Behaviour Research and Therapy 62 (2014) 37–46

Contents lists available at ScienceDirect

Behaviour Research and Therapy

journal homepage: www.elsevier.com/locate/brat

First steps in using machine learning on fMRI data to predict intrusive memories of traumatic film footage

Ian A. Clark [a, 1], Katherine E. Niehaus [b], Eugene P. Duff [c], Martina C. Di Simplicio [d], Gari D. Clifford [b], Stephen M. Smith [c], Clare E. Mackay [a], Mark W. Woolrich [e], Emily A. Holmes [d, f, *]

[a] University Department of Psychiatry, Warneford Hospital, University of Oxford, United Kingdom
[b] Institute of Biomedical Engineering, Department of Engineering Science, University of Oxford, United Kingdom
[c] FMRIB Centre, Nuffield Department of Clinical Neurosciences, John Radcliffe Hospital, University of Oxford, United Kingdom
[d] Medical Research Council Cognition and Brain Sciences Unit, 15 Chaucer Road, Cambridge CB2 7EF, United Kingdom
[e] Oxford Centre for Human Brain Activity (OHBA), Department of Psychiatry, Warneford Hospital, University of Oxford, United Kingdom
[f] Department of Clinical Neuroscience, Karolinska Institutet, Stockholm, Sweden

ARTICLE INFO

Article history:
Received 31 March 2014
Received in revised form
4 July 2014
Accepted 16 July 2014
Available online 4 August 2014

Keywords:
Intrusive memories
Trauma
Flashback
MVPA
Machine learning
Functional magnetic resonance imaging
Mental imagery

ABSTRACT

After psychological trauma, why do some only some parts of the traumatic event return as intrusive memories while others do not? Intrusive memories are key to cognitive behavioural treatment for post-traumatic stress disorder, and an aetiological understanding is warranted. We present here analyses using multivariate pattern analysis (MVPA) and a machine learning classifier to investigate whether peri-traumatic brain activation was able to *predict* later intrusive memories (i.e. before they had happened). To provide a methodological basis for understanding the context of the current results, we first show how functional magnetic resonance imaging (fMRI) *during* an experimental analogue of trauma (a trauma film) via a prospective event-related design was able to capture an individual's later intrusive memories. Results showed widespread increases in brain activation at encoding when viewing a scene in the scanner that would later return as an intrusive memory in the real world. These fMRI results were replicated in a second study. While traditional mass univariate regression analysis highlighted an association between brain processing and symptomatology, this is not the same as prediction. Using MVPA and a machine learning classifier, it was possible to predict later intrusive memories across participants with 68% accuracy, and within a participant with 97% accuracy; i.e. the classifier could identify out of multiple scenes those that would later return as an intrusive memory. We also report here brain networks key in intrusive memory prediction. MVPA opens the possibility of decoding brain activity to reconstruct idiosyncratic cognitive events with relevance to understanding and predicting mental health symptoms.

Introduction

The focus of the current paper is on using neuroimaging to understand the development of intrusive memories of trauma, that is "recurrent, involuntary and intrusive distressing memories of the traumatic event" (The Diagnostic and Statistical Manual of Mental Disorders, 5th ed.; DSM-5; American Psychiatric Association, 2013). Intrusive memories are a hallmark symptom from the re-experiencing cluster of Post-Traumatic Stress Disorder (PTSD). They have previously been defined as involuntary mental images that occur in a waking state (Frankel, 1994; Jones et al., 2003). Thus, key features of intrusive memories are that they are involuntary rather than deliberately retrieved, i.e. apparently spontaneous (Kvavilashvili, 2014); include perceptual aspects of the traumatic event, i.e. involve mental imagery rather than only verbal thought (Holmes, Grey, & Young, 2005); are in line with episodic and memory recall more broadly (Conway, 2001), and have distressing, i.e. emotional content (Hackmann, Ehlers, Speckens, & Clark, 2004). For example, after a motor vehicle accident, seeing scaffolding smashing through the car windscreen (see Grey & Holmes, 2008; Holmes et al., 2005 for further examples). In their most extreme form, re-experiencing symptoms can present as so-called dissociative 'flashbacks' where patients relive past events as if they are

* Corresponding author.
E-mail address: emily.holmes@mrc-cbu.cam.ac.uk (E.A. Holmes).
[1] Present address: Wellcome Trust Centre for Neuroimaging, Institute of Neurology, University College London, 12 Queen Square, London WC1N 3BG, United Kingdom.

http://dx.doi.org/10.1016/j.brat.2014.07.010

happening in the present (American Psychiatric Association, 2013). In contrast, during the experience of an intrusive memory the past events are spontaneously remembered while awareness of the present is maintained.

Due to the nature of this special issue, "*How neuroscience informs behavioural treatment*" within *Behaviour Research and Therapy*, we appreciate that many readers may not have a detailed understanding of neuroimaging terms and techniques. We therefore present a slightly longer than normal introduction to guide the reader through the steps taken before performing the main predictive analysis presented here. We first describe our initial study using traditional neuroimaging analysis techniques (Bourne, Mackay, & Holmes, 2013) and its subsequent replication (Clark, Holmes, Woolrich, & Mackay, submitted for publication). We then introduce the ideas of multivariate pattern analysis (MVPA) and machine learning, before next describing how we utilised these techniques in the current experiment. The aim of this is to provide a methodological basis for understanding the context of the current results and show that these findings are both replicable and reliable. We believe that by using neuroimaging techniques in addition to behavioural, cognitive and psychophysiological experiments we may be able to identify those neural and cognitive functions that are critical for intrusive memory formation. Understanding how intrusive memories are formed from multiple perspectives may allow future work to improve the ability to refine treatments which target the underlying mechanisms of intrusive memory (i.e. symptom) development. Indeed, by gaining the most comprehensive understanding of differences at an individual level, we may be able to open future possibilities of early screening for risk of PTSD, as well as the development of preventative approaches in the immediate aftermath of trauma and for targeted early interventions.

We also note that many different approaches to machine learning and MVPA are evolving, including (but not limited to) Random Forest Theory (Breiman, 2001), Graph theory (Power et al., 2011; Sporns, 2014) and Representational Similarity Analysis (Kriegeskorte, Mur, & Bandettini, 2008), in addition to that used here, a Support Vector Machine classifier (Pereira, Mitchell, & Botvinick, 2009). The current work represents only first steps in applying neuroimaging techniques to understand the neural impact of witnessing trauma and to inform behavioural treatment. We finish by exploring how such techniques might have implications for future cognitive behavioural therapy.

Intrusive memories and PTSD

Most people will experience a traumatic event during the course of their lifetime and a significant minority will go on to develop PTSD (Breslau et al., 1998; Kessler, Sonnega, Bromet, Hughes, & Nelson, 1995). We have successful treatments for the full blown disorder, those recommended by clinical guidelines (e.g. National Institute for Health and Clinical Excellence, 2005) are Cognitive Behavioural Therapy (CBT; e.g. Ehlers & Clark, 2000; Foa & Rothbaum, 1998) and Eye Movement Desensitisation and Reprocessing (EMDR; Shapiro, 1995). However, satisfactory preventative treatments against PTSD development are lacking (Roberts, Kitchiner, Kenardy, & Bisson, 2009). A greater understanding of the brain mechanisms that lead to the development of intrusive memories may help guide effective preventative interventions for the early aftermath of trauma.

We know little, in particular in terms of neuroscience, about why only certain events within a trauma return as intrusive memories when others do not. Processing at the time of trauma (peri-traumatic) is implicated in PTSD development (e.g. Brewin, 2014; Ehlers & Clark, 2000; Ozer, Best, Lipsey, & Weiss, 2003). Additionally, experimental findings implicate heightened emotional processing at the time of the event in intrusive memory development (Clark, Mackay, & Holmes, 2013, 2014). Interestingly, dissociation, defined within the DSM 5 as "a disruption of and/or discontinuity in the normal integration of consciousness, memory, identity, emotion …" (American Psychiatric Association, 2013, p. 291), can be a reaction to extreme emotion, and peri-traumatic dissociation has also been associated with intrusive memory formation (e.g. Daniels et al., 2012; Holmes, Brewin, & Hennessy, 2004). Seminal work on 'flashbulb' memories, defined as 'memories for the circumstances in which one first learned of a very surprising and consequential (or emotionally arousing) event' (Brown & Kulik, 1977) may also illuminate some of the mechanisms involved in intrusive memory formation. While flashbulb memories are a distinct phenomenon (and not exclusive to trauma, but part of autobiographical memory more generally), they may lie on a continuum with intrusive memories. Research suggests that memories that end up as flashbulb memories are psychophysiologically arousing, personally salient and unexpected and sudden (Brown & Kulik, 1977). Indeed, psychophysiology has been associated with intrusive memory development; at the time of viewing a specific film scene that is later recalled as an intrusive memory, heart rate has been shown to drop in comparison to the rest of film viewing (Chou, Marca, Steptoe, & Brewin, 2014; Holmes et al., 2004). Understanding the neural processes involved in intrusive memory formation adds another level of comprehension of this complex phenomenon.

Neuroimaging and established PTSD

The majority of studies using neuroimaging to investigate PTSD have done so once symptoms are already established in patients (Francati, Vermetten, & Bremner, 2007; Hughes & Shin, 2011; Pitman et al., 2012). Neurocircuitry models suggest that PTSD is characterised by reduced activity in the ventromedial prefrontal cortex, which is associated with decision making and emotional response inhibition, and increased activation in the amygdala and other limbic areas, which are associated with emotional processing (e.g. Rauch, Shin, & Phelps, 2006; Rauch, Shin, Whalen, & Pitman, 1998). A further recent model suggests that abnormalities in the amygdala and dorsal anterior cingulate cortex are pre-disposing, while abnormal interactions between the hippocampus and ventromedial prefrontal cortex arise after developing PTSD (Admon, Milad, & Hendler, 2013). While informative for understanding PTSD as a whole, these studies cannot tell us specifically about intrusive memories, that is, those events we need to target within a CBT treatment (e.g. Ehlers & Clark, 2000; Foa, Hembree, & Rothbaum, 2007). Further, studying symptoms once they are already established tells us little about the neural processes involved in intrusive memory *formation* (aetiology).

The trauma film paradigm: an experimental psychopathology approach

Electronic media offers a way to use neuroimaging to investigate the brain responses to experimental analogue trauma exposure and intrusive memory formation. Recent work has examined the effects of electronic media, for example television news film footage, on the development of PTSD symptoms. Individuals exposed for prolonged hours to media footage of terrorist attacks have been shown to present higher scores on stress and trauma related symptom scales both a month after the attack (Holman, Garfin, & Silver, 2014) and 2–3 years after the attack (Silver et al., 2013). Additionally, the DSM 5 (American Psychiatric Association, 2013) now includes exposure to trauma through electronic media in the definition of a traumatic event, with the caveat that the exposure is work related.

Together, this suggests that traumatic events transmitted through electronic media footage have the potential to induce PTSD-like symptomatology.

The trauma film paradigm is widely used as an experimental analogue of psychological trauma (see Holmes & Bourne, 2008; Lazarus, 1964) and involves healthy participants viewing traumatic footage in line with DSM 5 criteria for a traumatic event (e.g. real life footage depicting actual or threatened death and serious injury; American Psychiatric Association, 2013). The paradigm has been most commonly used in behavioural experiments. Examples include the investigation of cognitive tasks to reduce intrusive memory frequency (e.g. Tetris; Holmes, James, Coode-Bate, & Deeprose, 2009) or vulnerability factors for intrusive memory development (Laposa & Alden, 2008; Wessel, Overwijk, Verwoerd, & de Vrieze, 2008).

Recently, we conducted the first study, to our knowledge, to combine the trauma film paradigm with functional Magnetic Resonance Imaging (fMRI) (Bourne et al., 2013; $n = 22$). This provided a prospective measure of the brain activation at the moment of viewing a film scene that would later return as an intrusive memory during the following week. We then replicated this experiment, finding a near identical pattern of results (Clark et al., submitted for publication; $n = 35$). The importance of such replication studies has been particularly noted recently within the field of fMRI (e.g. Carp, 2013; Fletcher & Grafton, 2013).

In these studies, unlike most fMRI designs, we could not specify our neuroimaging 'events' of interest in advance (i.e. the specific time within stimuli presentation when brain activation is selected to be compared to the rest of stimuli presentation). This is due to intrusive memories being highly idiosyncratic; thus we did not know which scenes in the film would return involuntarily for each individual (just as after a real trauma we do not know which moments will be the hotspots and intrude). The film was created to include 20 scenes that had previously been found to induce intrusive memories. Participants recorded their intrusive memories (defined as mental images of the film content that involuntarily come to mind) for one week in daily life using a pen-and-paper diary. From written descriptions in the intrusive memory diary, intrusions were matched to specific scenes within the film (e.g. the car rolling over the hedge hitting the boy playing football in his garden). Film scenes were then classified on an individual participant basis as either 'Flashback scenes' – emotional scenes that returned as an intrusive memory for that individual, or 'Potential scenes' – emotional scenes that did not return as an intrusive memory for that individual, but did in other participants (see Fig. 1). On average, 3 of the possible 20 scenes became intrusive memories for each participant; a similar frequency to the number of different events experienced as intrusions after real life trauma (Grey & Holmes, 2008; Holmes et al., 2005).

Using a standard statistical mass univariate regression analysis approach (i.e. the analysis currently most used for fMRI data) we found that Flashback scenes, in comparison to Potential scenes, were characterised by widespread increases in brain activity including the anterior cingulate cortex, thalamus, putamen, insula, amygdala, ventral occipital cortex, left inferior frontal gyrus and bilateral middle temporal gyrus. In brief, brain regions that have previously been associated with emotional processing, visual/mental imagery and memory (see Bourne et al., 2013 for discussion). These results provided, to our knowledge, the first evidence of a 'neural signature' at the time of intrusive memory formation.

Predicting from fMRI; multivariate pattern analysis (MVPA) and machine learning

However, traditional univariate fMRI analysis only highlights an association of peri-traumatic brain responses with later intrusive memories *across a group* of individuals (see for details Jezzard, Matthews, & Smith, 2001; Smith et al., 2004). Additionally, traditional fMRI analysis relies on the self-report diary to identify the scene type. It would be useful to know the extent to which brain responses during exposure to analogue trauma can actually *predict* a specific moment of the traumatic footage that would later become an intrusive memory, for example, to inform preventative interventions against intrusive memory formation.

Machine learning and multivariate pattern analysis (MVPA) are neuroimaging analysis techniques that can be used to measure prediction accuracy. MVPA makes use of multivariate, spatially extensive patterns of activation across the brain. The patterns of activation across these larger regions can be "learned" through approaches from the field of machine learning. Supervised machine learning techniques optimise input "features" to best separate or describe the two labelled classes of data (i.e. Flashback scene or Potential scene). These "features" are simply summary measures of some aspects of the data. It is through these optimisation steps that machine learning approaches "learn" the patterns that best describe each class of data. Once the patterns have been identified, they can be used to predict the behaviour of new, previously unseen participants. Such approaches can provide greater discriminative ability than spatially localised mass-univariate regression analyses (see for further details, Haxby, 2012; Haynes & Rees, 2006;

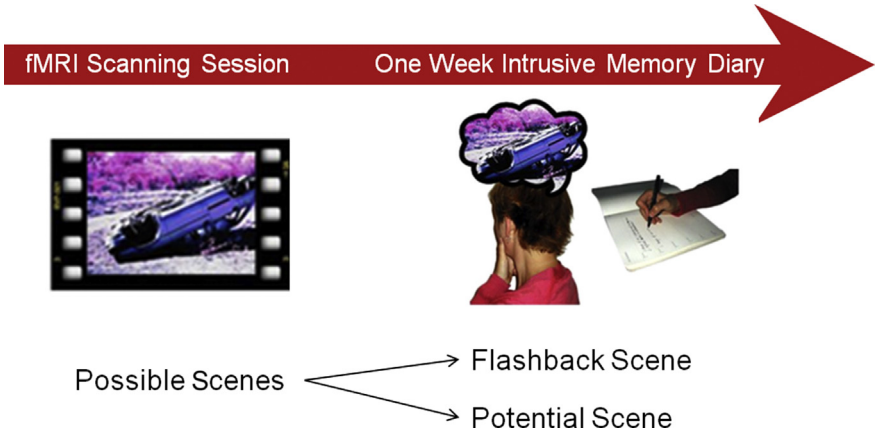

Fig. 1. Procedure diagram. Participants viewed traumatic footage while undergoing fMRI. Specific scenes in the film were determined to be 'Possible' scenes (scenes that had previously caused intrusive memories in other studies). As intrusive memories are idiosyncratic, Possible scenes became either 'Flashback' scenes or 'Potential' scenes for each individual. Scene type was determined for each participant retrospectively from the 1 week intrusive memory diaries.

McIntosh & Mišić, 2013; Mur, Bandettini, & Kriegeskorte, 2009; Norman, Polyn, Detre, & Haxby, 2006). Machine learning can then be used to learn these patterns of activity to accurately predict the occurrence of a new, unseen example of the same event (Lemm, Blankertz, Dickhaus, & Müller, 2011; Pereira et al., 2009).

To highlight just a few examples of MVPA techniques applied to fMRI, neural patterns identified by MVPA while participants were exposed to a shock during the presentation of picture stimuli have predicted the later behavioural expression of fear memory (pupil dilation response) between 2 and 6 weeks after encoding (Visser, Scholte, Beemsterboer, & Kindt, 2013). Additionally, MVPA techniques have identified patterns of activation at encoding that can predict later deliberate memory recall (see Rissman & Wagner, 2012).

We hypothesised that machine learning may be able to predict an intrusive memory from just the peri-traumatic brain activation. We aimed first, to investigate whether specific scenes in the film could be identified as later becoming intrusive memories solely from brain activation at the time of viewing traumatic footage by applying machine learning with MVPA. Second, we explore which brain networks are key in MVPA-based prediction of intrusive memory formation, and when the activation of these brain networks in relation to the timing of the intrusive memory scene is important.

Methods

Overview

To investigate whether differences in brain activation during the encoding of the trauma film stimuli could predict later intrusive memories of the film, we first trained a machine learning classifier (a support vector machine, SVM) to identify the specific brain activation pattern associated with viewing a film scene that was later involuntarily recalled as an intrusive memory. To do this, the classifier was provided with the timings of the intrusions (from scenes within the original film footage) from the diary data (i.e. from the intrusion content description once we knew which section(s) of the film became an intrusive memory for a given participant). We used a SVM since this approach has been shown to be reliable across multiple studies (see Mourão-Miranda, Bokde, Born, Hampel, & Stetter, 2005; Pereira et al., 2009). We tested a number of pre-processing and feature generation methods, as these choices have been shown to impact prediction accuracy more consistently than the choice of classifier (Duff et al., 2012; Ku, Gretton, Macke, & Logothetis, 2008).

Following training, the SVM was used to examine novel data – i.e. brain activation data from a new participant viewing traumatic footage – to pick out the scenes(s) in the footage which would be later experienced as an intrusive memory. Accuracy of prediction was evaluated by the classifiers predictions to those events reported in the participant's diary. Analysis was performed on our previously collected fMRI data (Bourne et al., 2013; Clark et al., submitted for publication).

What is key here is the prediction of which scenes in the film will later return as an intrusive memory in a *new* participant (something even the participant themselves cannot know at this point in time since they have not yet lived the week in which they will experience an intrusive memory). For details of the engineering aspects of the machine learning classifier development we refer the reader to Niehaus et al. (2014).

Participants

Participants were recruited from the local community separately for the two studies. Twenty-two participants took part in Bourne et al. (2013; mean age = 22 years, SD = 3.08; 17 female), and 35 in Clark et al. (submitted for publication; mean age = 22.43 years, SD = 7.58; 29 female). Inclusion criteria were: participants were aged over 18, had no metal implants, had not taken part in a similar study involving viewing traumatic footage and declared no previous or current psychiatric illness. In the Clark et al. (submitted for publication) study, data could not be analysed for additionally recruited participants where 0 intrusive memories were reported in the diary ($n = 2$) or insufficient performance on a visual recognition memory test ($n = 1$). For 3 further participants the full data was not acquired due to one participant stopping the scan during film viewing, one failing to return to follow up and for one technical issues stopped the scan before film completion. Recruitment material contained information about the potentially distressing content of the film material. Ethical approval was received from NHS Oxfordshire Research Ethics Committee 'B' (Bourne et al., 2013) and the University of Oxford Central University Research Ethics Committee (Clark et al., submitted for publication). All participants provided written informed consent and were reimbursed £25 (US $40).

Data acquisition

fMRI imaging data were acquired on a 3-T Siemens TIM Trio System with a 12-channel head coil [voxel resolution = $3 \times 3 \times 3$ mm^3; repetition time (TR) = 3 s, echo time (TE) = 30 ms]. T1-weighted structural images were acquired for subject registration using a magnetisation prepared rapid gradient echo (MPRAGE) sequence (voxel resolution = $1 \times 1 \times 1$ mm^3; TR = 2040 ms; TE = 4.7 ms]. Field maps were obtained for Clark et al. (submitted for publication) with .49 ms echo spacing and 22 ms TE.

Pre-processing

Data was pre-processed using FEAT (part of FSL – FMRIB's Software Library) version 6.0 (www.fmrib.ox.ac.uk/fsl). Brain extraction was performed using BET (Smith, 2002). High pass filtering was applied with a 100-s cut-off and spatial smoothing with a 5 mm full width half maximum Gaussian kernel. Motion correction was applied with MCFLIRT (Jenkinson, Bannister, Brady, & Smith, 2002). Field map based unwarping was applied to the data from Clark et al. (submitted for publication). Independent Component Analysis (ICA) was performed on all data using MELODIC. Components likely due to noise were removed by the FSL tool FIX. Images were registered to Montreal Neurological Institute (MNI) standard space.

The machine learning classifier

Classifier input features

The raw data from an fMRI study consists of activation levels for each voxel in the brain at every time-point during the study (here, images were captured every 3 s). In order to examine patterns across wider spatial regions, a group level Independent Component Analysis (ICA) was conducted. ICA is a statistical technique that separates the brain signals into independent spatial maps, clustering areas characterised by concurrent activation. This produces independent networks of brain regions that may be activated differentially during different tasks. The group ICA performed here is different to the ICA MELODIC analysis conducted during pre-processing as it identifies regions of concurrent activity across all participants rather than for individual participants (Beckmann & Smith, 2004). Following ICA decomposition, the spatial independent components (ICs) were projected back onto each participant to obtain participant-specific activation levels throughout the

spatial region of each IC. The number of ICs was varied to determine the optimal number for predicting flashbacks (detailed in Niehaus et al., 2014). These steps produced a set of activation time-courses for each IC for each participant.

In order to further summarise this data across time, the average level of activation was calculated for three different time periods for each scene type (i.e., for all Flashback and all Potential scenes): the first 6 s of each scene, the remaining duration of the scene, and the 12 s following the conclusion of the scene. In other words, this produced a set of (number of ICs)*(3) values, for each participant, which were used as input features into the machine learning classifiers.

Classifier optimisation

The support vector machine (SVM) classifier was first optimised on the larger of the 2 data sets (Clark et al., submitted for publication; 35 participants). A labelled sequence of Flashback and Potential scene time points in the film was created from the diaries for each individual participant (as each person may have different intrusions). The input features detailed above, reflecting activation across the brain, were extracted from the fMRI data during these Flashback and Potential time points (see Niehaus et al., 2014 for details). The SVM was then trained on this data to learn the patterns for both scene types, using a leave-one-out methodology to provide a test case: for 1 participant brain activation was not included in the training. Based upon the learned patterns of activity from all other participants, the classifier then attempted to identify the film scenes that later induced intrusive memories for the left-out participant. Identification based on brain activation patterns was the checked against the participant's diary entries (see Fig. 2). This leave-one-out 'cross-validation loop' was conducted 35 times, each one with a different participant left out of the training set. Results were averaged over the performance of the SVM on the left-out participant.

Various parameters were examined in order to optimise the predictive ability of the classifier. We compared both linear discriminant analysis and support vector machines as classifiers. Other supervised learning classifiers, such as random forests, could also have been employed, but here we limited our focus for this initial study. Due to the large number of Potential scenes in comparison to the number of Flashback scenes (approximately 5:1), we also compared various balancing techniques. Discussion of classifier optimisation is detailed in Niehaus et al. (2014).

As accuracy alone is not a good indicator of performance within imbalanced data sets (the classifier could achieve high accuracy by always classifying scenes as Potentials) we also assessed sensitivity. We define sensitivity here as the number of true Flashback scenes identified by the classifier out of the total number of Flashback scenes for that participant.

We then tested our ability to predict intrusive memories on our other data set (Bourne et al., 2013; 22 participants). Given our small number of participants, this step was important to test whether prediction performance would generalise to a separate data set.

Finally, we investigated the ability of machine learning to predict intrusive memory formation within a single participant. This within-participant analysis used only those participants within Clark et al. (submitted for publication) that experienced 4 or more different intrusive memories ($n = 16$; mean age 23 years, SD = 7.16;

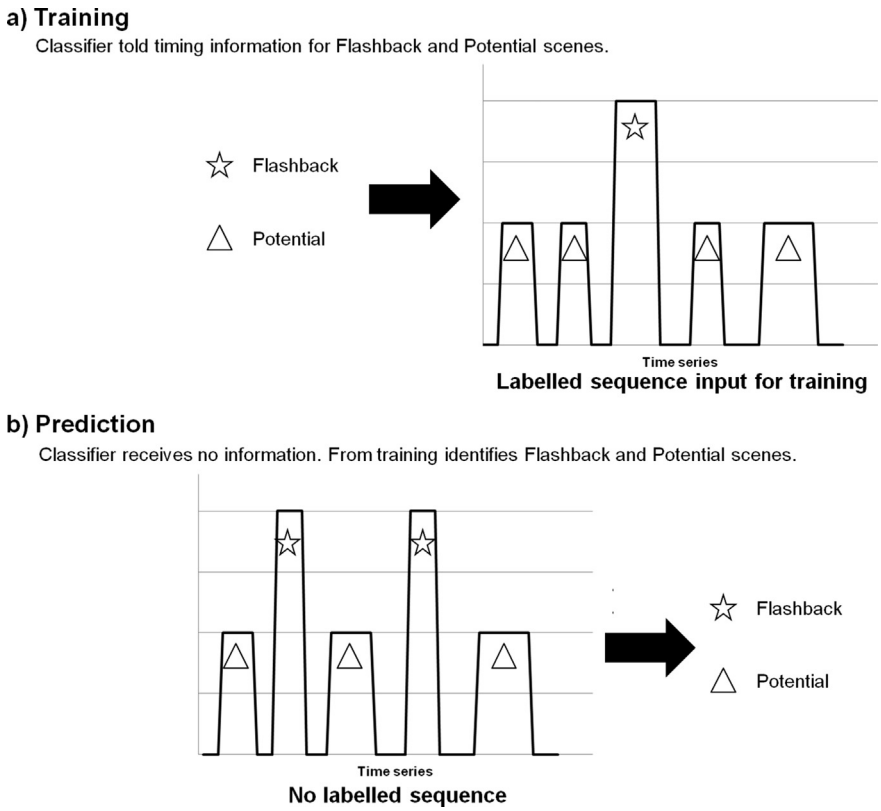

Fig. 2. Illustration of the prediction aspect of the machine learning analysis. a. Shows the training element of the machine learning approach. The classifier was provided with information concerning the timing of the Flashback scenes (emotional scenes that returned as a intrusive memory for that individual) and Potential scenes (emotional scenes that did not return as a intrusive memory for that individual, but did in other participants) from which to learn the patterns of brain activation for each scene type. Training was performed on all but 1 participant. b. Shows the predictive element of the machine learning approach. For the 1 participant not included in training the machine learning classifier goes through the brain activation data and attempts to identify the Flashback and Potential scenes.

13 female) leaving one Flashback scene and one Potential scene out for each participant. For within participant analysis, activation levels within individual voxels were used as input features. Voxels were selected with a *t*-test, and brain activity levels were averaged across the entire duration of each scene.

Identification of brain network functions

Possible functions of the networks identified in the input features (i.e. the ICA components at specific time points), and the names used to describe the cognitive functions of these networks were identified from Smith et al. (2009). Smith et al. (2009) utilised an online repository of published neuroimaging results containing around 30,000 participants from over 1600 published articles (the BrainMap database; Fox & Lancaster, 2002; Laird, Lancaster, & Fox, 2005) to map behavioural tasks (and their proposed corresponding cognitive functions) onto brain regions and networks.

Results

Prediction accuracy

In the original training data set the average accuracy of classification within each left-out participant (averaged across the training loops) was 70.1% (SE = 1.8%) with a sensitivity of 60.0% (SE = 5.9%). During replication in the second data set (Bourne et al., 2013); the classifier had a leave-one-out average performance accuracy of 68.0% (SE = 2.4%) and sensitivity of 58.7% (SE = 7.0%). Within a given participant the average accuracy was 97.3% (SE = .93%) and sensitivity of 90.3% (SE = 3.07%).

The best performance for predicting the scenes that would later become intrusive memories was found by using a linear discriminate analysis classifier with 39 independent components. It was found that predictive accuracy significantly decreased when the number of ICs was reduced to below 10 or increased to greater than 70. The best approach for managing the unbalanced class sizes was to apply an increased cost weighting for misclassifying Flashback scenes.

The best performance for predicting which scenes would become intrusive memories within participants was with a support vector machine classifier using 1000 voxels as input features.

Network identification

A total of 117 input features (i.e. averaged activation across the 39 ICA brain networks during the 3 defined time points of the scenes; the initial 6 s of the scene, the remaining duration of the scene after the initial 6 s, and the 12 s post scene) contributed to intrusive memory scene prediction. Below we describe the top weighted input features of the classifier for predicting Flashback versus Potential events (i.e. the features contributing most strongly towards prediction in terms of their weighting within the classifier). We also note their possible cognitive function. While these networks are those top weighted by the classifier, this is not a statistical measure and can only provide a guide towards their predictive contribution. There are 2 components of each feature; the location in the brain (i.e. the ICA component) and the timing of activation. The top weighted input features comprise 8 ICA components, 3 of which were important for intrusive memory prediction at 2 time points (see Fig. 3; ICA components (a–h) are displayed according to their weighting, activation time points are displayed in brackets).

The number of ICA brain networks included in the classifier was restricted so that maximum predictive ability was obtained (increasing from 39 to 70 independent components decreased sensitivity to 47.3%, SE = 7.27). This resulted in relatively widespread brain networks rather than specific brain areas, for which it is harder to attribute a specific function.

The highest weighted input feature (Fig. 3(a)), included the lingual gyrus, left hippocampus, middle temporal cortex, inferior frontal gyrus, supramarginal gyrus, left thalamus, precuneus, middle frontal cortex, left superior frontal cortex and posterior cingulate cortex. Networks within this feature (identified using Smith et al., 2009) have been previously associated with Cognition–Language–Semantics, Cognition–Language–Phonology and Cognition–Memory–Explicit. Activation of this input feature was important for prediction during the remaining duration of the scene (after the initial 6 s) and the 12 s post scene.

The next weighted feature (Fig. 3(b)) included the frontal orbital cortex, insula, frontal, central and parietal operculum, putamen, inferior frontal gyrus, anterior cingulate cortex, thalamus, supramarginal gyrus, middle frontal cortex, pre central cortex and the lateral occipital cortex. Networks within the feature have been associated with a number of functions termed 'Executive Control' in addition to Emotion, Perception–Somesthesis–Pain and Action–Inhibition (Fig. 3(b)). Activation of the feature was important for prediction during the initial 6 s of the scene.

The third weighted feature (Fig. 3(c)), involved the thalamus, insula, central and parietal operculum, putamen, inferior frontal gyrus and the anterior and posterior cingulate cortex. Networks in these areas have been associated with Emotion and Perception–Somesthesis–Pain. The feature was predictive in the 12 s post scene.

The fourth weighted feature (Fig. 3(d)) involved the lateral occipital cortex, occipital fusiform, amygdala, right putamen, right inferior frontal gyrus, right insula, right thalamus and occipital pole. Networks in the feature have been associated with Perception–Vision–Shape and Emotion. Activation levels were important for prediction during the remaining duration of the scene (after the initial 6 s) and the 12 s post scene.

The fifth feature (Fig. 3(e)) predominantly involved occipital fusiform gyrus, temporal occipital fusiform gyrus, lateral occipital cortex, occipital pole and intracalcarine cortex. This network has been associated with Perception–Vision–Shape. Activation of the feature was important for prediction in the 12 s post scene.

The sixth weighted feature (Fig. 3(f)) involved a wide range of regions including the parahippocampal gyrus, middle temporal cortex, right hippocampus, insula, thalamus, lingual gyrus, occipital pole, putamen, precuneus, frontal operculum, middle frontal cortex, left inferior frontal gyrus, angular gyrus, lateral occipital cortex, supramarginal gyrus and the anterior and posterior cingulate. Networks involved have been associated with Cognition–Language–Semantics, Cognition–Language–Phonology, Cognition–Memory–Explicit, Emotion, and the Default Mode Network. Activation of the feature was important for prediction during the remaining duration of the scene (after the initial 6 s).

The seventh weighted feature (Fig. 3(g)) involved the insula, left parahippocampus, left hippocampus left middle temporal cortex, planum polare (part of Wernicke's area), thalamus, posterior cingulate cortex and lateral occipital cortex. Networks have been associated with Emotion and Cognition–Memory–Explicit. Activation was important during the initial 6 s of the scene and the remaining duration of the scene (after the initial 6 s).

The final feature shown here (Fig. 3(h)) involved the lateral occipital cortex, amygdala, thalamus, accumbens, putamen, frontal operculum, inferior frontal gyrus, supramarginal gyrus, superior and middle frontal cortices, and the precuneus. Networks have been associated with Cognition–Language–Semantics, Cognition–Language–Phonology, Cognition–Memory–Explicit and Perception–Somesthesis–Pain. Activation of the feature was important for prediction in the 12 s post scene.

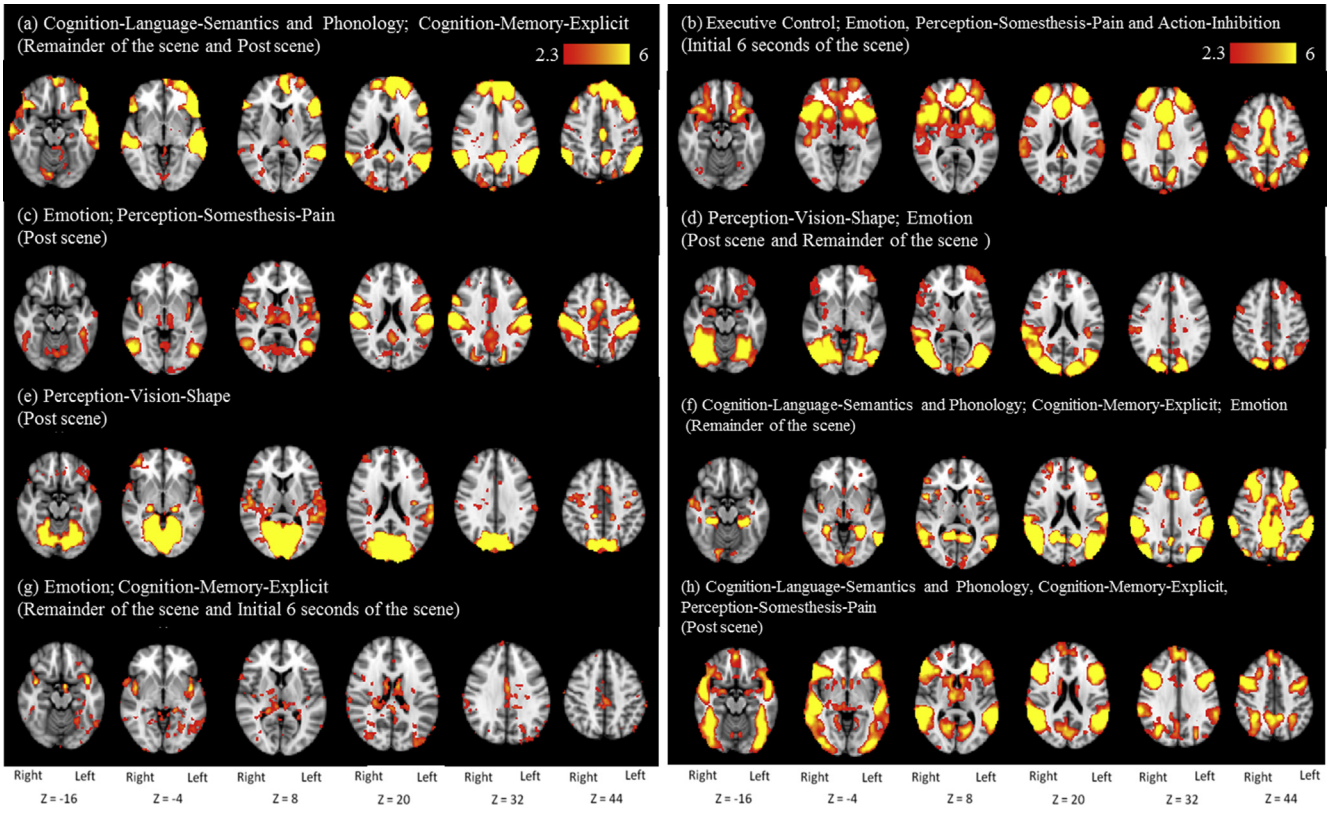

Fig. 3. The top weighted input features compromising 8 ICA components (a–h) and their corresponding time points (in brackets) involved in the prediction of a Flashback scene at the time of viewing traumatic footage. The ICA components are presented in the weighted order of the features used in the classifier. Features could be involved at 1 or all of 3 time points; i) the initial 6 s of the Flashback scene, ii) the remainder of the Flashback scene or iii) the 12 s post Flashback scene. Proposed functions of networks within the feature are included to provide a guide to their potential role in intrusive memory formation with names taken from Smith et al. (2009). 6 images are taken for each ICA component and are shown in the axial plane with their corresponding *z* coordinate. The underlying image is the Montreal Neurological Institute (MNI) 152 template, *z*-statistic images are thresholded at $z > 2.3$. *z*-Statistic range is represented by the change in colour.

Discussion

Intrusive memories are a target in CBT for Post-Traumatic Stress Disorder. This paper has presented an experimental psychopathology approach to understanding the underlying neural mechanisms of intrusive memories using recent advances in brain imaging analysis techniques. Here we show that intrusive memory formation (that is, which moments within an analogue trauma will be spontaneously recalled in the week after viewing the trauma) can be *predicted* solely from brain activation at the time of viewing the traumatic film footage. Intrusive memories are highly idiosyncratic; on average 3 of a possible 20 scenes within the trauma film returned as an intrusive memory, but which 3 varied according to each individual. The machine learning (Support Vector Machine) classifier, using MVPA for the input variables, was able to predict the later occurrence of a specific intrusive memory in an unseen participant from the data set with 70.1% accuracy and 60% sensitivity. This generalised to a novel data set of new participants with 68% accuracy and 58.7% sensitivity showing good replication. Further, we could predict intrusive memory development within a given participant with 97% accuracy and 90.3% sensitivity (i.e. if we know someone's brain reaction associated with intrusive memory development within the trauma trained on all their intrusions except one, we could then accurately predict a new example of intrusive memory formation – the missing intrusion).

These results provide support for a hypothesised 'intrusive memory signature' within brain activation at the time of the original analogue trauma encoding (Bourne et al., 2013; Clark et al., submitted for publication). Our results suggest that not only is brain activation at encoding *associated* with the occurrence of intrusive memories, but we can measure the accuracy with which brain activity can be used to *predict* specific scenes that will become intrusive memories. That is, a specific pattern of brain response during trauma exposure contributes to determining if a certain moment during the trauma will be later re-experienced as an intrusive memory or not. A related effect has previously been noted in the non-clinical memory literature, called the subsequent memory effect (Dobbins & Wagner, 2005; Paller & Wagner, 2002; Rissman & Wagner, 2012) albeit for non-intrusive types of memory.

Our data indicate a number of brain networks where analogue peri-traumatic activation appears crucial for intrusive memory prediction. The networks used by the machine learning classifier for intrusive memory prediction are in line with neurocircuitry models of PTSD patients (Admon et al., 2013; Rauch et al., 2006): hyper-responsivity in the amygdala and associated limbic regions involved in emotional processing and the dorsal anterior cingulate cortex have been found in PTSD samples. These regions are also active in the networks implicated in the current machine learning analysis. In particular, increased activation in emotional processing regions was involved in 5 of our 8 top weighted networks used to predict intrusive memory formation after analogue trauma. Findings are in line with fMRI results for pre-disposing factors for later clinical PTSD symptom development (see Admon et al., 2013).

Interestingly, both our univariate and multivariate analyses highlight the involvement of possible language related networks in intrusive memory formation. This is interesting clinically since

early Positron Emission Tomography (PET) studies on Vietnam veterans revealed decreased activation in Broca's area (Shin et al., 1997, 1999). As cognitive behavioural therapies are language based, further understanding of the involvement of language in intrusive sensory memory development may be relevant to optimising therapeutic interventions. Additionally, it may help us to experimentally explore why some early aftermath counselling interventions, such as critical incident stress debriefing, have been found to be harmful (Roberts et al., 2009; Rose, Bisson, Churchill, & Wessely, 2002).

Overall, our results suggest that we were able to so-called 'mind read' (Norman et al., 2006), or in more literal terms decode the brain activity during film viewing to identify which scenes of the film would later intrude. This new approach of using machine learning and MVPA strengthens our understanding of neural mechanisms underpinning intrusive memory formation with clinical relevance. At a general process level we can derive information from the specific brain networks predictive of intrusive memories, suggesting which cognitive functions may be most relevant for intrusive memory formation, and present possible mechanistic targets for preventative interventions. Additionally, differences at an individual level may open future possibilities of early screening for risk of PTSD development in the immediate aftermath of trauma for targeted early intervention. A trauma film paradigm with fMRI might even be developed for use prior to real trauma exposure for identifying those who may be more vulnerable to trauma generally (e.g. within army recruits or emergency personnel).

Future work applying machine learning and fMRI to clinical psychology more broadly

How else may we be able to use advanced neuroimaging techniques within clinical psychology? MVPA predictive techniques may be able to use neuroimaging data to predict (among others) likelihood of illness occurrence in at-risk groups. For example, in depression, meta-analysis of fMRI studies indicates abnormal activity across various brain regions (e.g. amygdala, dorsal anterior cingulate cortex, insula) in depressed patients compared to healthy controls in response to negative stimuli (Hamilton et al., 2012). Machine learning classifiers have been able to utilise these differences to predict whether participants are grouped as patients or healthy controls solely from differences in brain activity at the time of viewing sad faces (Fu et al., 2008). Extending this to at-risk groups may help target resources and treatments, and possibly in the future could even aid diagnosis. Above, for example, we have suggested how our line of enquiry could be developed to aid identification those of at risk for PTSD, e.g. in emergency personnel.

Cognitive Bias Modification (CBM) is a procedure which aims to retune dysfunctional attentional and emotional biases (e.g. Browning, Holmes, & Harmer, 2010; Mathews & MacLeod, 2005; Niles, Mesri, Burklund, Lieberman, & Craske, 2013; Waters, Pittaway, Mogg, Bradley, & Pine, 2013). However, we lack objective methods to test whether an individual has altered their cognitive bias. If machine learning were able to classify cognitive biases it may be possible for the therapist to objectively observe whether a patient is able to modulate and reduce a cognitive bias by observing alterations in the underpinning brain response. Future studies could readily apply work to this area given the ease of studying cognitive bias modification during fMRI (Browning, Holmes, Murphy, Goodwin, & Harmer, 2010).

Further work using MVPA and machine learning may be able to identify brain activity at an individual participant level. Understanding the presentation of symptoms at an individual level may help assess the effects of a treatment for that patient by performing neuroimaging before and after treatments (e.g. exposure based therapy; Foa et al., 2007). MVPA techniques could compare brain response to trauma related stimuli, hypothesising that successful treatment would be signalled by a change in brain activation patterns compared to pre-treatment in those specific networks that were predictive of intrusive memory formation (e.g. as in Kriegeskorte, 2009, 2011). This may also be applicable to fear extinction and return of fear; while initial fear extinction is relatively easy to induce, ensuring that the extinction remains permanent is more difficult (Vervliet, Craske, & Hermans, 2013). MVPA utilising the brain activations involved in extinction (e.g. recruitment of the ventromedial prefrontal cortex and hippocampus; Milad et al., 2007) may be able to suggest whether a fear memory has undergone permanent extinction.

Advanced neuroimaging techniques may provide an avenue to overcome the occasional limitations of subjective reports of symptomatology, such as in patients who are mute, or difficulties that some patients have with verbally describing their precise symptoms. For example, work outside of clinical psychology has demonstrated the potential of MVPA to identify a specific image seen by a participant undergoing fMRI (Kay, Naselaris, Prenger, & Gallant, 2008). After examining the brain activity associated with viewing neutral images (picture stills), of which the content was known to the computer model, the model was able to pick out, from a large set of new picture stimuli, which specific image was seen by the participant. More recently, this technique was extended to film stimuli, following the same procedure but using dynamic neutral movies (Nishimoto et al., 2011). Further, by comparing brain activity identified to specific visual content and the brain activity during sleep, it has been possible to describe the content of a participant's dream (Horikawa, Tamaki, Miyawaki, & Kamitani, 2013).

Methodological and conceptual limitations

Viewing traumatic film footage is not the same as experiencing an actual trauma and findings need to be extended to clinical samples. However, we note that intrusive memories experienced after traumatic events and intrusive memories in everyday life (such as those in our experimental procedure) can be considered on a continuum (Kvavilashvili, 2014). Additionally, PTSD symptoms have been reported following exposure to traumatic media footage (Holman et al., 2014; Silver et al., 2013). Changes to DSM 5 (American Psychiatric Association, 2013) now include exposure to traumatic content via electronic media (e.g. films) as sufficient for a diagnosis when the exposure is work related which suggests, at least at times, film footage can create real PTSD symptoms.

Additionally, we note that our study has other limitations. The number of participants was limited, reducing the extent it was possible to test different machine learning strategies. The unusual scarce-event study design meant that it was nevertheless crucial to test and optimise pre-processing and feature generation approaches on the first study participants, and then test the optimised approach on the independent sample.

Links between brain activations and cognitive function have been made here with what is termed in the fMRI literature as 'reverse inference', i.e. a brain region is identified as being predictive of a later event (e.g. an intrusive memory); in other studies that region was active when participants were performing a task engaging a particular cognitive process; it is therefore likely that this cognitive process is involved in intrusive memory formation (see Poldrack, 2006, 2011). However, the problem arises as to the specificity of the identified region in that specific cognitive process – it is unlikely that a brain region has just one cognitive function. On the other hand, we note that the predictive capabilities of the machine learning are not in question, only potential interpretations of the brain regions involved.

Finally, the DSM-5 distinguishes between intrusive memories and dissociative 'flashbacks'. Dissociation has been studied previously in behavioural experiments using the trauma film paradigm (e.g. Hagenaars & Krans, 2010; Hagenaars, van Minnen, Holmes, Brewin, & Hoogduin, 2008; Holmes et al., 2004). However, no measure of dissociation was taken in the current study and thus we could not examine any possible effects of dissociation to the current work. A continuum has been proposed ranging from involuntary autobiographical memories in everyday life to recurrent intrusive memories and in the most extreme (and rarest) form dissociative flashbacks (Kvavilashvili, 2014). Investigating dissociation in combination with fMRI is therefore an important step for future work (e.g. Daniels et al., 2012).

Conclusions

Using machine learning and MVPA on fMRI data of trauma film encoding, we have demonstrated that peri-traumatic brain activation is able to *predict* moments that would later return as an intrusive memory with 68% accuracy across participants and within a given participant with 97% accuracy. Here, we make an attempt to import ideas from basic neuroscience to contribute to an area of mental health – intrusive trauma memories. We suggest certain advance neuroimaging techniques may even be developed for use in studying relatively infrequently occurring and idiosyncratic events in mental health symptomatology (such as intrusive memories) and be used to predict individual's future symptom response.

Acknowledgements

Ian Clark is supported by a United Kingdom Medical Research Council Centenary Early Career Award. Katherine Niehaus is supported by the Rhodes Trust and the RCUK Digital Economy Programme [EP/G036861/1]. Mark Woolrich is supported by the Wellcome Trust; the MRC/EPSRC UK MEG Partnership award. Emily Holmes is supported by the United Kingdom Medical Research Council intramural programme [MC-A060-5PR50]; a Wellcome Trust Clinical Fellowship [WT088217]. Clare Mackay, Emily Holmes, Mark Woolrich are supported by the National Institute for Health Research (NIHR) Oxford Biomedical Research Programme. The views expressed are those of the author(s) and not necessarily those of the Rhodes Trust, RCUK, NHS, NIHR or the Department of Health. Funding to pay the Open Access publication charges for this article are provided by the United Kingdom Medical Research Council. None of the authors have any financial interest or benefit arising from the direct applications of their research.

References

Admon, R., Milad, M. R., & Hendler, T. (2013). A causal model of post-traumatic stress disorder: disentangling predisposed from acquired neural abnormalities. *Trends in Cognitive Sciences, 17*(7), 337–347. http://dx.doi.org/10.1016/j.tics.2013.05.005.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington D.C.: American Psychiatric Association.

Beckmann, C. F., & Smith, S. M. (2004). Probabilistic independent component analysis for functional magnetic resonance imaging. *IEEE Transactions on Medical Imaging, 23*(2), 137–152. http://dx.doi.org/10.1109/tmi.2003.822821.

Bourne, C., Mackay, C. E., & Holmes, E. A. (2013). The neural basis of flashback formation: the impact of viewing trauma. *Psychological Medicine, 43*(7), 1521–1533. http://dx.doi.org/10.1017/S0033291712002358.

Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5–32. http://dx.doi.org/10.1023/A:1010933404324.

Breslau, N., Kessler, R. C., Chilcoat, H. D., Schultz, L. R., Davis, G. C., & Andreski, A. (1998). Trauma and posttraumatic stress disorder in the community. *Archives of General Psychiatry, 55*(7), 626–632.

Brewin, C. R. (2014). Episodic memory, perceptual memory, and their interaction: foundations for a theory of posttraumatic stress disorder. *Psychological Bulletin, 140*(1), 69–97. http://dx.doi.org/10.1037/a0033722.

Brown, G., & Kulik, J. (1977). Flashbulb memories. *Cognition, 5*, 73–99.

Browning, M., Holmes, E. A., & Harmer, C. J. (2010). The modification of attentional bias to emotional information: a review of techniques, mechanisms and relevance to emotional disorders. *Cognitive, Affective and Behavioral Neuroscience, 10*(1), 8–20.

Browning, M., Holmes, E. A., Murphy, S. E., Goodwin, G. M., & Harmer, C. J. (2010). Lateral prefrontal cortex mediates the cognitive modification of attentional bias. *Biological Psychiatry, 67*(10), 919–925. http://dx.doi.org/10.1016/j.biopsych.2009.10.031.

Carp, J. (2013). Better living through transparency: improving the reproducibility of fMRI results through comprehensive methods reporting. *Cognitive, Affective, & Behavioral Neuroscience, 13*(3), 660–666. http://dx.doi.org/10.3758/s13415-013-0188-0.

Chou, C.-Y., Marca, R. L., Steptoe, A., & Brewin, C. R. (2014). Heart rate, startle response, and intrusive trauma memories. *Psychophysiology, 51*(3), 236–246. http://dx.doi.org/10.1111/psyp.12176.

Clark, I. A., Holmes, E. A., Woolrich, M. W., & Mackay, C. E. The neural basis of the encoding and involuntary recall of intrusive memories to traumatic footage (submitted for publication).

Clark, I. A., Mackay, C. E., & Holmes, E. A. (2013). Positive involuntary autobiographical memories: you first have to live them. *Consciousness and Cognition, 22*(2), 402–406. http://dx.doi.org/10.1016/j.concog.2013.01.008.

Clark, I. A., Mackay, C. E., & Holmes, E. A. (2014). Low emotional response to traumatic footage is associated with an absence of analogue flashbacks: An individual participant data meta-analysis of 16 trauma film paradigm experiments. *Cognition and Emotion*. Advance online publication http://dx.doi.org/10.1080/02699931.2014.926861.

Conway, M. A. (2001). Sensory-perceptual episodic memory and its context: autobiographical memory. *Philosophical Transactions of the Royal Society of London Series B – Biological Sciences, 356*(1413), 1375–1384.

Daniels, J. K., Coupland, N. J., Hegadoren, K. M., Rowe, B. H., Densmore, M., Neufeld, R. W., et al. (2012). Neural and behavioral correlates of peritraumatic dissociation in an acutely traumatized sample. *Journal of Clinical Psychiatry, 73*(4), 420–426. http://dx.doi.org/10.4088/JCP.10m06642.

Dobbins, I. G., & Wagner, A. D. (2005). Domain-general and domain-sensitive prefrontal mechanisms for recollecting events and detecting novelty. *Cerebral Cortex, 15*(11), 1768–1778. http://dx.doi.org/10.1093/cercor/bhi054.

Duff, E. P., Trachtenberg, A. J., Mackay, C. E., Howard, M. A., Wilson, F., Smith, S. M., et al. (2012). Task-driven ICA feature generation for accurate and interpretable prediction using fMRI. *NeuroImage, 60*(1), 189–203. http://dx.doi.org/10.1016/j.neuroimage.2011.12.053.

Ehlers, A., & Clark, D. M. (2000). A cognitive model of posttraumatic stress disorder. *Behaviour Research and Therapy, 38*(4), 319–345.

Fletcher, P. C., & Grafton, S. T. (2013). Repeat after me: replication in clinical neuroimaging is critical. *NeuroImage: Clinical, 2*(0), 247–248. http://dx.doi.org/10.1016/j.nicl.2013.01.007.

Foa, E. B., Hembree, E. A., & Rothbaum, B. O. (2007). *Prolonged exposure therapy for PTSD: Emotional processing of traumatic experiences*. NY: Oxford University Press.

Foa, E. B., & Rothbaum, B. (1998). *Treating the trauma of rape: Cognitive-behavioral therapy for PTSD*. New York: Guilford Press.

Fox, P., & Lancaster, J. (2002). Mapping context and content: the BrainMap model. *Nature Reviews Neuroscience, 3*(4), 319–321. http://dx.doi.org/10.1038/nrn789.

Francati, V., Vermetten, E., & Bremner, J. D. (2007). Functional neuroimaging studies in posttraumatic stress disorder: review of current methods and findings. *Depression and Anxiety, 24*(1), 202–218.

Frankel, F. H. (1994). The concept of flashbacks in historical perspective. *International Journal of Clinical and Experimental Hypnosis, 42*(4), 321–336. http://dx.doi.org/10.1080/00207149408409362.

Fu, C. H. Y., Mourao-Miranda, J., Costafreda, S. G., Khanna, A., Marquand, A. F., Williams, S. C. R., et al. (2008). Pattern classification of sad facial processing: toward the development of neurobiological markers in depression. *Biological Psychiatry, 63*(7), 656–662. http://dx.doi.org/10.1016/j.biopsych.2007.08.020.

Grey, N., & Holmes, E. A. (2008). "Hotspots" in trauma memories in the treatment of post-traumatic stress disorder: a replication. *Memory, 16*(7), 788–796.

Hackmann, A., Ehlers, A., Speckens, A., & Clark, D. M. (2004). Characteristics and content of intrusive memories in PTSD and their changes with treatment. *Journal of Traumatic Stress, 17*(3), 231–240.

Hagenaars, M. A., & Krans, J. (2010). Trait and state dissociation in the prediction of intrusive images. *International Journal of Cognitive Therapy*.

Hagenaars, M. A., van Minnen, A., Holmes, E. A., Brewin, C. R., & Hoogduin, K. A. L. (2008). The effect of hypnotically induced somatoform dissociation on the development of intrusions after an aversive film. *Cognition & Emotion, 22*(5), 944–963.

Hamilton, J. P., Etkin, A., Furman, D. J., Lemus, M. G., Johnson, R. F., & Gotlib, I. H. (2012). Functional neuroimaging of major depressive disorder: a meta-analysis and new integration of base line activation and neural response data. *The American Journal of Psychiatry, 169*(7), 693–703.

Haxby, J. V. (2012). Multivariate pattern analysis of fMRI: the early beginnings. *NeuroImage, 62*(2), 852–855. http://dx.doi.org/10.1016/j.neuroimage.2012.03.016.

Haynes, J.-D., & Rees, G. (2006). Decoding mental states from brain activity in humans. *Nature Reviews Neuroscience, 7*(7), 523–534. http://dx.doi.org/10.1038/nrn1931.

Holman, E. A., Garfin, D. R., & Silver, R. C. (2014). Media's role in broadcasting acute stress following the Boston Marathon bombings. *Proceedings of the National Academy of Sciences of the United States of America, 111*(1), 93–98. http://dx.doi.org/10.1073/pnas.1316265110.

Holmes, E. A., & Bourne, C. (2008). Inducing and modulating intrusive emotional memories: a review of the trauma film paradigm. *Acta Psychologica, 127*(3), 553–566.
Holmes, E. A., Brewin, C. R., & Hennessy, R. G. (2004). Trauma films, information processing, and intrusive memory development. *Journal of Experimental Psychology: General, 133*(1), 3–22.
Holmes, E. A., Grey, N., & Young, K. A. (2005). Intrusive images and "hotspots" of trauma memories in posttraumatic stress disorder: an exploratory investigation of emotions and cognitive themes. *Journal of Behavior Therapy and Experimental Psychiatry, 36*(1), 3–17.
Holmes, E. A., James, E. L., Coode-Bate, T., & Deeprose, C. (2009). Can playing the computer game 'Tetris' reduce the build-up of flashbacks for trauma? A proposal from cognitive science. *PLoS ONE, 4*(1), e4153. http://dx.doi.org/10.1371/journal.pone.0004153.
Horikawa, T., Tamaki, M., Miyawaki, Y., & Kamitani, Y. (2013). Neural decoding of visual imagery during sleep. *Science, 340*(6132), 639–642. http://dx.doi.org/10.1126/science.1234330.
Hughes, K. C., & Shin, L. M. (2011). Functional neuroimaging studies of post-traumatic stress disorder. *Expert Review of Neurotherapeutics, 11*, 275–285. http://dx.doi.org/10.1586/ern.10.198.
Jenkinson, M., Bannister, P., Brady, M., & Smith, S. (2002). Improved optimization for the robust and accurate linear registration and motion correction of brain images. *NeuroImage, 17*(2), 825–841. http://dx.doi.org/10.1006/nimg.2002.1132.
Jezzard, P., Matthews, P. M., & Smith, S. M. (Eds.). (2001). *Functional MRI: An introduction to methods*. Oxford: Oxford University Press.
Jones, E., Hodgins Vermaas, R., McCartney, H., Beech, C., Palmer, I., Hyams, K., et al. (2003). Flashbacks and post-traumatic stress disorder: the genesis of a 20th-century diagnosis. *The British Journal of Psychiatry, 182*, 158–163. http://dx.doi.org/10.1192/bjp.02.231.
Kay, K. N., Naselaris, T., Prenger, R. J., & Gallant, J. L. (2008). Identifying natural images from human brain activity. *Nature, 452*(7185), 352–355. http://www.nature.com/nature/journal/v452/n7185/suppinfo/nature06713_S1.html.
Kessler, R. C., Sonnega, A., Bromet, E., Hughes, M., & Nelson, C. B. (1995). Post-traumatic stress disorder in the national comorbidity survey. *Archives of General Psychiatry, 52*(12), 1048–1060.
Kriegeskorte, N. (2009). Relating population-code representations between man, monkey, and computational models. *Frontiers in Neuroscience, 3*. http://dx.doi.org/10.3389/neuro.01.035.2009.
Kriegeskorte, N. (2011). Pattern-information analysis: from stimulus decoding to computational-model testing. *NeuroImage, 56*(2), 411–421. http://dx.doi.org/10.1016/j.neuroimage.2011.01.061.
Kriegeskorte, N., Mur, M., & Bandettini, P. A. (2008). Representational similarity analysis – connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience, 2*. http://dx.doi.org/10.3389/neuro.06.004.2008.
Ku, S.-P., Gretton, A., Macke, J., & Logothetis, N. K. (2008). Comparison of pattern recognition methods in classifying high-resolution BOLD signals obtained at high magnetic field in monkeys. *Magnetic Resonance Imaging, 26*(7), 1007–1014. http://dx.doi.org/10.1016/j.mri.2008.02.016.
Kvavilashvili, L. (2014). Solving the mystery of intrusive flashbacks in posttraumatic stress disorder: comment on Brewin. *Psychological Bulletin, 140*(1), 98–104. http://dx.doi.org/10.1037/a0034677.
Laird, A., Lancaster, J., & Fox, P. (2005). BrainMap. *Neuroinformatics, 3*(1), 65–77. http://dx.doi.org/10.1385/ni:3:1:065.
Laposa, J. M., & Alden, L. E. (2008). The effects of pre-existing vulnerability factors on laboratory analogue trauma experience. *Journal of Behavior Therapy and Experimental Psychiatry, 39*(4), 424–435.
Lazarus, R. S. (1964). A laboratory approach to the dynamics of psychological stress. *American Psychologist, 19*(6), 400–411.
Lemm, S., Blankertz, B., Dickhaus, T., & Müller, K.-R. (2011). Introduction to machine learning for brain imaging. *NeuroImage, 56*(2), 387–399. http://dx.doi.org/10.1016/j.neuroimage.2010.11.004.
Mathews, A., & MacLeod, C. (2005). Cognitive vulnerability to emotional disorder. *Annual Review of Clinical Psychology, 1*, 167–195. http://dx.doi.org/10.1146/annurev.clinpsy.1.102803.143916.
McIntosh, A. R., & Mišić, B. (2013). Multivariate statistical analyses for neuroimaging data. *Annual Review of Psychology, 64*(1), 499–525. http://dx.doi.org/10.1146/annurev-psych-113011-143804.
Milad, M. R., Wright, C. I., Orr, S. P., Pitman, R. K., Quirk, G. J., & Rauch, S. L. (2007). Recall of fear extinction in humans activates the ventromedial prefrontal cortex and hippocampus in concert. *Biological Psychiatry, 62*(5), 446–454. http://dx.doi.org/10.1016/j.biopsych.2006.10.011.
Mourão-Miranda, J., Bokde, A. L. W., Born, C., Hampel, H., & Stetter, M. (2005). Classifying brain states and determining the discriminating activation patterns: support vector machine on functional MRI data. *NeuroImage, 28*(4), 980–995. http://dx.doi.org/10.1016/j.neuroimage.2005.06.070.
Mur, M., Bandettini, P. A., & Kriegeskorte, N. (2009). Revealing representational content with pattern-information fMRI—an introductory guide. *Social Cognitive and Affective Neuroscience, 4*(1), 101–109. http://dx.doi.org/10.1093/scan/nsn044.
National Institute for Health and Clinical Excellence. (2005). Post-traumatic stress disorder (PTSD): the management of PTSD in adults and children in primary and secondary care. From http://www.nice.org.uk/CG26.
Niehaus, K. E., Clark, I. A., Bourne, C., Mackay, C. E., Holmes, E. A., Smith, S. M., et al. (2014). *MVPA to enhance the study of rare cognitive events: An investigation of experimental PTSD*. Paper presented at the Pattern Recognition in Neuroimaging, Tubingen, Germany.
Niles, A. N., Mesri, B., Burklund, L. J., Lieberman, M. D., & Craske, M. G. (2013). Attentional bias and emotional reactivity as predictors and moderators of behavioral treatment for social phobia. *Behaviour Research and Therapy, 51*(10), 669–679. http://dx.doi.org/10.1016/j.brat.2013.06.005.
Nishimoto, S., Vu, An T., Naselaris, T., Benjamini, Y., Yu, B., & Gallant, J. L. (2011). Reconstructing visual experiences from brain activity evoked by natural movies. *Current Biology, 21*(19), 1641–1646. http://dx.doi.org/10.1016/j.cub.2011.08.031.
Norman, K. A., Polyn, S. M., Detre, G. J., & Haxby, J. V. (2006). Beyond mind-reading: multi-voxel pattern analysis of fMRI data. *Trends in Cognitive Sciences, 10*(9), 424–430. http://dx.doi.org/10.1016/j.tics.2006.07.005.
Ozer, E. J., Best, S. R., Lipsey, T. L., & Weiss, D. S. (2003). Predictors of posttraumatic stress disorder and symptoms in adults: a meta-analysis. *Psychological Bulletin, 129*(1), 52–73.
Paller, K. A., & Wagner, A. D. (2002). Observing the transformation of experience into memory. *Trends in Cognitive Science, 6*(2), 93–102.
Pereira, F., Mitchell, T., & Botvinick, M. (2009). Machine learning classifiers and fMRI: a tutorial overview. *NeuroImage, 45*(1 Suppl. 1), S199–S209. http://dx.doi.org/10.1016/j.neuroimage.2008.11.007.
Pitman, R. K., Rasmusson, A. M., Koenen, K. C., Shin, L. M., Orr, S. P., Gilbertson, M. W., et al. (2012). Biological studies of post-traumatic stress disorder. *Nature Reviews Neuroscience, 13*(11), 769–787.
Poldrack, R. A. (2006). Can cognitive processes be inferred from neuroimaging data? *Trends in Cognitive Sciences, 10*(2), 59–63. http://dx.doi.org/10.1016/j.tics.2005.12.004.
Poldrack, R. A. (2011). Inferring mental states from neuroimaging data: from reverse inference to large-scale decoding. *Neuron, 72*(5), 692–697. http://dx.doi.org/10.1016/j.neuron.2011.11.001.
Power, J. D., Cohen, A. L., Nelson, S. M., Wig, G. S., Barnes, K. A., Church, J. A., et al. (2011). Functional network organization of the human brain. *Neuron, 72*(4), 665–678. http://dx.doi.org/10.1016/j.neuron.2011.09.006.
Rauch, S. L., Shin, L. M., & Phelps, E. A. (2006). Neurocircuitry models of post-traumatic stress disorder and extinction: human neuroimaging research – past, present, and future. *Biological Psychiatry, 60*(4), 376–382.
Rauch, S. L., Shin, L. M., Whalen, P. J., & Pitman, R. K. (1998). Neuroimaging and the neuroanatomy of PTSD. *CNS Spectrums, 3*(Suppl. 2), 30–41.
Rissman, J., & Wagner, A. D. (2012). Distributed representations in memory: insights from functional brain imaging. *Annual Review of Psychology, 63*(1), 101–128. http://dx.doi.org/10.1146/annurev-psych-120710-100344.
Roberts, N. P., Kitchiner, N. J., Kenardy, J., & Bisson, J. I. (2009). Multiple session early psychological interventions for the prevention of post-traumatic stress disorder. *Cochrane Database of Systematic Reviews*, (3)http://dx.doi.org/10.1002/14651858.CD006869.pub2. http://www.mrw.interscience.wiley.com/cochrane/clsysrev/articles/CD006869/frame.html.
Rose, S. C., Bisson, J., Churchill, R., & Wessely. (2002). Psychological debriefing for preventing post traumatic stress disorder (PTSD). *Cochrane Database of Systematic Reviews*, (2)http://dx.doi.org/10.1002/14651858.CD000560. Art. no.: CD000560.
Shapiro, F. (1995). *Eye movement desensitization and reprocessing: Basic principles, protocols and procedures*. New York: Guilford Press.
Shin, L. M., Kosslyn, S. M., McNally, R. J., Alpert, N. M., Thompson, W. L., Rauch, S. L., et al. (1997). Visual imagery and perception in posttraumatic stress disorder. A positron emission tomographic investigation. *Archives of General Psychiatry, 54*(3), 233–241.
Shin, L. M., McNally, J., Kosslyn, S. M., Thompson, W. L., Rauch, S. L., Alpert, N. M., et al. (1999). Regional cerebral blood flow during script-driven imagery in childhood sexual abuse-related PTSD: a PET investigation. *American Journal of Psychiatry, 156*(4), 575–584.
Silver, R. C., Holman, E. A., Andersen, J. P., Poulin, M., McIntosh, D. N., & Gil-Rivas, V. (2013). Mental- and physical-health effects of acute exposure to media images of the September 11, 2001, attacks and the Iraq war. *Psychological Science, 24*(9), 1623–1634. http://dx.doi.org/10.1177/0956797612460406.
Smith, S. M. (2002). Fast robust automated brain extraction. *Human Brain Mapping, 17*(3), 143–155. http://dx.doi.org/10.1002/hbm.10062.
Smith, S. M., Fox, P., Miller, K., Glahn, D., Fox, M., Mackay, C., et al. (2009). Correspondence of the brain's functional architecture during activation and rest. *Proceedings of the National Academy of Sciences of the United States of America, 106*(31), 13040–13045. http://dx.doi.org/10.1073/pnas.0905267106. citeulike-article-id: 5282790.
Smith, S. M., Jenkinson, M., Woolrich, M. W., Beckmann, C. F., Behrens, T. E. J., Johansen-Berg, H., et al. (2004). Advances in functional and structural MR image analysis and implementation as FSL. *NeuroImage, 23*(Suppl. 1), S208–S219. http://dx.doi.org/10.1016/j.neuroimage.2004.07.051.
Sporns, O. (2014). Contributions and challenges for network models in cognitive neuroscience. *Nature Neuroscience, 17*(5), 652–660. http://dx.doi.org/10.1038/nn.3690.
Vervliet, B., Craske, M. G., & Hermans, D. (2013). Fear extinction and relapse: state of the art. *Annual Review of Clinical Psychology, 9*(1), 215–248. http://dx.doi.org/10.1146/annurev-clinpsy-050212-185542.
Visser, R. M., Scholte, H. S., Beemsterboer, T., & Kindt, M. (2013). Neural pattern similarity predicts long-term fear memory. *Nature Neuroscience, 16*(4), 388–390. http://dx.doi.org/10.1038/nn.3345.
Waters, A. M., Pittaway, M., Mogg, K., Bradley, B. P., & Pine, D. S. (2013). Attention training towards positive stimuli in clinically anxious children. *Developmental Cognitive Neuroscience, 4*(0), 77–84. http://dx.doi.org/10.1016/j.dcn.2012.09.004.
Wessel, I., Overwijk, S., Verwoerd, J., & de Vrieze, N. (2008). Pre-stressor cognitive control is related to intrusive cognition of a stressful film. *Behavioural Research and Therapy, 46*(4), 496–513.