a unified autobiographical narrative; and a Situationally-Accessible System where memories are encoded in sensory modalities (e.g., as visual images or auditory sounds), are hard to retrieve consciously but tend to emerge automatically, and are poorly integrated with one another. While both systems would underpin everyday memories, Dual Representation Theory argues that in PTSD traumatic memories are primarily under the control of the Situationally-Accessible System, resulting in the emergence of intrusive symptoms (consistent with the description of intrusive experiences as highly sensorial, automatic, and fragmented). In light of available evidence, whether traumatic memories in PTSD have a different nature compared to

everyday memories remain controversial.[41-44] The LCT-PTSD is agnostic about this debate, as its logic applies equally well independent of whether PTSD memories are of different kind or not. Certainly, the picture depicted by the LCT-PTSD is one where traumatic memories cluster around a traumatic LC which is segregated from other LCs, and thus where traumatic memories are poorly integrated with everyday memories. This captures the phenomenology of intrusive symptoms, described by patients as belonging to a "different" state of mind compared to everyday life. The separation between traumatic and everyday memories is proposed to encompass, above all, emotional aspects, being traumatic memories imbued with a level of fear (due to the extreme nature of the traumatic event) completely beyond the level characterising everyday memories. However, the segregation between traumatic and everyday memories implied by the LCT-PTSD does not entail, as Dual Representation Theory does, that these are different in kind (one verbal, the other sensory), nor that traumatic memories are harder to retrieve consciously.

An important aspect to consider is how the LCT-PTSD accounts for the multifaceted symptomatology of PTSD. Hierarchical models of psychopathology identify a non-specific component (often labelled as negative affectivity) underlying symptoms in multiple types of anxiety and depressive disorders.[45-47] Together with such general component, each specific disorder would be characterised by a particular set of symptoms.[45-47] This logic has been applied also to interpret PTSD symptoms.[48,49] Factor analytic investigations extract four factors (intrusions, avoidance, hyperarousal, and dysphoria) from PTSD symptoms, mapping three of these (avoidance, hyperarousal, and dysphoria) to a non-specific component (common to other anxiety and depressive disorders).[48,49] According to this perspective, intrusive symptoms would be those distinguishing PTSD from other disorders.[49] This picture fits with the LCT-PTSD, where (like in most models of the illness) the primary focus is on the origin of intrusive symptoms, deemed to be at the centre of PTSD. The LCT-PTSD is also consistent with mapping avoidance symptoms to a distinct factor (as in the example above comparing Jack and Marc, reporting low and high avoidance, respectively). Regarding a non-specific general distress component (captured by the dysphoria and hyperarousal factors), this can be interpreted by the LCT-PTSD as corresponding to a general sensitivity to punishment (i.e., as a tendency to be distressed by

negative events).[49] This sensitivity to punishment might predispose people to anxiety and depressive disorders of different kind, including PTSD. In the context of the LCT-PTSD, the predisposition towards PTSD might occur because higher punishment sensitivity might facilitate the interpretation of a negative event as extreme, thus favouring the formation of a traumatic LC (and the establishment of PTSD).

In conclusion, the contribution of the paper is twofold. First, it offers an interpretation of PTSD informed by latent cause theory, a framework based on cutting-edge discoveries in the animal literature. Second, it offers an instance of how latent cause theory can be applied to interpret mental illness, which (given the generality of the theory) might inspire applications of the theory to other forms of psychopathology.[50]

References

1. American Psychiatric Association. (2013). Diagnostic and statistical manual of mental disorders (5th ed.). Arlington, VA: Author.
2. Cahill SP, Foa EB. Psychological theories of PTSD. In Friedman MJ, Keane TM, Resick PA (Eds.), Handbook of PTSD: Science and practice (pp. 55–77). The Guilford Press, 2007.
3. Foa EB, McLean CP. The efficacy of exposure therapy for anxiety-related disorders and its underlying mechanisms: The case of OCD and PTSD. Ann Rev Clin Psychol, 2016; 12.
4. Gonçalves R, Pedrozo AL, Coutinho ESF., Figueira I, Ventura P. Efficacy of virtual reality exposure therapy in the treatment of PTSD: a systematic review. PloS one 2012; 7(12), e48469.
5. Peterson AL, Foa EB, Riggs DS. Prolonged exposure therapy. In Moore BA & Penk W (Eds.), Treating PTSD in military personnel: A clinical handbook (pp. 46–62). The Guilford Press, 2019.
6. Rauch SA, Eftekhari A, Ruzek JI. Review of exposure therapy: a gold standard for PTSD treatment. J Rehabil Res Dev 2012; 49(5), 679-688.
7. Wolpe, J. (1969). The practice of behavior therapy. Pergamon Press.
8. Collins AG, Frank MJ. Cognitive control over learning: creating, clustering, and generalizing task-set structure. Psychol Rev, 2013; 120(1), 190.

9. Courville AC, Daw ND, Touretzky DS. Bayesian theories of conditioning in a changing world. Trends Cogn Sci, 2006; 10(7), 294-300.
10. Gershman SJ, Hartley CA. Individual differences in learning predict the return of fear. Learn Behav, 2015; 43(3), 243-250.
11. Gershman SJ, Niv Y. Exploring a latent cause theory of classical conditioning. Learn Behav, 2012; 40(3), 255-268.
12. Gershman SJ, Blei DM, Niv Y. Context, learning, and extinction. Psychol Rev, 2010; 117(1), 197.
13. Gershman SJ, Jones CE, Norman KA, Monfils MH, Niv Y. Gradual extinction prevents the return of fear: implications for the discovery of state. Front Beha Neurosci, 2013; 7, 164.
14. Gershman SJ, Radulescu A, Norman KA, Niv Y. Statistical computations underlying the dynamics of memory updating. PLoS Comput Bio, 2014; 10(11), e1003939.
15. Gershman SJ, Norman KA, Niv Y. Discovering latent causes in reinforcement learning. Curr Opin Beha Sci, 2015; 5, 43-50.
16. Rescorla RA, Wagner AR. A theory of Pavlovian conditioning: variations in the effectiveness of reinforcement and nonreinforcement. In Black AH, Prokasy WF (Eds) Classical Conditioning II: Current Research and Theory. New York, NY: Appleton-Century-Crofts, 1972.
17. Bouton ME, Bolles RC. Role of conditioned contextual stimuli in reinstatement of extinguished fear. J Exp Psychol Anim Beha Process, 1979; 5(4), 368.
18. Rescorla RA. Spontaneous recovery. Learn Mem, 2004; 11(5), 501-509.
19. Rescorla RA, Heth CD. Reinstatement of fear to an extinguished conditioned stimulus. J Exp Psychol Anim Beha Process, 1975; 1(1), 88.
20. Gershman SJ, Blei DM. A tutorial on Bayesian nonparametric models. J Math Psychol, 2012; 56(1), 1-12.
21. Bremner JD, Southwick SM, Johnson DR, Yehuda R, Charney DS. Childhood physical abuse and combat-related posttraumatic stress disorder in Vietnam veterans. Am J Psychiatry, 1993.
22. Lanius RA, Vermetten E, Pain C. The impact of early life trauma on health and disease: The hidden epidemic. Cambridge, UK, 2010.
23. McLean CP, Foa EB. Prolonged exposure therapy for post-traumatic stress disorder: A review of evidence and dissemination. Expert Rev Neurother, 2011; 11(8), 1151-1163.
24. Hembree EA, Rauch SA, Foa EB. Beyond the manual: The insider's guide to prolonged exposure therapy for PTSD. Cogn Behav Pract, 2003; 10(1), 22-30.

25. Nijenhuis ERS., den Boer JA (2009). Psychobiology of traumatization and trauma-related structural dissociation of the personality. In Dell PF, O'Neil JA (Eds.), Dissociation and the dissociative disorders: DSM-V and beyond (pp. 337–365). Routledge/Taylor & Francis Group, 2009.
26. Nijenhuis E, van der Hart O, Steele K. Trauma-related structural dissociation of the personality. Act Nerv Super, 2010; 52(1), 1-23.
27. Norbury A, Brinkman H, Kowalchyk M, Monti E, Pietrzak RH, Schiller D, Feder A. Latent cause inference during extinction learning in trauma-exposed individuals with and without PTSD. Psychol Med, 2020; 1-12.
28. Dretsch MN, Thiel KJ, Athy JR, Born S, Prue-Owens K. Posttraumatic stress disorder in the US Warfighter: Sensitivity to punishment and antidepressant use contribute to decision-making performance. Traumatol, 2013; 19(2), 118-125.
29. Duits P, Cath DC, Lissek S, Hox JJ, Hamm AO, Engelhard IM, Baas JM. Updated meta-analysis of classical fear conditioning in the anxiety disorders. Depress Anxiety, 2015; 32(4), 239-253.
30. Marin MF, Hammoud MZ, Klumpp H, Simon NM, Milad MR. Multimodal categorical and dimensional approaches to understanding threat conditioning and its extinction in individuals with anxiety disorders. JAMA Psychiatry, 2020; 77(6), 618–627.
31. Kaczkurkin AN, Burton PC, Chazin SM, Manbeck AB, EspensenSturges T, Cooper SE., Lissek S. Neural substrates of overgeneralized conditioned fear in PTSD. Am J Psychiatry, 2016; 174(2), 125–134.
32. Lissek S, van Meurs B. Learning models of PTSD: Theoretical accounts and psychobiological evidence. Int J Psychophysiol, 2015; 98(3), 594-605.
33. Norbury A, Robbins TW, Seymour B. Value generalization in human avoidance learning. ELife, 2018; 7, e34779
34. Dunsmoor JE, Mitroff SR, LaBar KS. Generalization of conditioned fear along a dimension of increasing fear intensity. Learn Mem, 2009; 16(7), 460-469.
35. Gray, JA. The neuropsychology of emotion and personality. In Stahl SM, Iversen SD, Goodman EC (Eds.), Cogn Neurochem (pp. 171–190); 1987, Oxford University Press.
36. Bouton, ME. Context and behavioral processes in extinction. Learn Mem, 2004; 11(5), 485-494.
37. Janet P. The Mental State of Hystericals: A Study of Mental Stigmata and Mental Accidents. Putnam's Sons, 1901
38. Van der Hart O, Horst R. The dissociation theory of Pierre Janet. J Trauma Stress 1989; 2(4), 397-412.

39. Brewin CR, Dalgleish T, Joseph S. A dual representation theory of posttraumatic stress disorder. Psychol Rev, 1996; 103(4), 670.
40. Brewin CR, Gregory JD, Lipton M, Burgess N. Intrusive images in psychological disorders: characteristics, neural mechanisms, and treatment implications. Psychol Rev, 2010; 117(1), 210.
41. Zoellner LA, Bittenger JN. On the uniqueness of trauma memories in PTSD. Posttraumatic stress disorder: Issues and controversies, 2004.
42. Rubin DC, Deffler SA, Ogle CM, Dowell NM, Graesser AC, Beckham JC. Participant, rater, and computer measures of coherence in posttraumatic stress disorder. J Abnorm Psychol, 2016; 125(1), 11.
43. Berntsen D, Rubin DC. Involuntary memories and dissociative amnesia: Assessing key assumptions in posttraumatic stress disorder research. Clin Psychol Sci, 2014; 2(2), 174-186.
44. Lynn SJ, Maxwell R, Merckelbach H, Lilienfeld SO, van Heugten-van der Kloet D, Miskovic V. Dissociation and its disorders: Competing models, future directions, and a way forward. Clin Psychol Rev, 2019; 73, 101755.
45. Mineka S, Watson D, Clark LA. Comorbidity of anxiety and unipolar mood disorders. Ann Rev Psychol, 1998; 49(1), 377-412.
46. Brown TA, Chorpita BF, Barlow DH. Structural relationships among dimensions of the DSM-IV anxiety and mood disorders and dimensions of negative affect, positive affect, and autonomic arousal. J Abnorm Psychol, 1998; 107(2), 179.
47. Clark LA, Watson D. Tripartite model of anxiety and depression: psychometric evidence and taxonomic implications. J Abnorm Psychol, 1991; 100(3), 316.
48. Elklit A, Shevlin M. The structure of PTSD symptoms: A test of alternative models using confirmatory factor analysis. Br J Clin Psychol, 2007; 46(3), 299-313.
49. Simms LJ, Watson D, Doebbelling BN. Confirmatory factor analyses of posttraumatic stress symptoms in deployed and nondeployed veterans of the Gulf War. J Abn Psychol, 2002; 111(4), 637.
50. Erdmann T, Mathys C. A generative framework for the study of delusions. Schizophr Res, 2021.

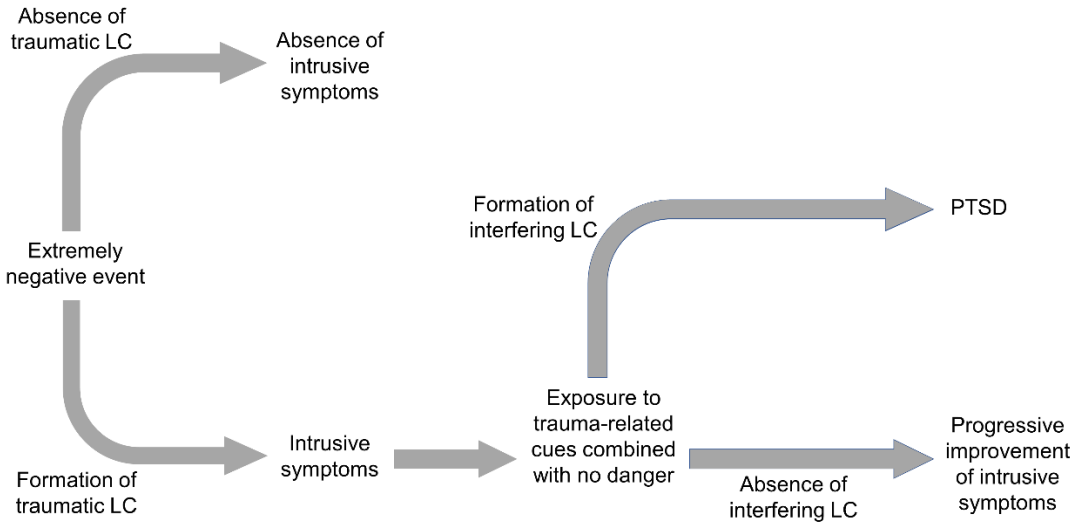

Fig. 1. Diagram describing the processes underlying the development of PTSD according to the LCT-PTSD. After an extremely negative event occurs, some people would link the event to old LCs, thus never developing intrusive symptoms. Conversely, other people would form a new traumatic LC, leading to the emergence of these symptoms. Exposure to trauma-related cues combined with absence of danger would follow. Some people would not link this to a new interfering LC, leading to a progressive improvement of the symptoms. Conversely, other people would interpret the same experience by forming a new interfering LC, which would prevent extinction to occur. By continuing to exhibit the symptoms for months after the negative event has occurred, the latter people would eventually receive a PTSD diagnosis.