# Neural mechanisms of harm avoidance learning – a model for obsessive-compulsive disorder?

Tobias U. Hauser PhD[1,2], Eran Eldar PhD[1,2] & Raymond J. Dolan FRS MD[1,2]

[1] Wellcome Trust Centre for Neuroimaging, University College London, London WC1N 3BG, United Kingdom
[2] Max Planck University College London Centre for Computational Psychiatry and Ageing Research, London WC1B 5EH, United Kingdom

**Corresponding author**
Tobias U. Hauser
Wellcome Trust Centre for Neuroimaging
University College London
12 Queen Square
London WC1N 3BG
United Kingdom
Phone: +44 / 20 3448 4409
Email: t.hauser@ucl.ac.uk

Tom is a middle-aged man who suffers with obsessive-compulsive disorder (OCD). Since early adulthood he has had a fear of being infected with the HI-Virus. In turn, he developed a variety of compulsive behaviours, such as excessive hand washing and cleaning rituals, in order to reduce what he perceived as risks for infection. Despite being aware of the irrationality of his behaviour, Tom's compulsions dominate his life. This type of scenario raises a question as to why some people with distressing thoughts develop excessive compulsions while others do not?

The traditional view of OCD, which sees the intrusive thoughts – for instance, fear of an HIV infection – as the primary cause of the disorder, has limited validity. There are far more people with intrusive thoughts about potential harm than there are OCD patients[1]. Furthermore, the most effective therapy for OCD, cognitive behavioural therapy, does not target the occurrence of intrusive thoughts, but focuses on the ensuing harm avoidance behaviours. In fact recent interpretations propose that compulsive behaviour is the actual trigger for the intrusions[2]. Thus, a tendency towards excessive harm avoidance behaviour is increasingly seen as a primary cause of OCD.

However, increased harm avoidance is not specific to OCD, and occurs in other anxiety disorders such as social anxiety or panic disorder[3]. We propose that a critical distinction relates to two types of harm avoidance, passive and active (Fig. A). We define passive avoidance as the cessation of a behaviour that might expose a person to potential harm. This is the type of avoidance commonly seen in a wide variety of anxiety disorders, and the paradigmatic case is agoraphobia, where patients passively avoid exposure to crowded places. We define active harm avoidance as behaviour where instrumental actions are enacted to prevent a potential occurrence of harm. In this framework compulsions constitute a form of active harm avoidance (e.g., excessive washing) where instrumental (physical or mental) behaviours are enacted to prevent a feared future state (e.g., HIV infection).

A number of studies have recently employed learning tasks to examine how OCD patients develop active harm avoidance behaviours. Gillan and colleagues[4,5] examined the emergence of active harm avoidance behaviour using a task where electric shocks, paired to a conditioned stimulus, could be prevented by executing an action. The authors found that patients with OCD continued to perform the learned avoidance behaviour even when it no longer resulted in any benefit, i.e. when the cables of the stimulator were detached. In a related study with OCD patients, Endrass et al.[6] compared learning of active harm avoidance and reward approach behaviour. After training, the OCD patients performed better than controls in active harm avoidance, but not in reward approach behaviour. These findings suggest that aspects of the OCD syndrome might be caused by a specific imbalance in avoidance learning, which engenders excessive and persistent active harm avoidance. However, why such imbalances emerge and what the underlying neurocognitive learning mechanisms are remains elusive.

A new study examining individual differences in passive and active harm avoidance learning provides novel insight into potential mechanisms, which we suggest has implications for understanding an excess of active avoidance behaviour in OCD. Eldar et al.[7] investigated how humans differ in their learning to avoid painful electric shocks. In a probabilistic gambling task (Fig. B), participants had to learn whether gambling with a deck of cards was likely to prevent a shock. This learning could be achieved by pursuing two possible learning strategies. One strategy involves learning from successful gambles how to actively avoid shocks. The alternative is to learn from unsuccessful gambles that resulted in shock, in other words learning when it is better to stay passive and not engage in gambling. Computational modelling highlighted a substantial heterogeneity across subjects' learning styles. About half learned more about what gambles they should take (active avoidance learners), whereas the other half learned more about what gambles they should not take (passive avoidance learners). These

individual differences were also reflected in neural responses in the striatum: shocks elicited increased activation in active avoidance learners, but an activation decrease in passive avoidance learners. Additionally, multivariate regression analysis showed that gray matter structure in the striatum significantly predicted whether someone belonged to an active or passive avoidance group, suggesting that harm avoidance learning style is a stable trait with a neuroanatomical basis in the striatum (Fig. C).

The striatum is of particular interest for OCD. Altered structure and function are often reported in this region[8], and this has led to it being a primary target for deep-brain stimulation (DBS) in OCD. This region also plays a critical role in the persistence of avoidance behaviours. Using functional MRI, Gillan et al.[5] found that those patients who persisted in their avoidance habit, showed increased activation in the striatum during exposure to conditioned stimuli, compared to patients that did not express the habit. These findings in humans are nicely complemented by a recent study investigating active avoidance in an OCD rodent model[9]. Similar to humans, Rodriguez-Romaguera et al.[9] found that a considerable section of their rats showed persisting avoidance habits after devaluation, similar to the findings by Gillan et al.[4,5]. Remarkably, these persisting avoidance habits ceased when these animals received DBS in the striatum. These converging findings thus suggest that the striatum is key to the formation and persistence of active avoidance behaviours.

These studies have theoretical implications for OCD research as they highlight how individuals differ in the way they deal with potential aversive outcomes. Intrusive thoughts, such as fear of an HIV infection, might result in different disorders conditional on the harm avoidance learning strategy a person is disposed to adopt. If one employs passive avoidance learning, a source of fear could result, for instance, in a specific phobia that manifests in avoiding needles and related stimuli. However, if a person, perhaps someone like Tom, employs active avoidance learning, this could result in active compulsion-like harm avoidance

behaviours, which might evolve into full-blown OCD. Moreover, a considerable heterogeneity within OCD patients[5] could also be indicative of different avoidance learning subgroups, which may be reflected by specific symptom patterns[10]. By assessing the relationship between active and passive avoidance learning in OCD patients in future studies, we might improve our understanding of OCD and its neurobiological foundations.

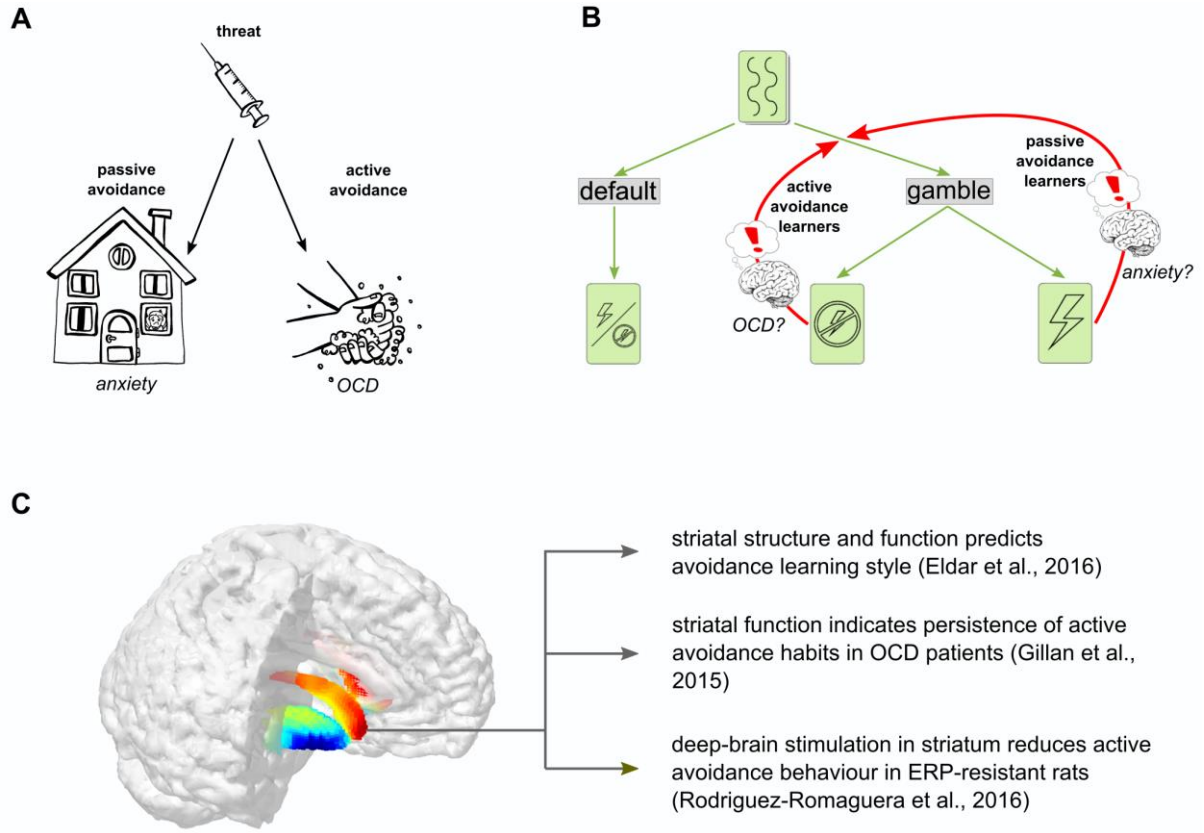


**Figure**. Neurocognitive mechanisms in harm avoidance learning. (A) A potential threat, such as being infected by a transmittable disease, can be prevented by pursuing different strategies. A passive avoidance strategy, often seen in anxiety disorders, is characterised by the cessation of a behaviour to prevent exposure. In contrast, active avoidance behaviour involves an instrumental behaviour to prevent harm, often observed in OCD patients. (B) A recent study[7] reveals how these distinct avoidance strategies emerge as a result of different learning styles. In a gambling task, participants had to avoid painful electric shocks. They could either bet on a deck of cards ('gamble') or default, in which case they received an electric shock in 50% of the trials. Subjects differed in how they learned from gambles: half of the participants – active avoidance learners – learned from successfully avoided shocks, i.e. when to act to prevent harm. The other half – passive avoidance learners – learned from trials where they received shocks, i.e. when to remain passive. These learning styles led to a bias with active avoidance gamblers having a higher propensity to gamble. (C) Individual avoidance learning styles were also reflected in the striatal response to shocks, and the gray matter (GM) structure in this area (red colours: more GM in active avoiders, blue colours: more GM in passive avoiders). Likewise, striatal activations were indicative of a persistence of active avoidance habits in OCD patients[5]. This is in line with rodent findings showing that stimulation of the striatum reduced active avoidance habits in rodents that persisted in these habits after an extinction with response prevention (ERP) procedure[9].

This key role of the striatum in active avoidance learning and the persistence of active avoidance habits may also explain why deep-brain stimulation can help treating patients with OCD.

**Acknowledgements**

All authors were supported by the Wellcome Trust Cambridge-UCL Mental Health and Neurosciences Network grant (095844/Z/11/Z). RJD holds a Wellcome Trust Senior Investigator Award (098362/Z/12/Z). The UCL-Max Planck Centre is a joint initiative supported by UCL and the Max Planck Society. The Wellcome Trust Centre for Neuroimaging is supported by core funding from the Wellcome Trust (091593/Z/10/Z). We thank Francesca Hauser-Piatti, MSc for creating some of the figures. All authors declare no conflict of interest. The funders had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

1. Julien D, O'Connor KP, Aardema F. Intrusive thoughts, obsessions, and appraisals in obsessive-compulsive disorder: a critical review. *Clin Psychol Rev*. 2007;27(3):366-383. doi:10.1016/j.cpr.2006.12.004.

2. Robbins TW, Gillan CM, Smith DG, de Wit S, Ersche KD. Neurocognitive endophenotypes of impulsivity and compulsivity: towards dimensional psychiatry. *Trends Cogn Sci*. 2012;16(1):81-91. doi:10.1016/j.tics.2011.11.009.

3. Kampman O, Viikki M, Järventausta K, Leinonen E. Meta-analysis of anxiety disorders and temperament. *Neuropsychobiology*. 2014;69(3):175-186. doi:10.1159/000360738.

4. Gillan CM, Morein-Zamir S, Urcelay GP, et al. Enhanced Avoidance Habits in Obsessive-Compulsive Disorder. *Biol Psychiatry*. March 2013. doi:10.1016/j.biopsych.2013.02.002.

5. Gillan CM, Apergis-Schoute AM, Morein-Zamir S, et al. Functional neuroimaging of avoidance habits in obsessive-compulsive disorder. *Am J Psychiatry*. 2015;172(3):284-293. doi:10.1176/appi.ajp.2014.14040525.

6. Endrass T, Kloft L, Kaufmann C, Kathmann N. Approach and avoidance learning in obsessive–compulsive disorder. *Depress Anxiety*. 2011;28(2):166-172. doi:10.1002/da.20772.

7. Eldar E, Hauser TU, Dayan P, Dolan RJ. Striatal structure and function predict individual biases in learning to avoid pain. *Proc Natl Acad Sci U S A*. 2016;113(17):4812-4817. doi:10.1073/pnas.1519829113.

8. Burguière E, Monteiro P, Mallet L, Feng G, Graybiel AM. Striatal circuits, habits, and implications for obsessive-compulsive disorder. *Curr Opin Neurobiol*. 2015;30:59-65. doi:10.1016/j.conb.2014.08.008.

9. Rodriguez-Romaguera J, Greenberg BD, Rasmussen SA, Quirk GJ. An Avoidance-Based Rodent Model of Exposure With Response Prevention Therapy for Obsessive-Compulsive Disorder. *Biol Psychiatry*. 2016. doi:10.1016/j.biopsych.2016.02.012.

10. McGuire JF, Storch EA, Lewin AB, Price LH, Rasmussen SA, Goodman WK. The role of avoidance in the phenomenology of obsessive-compulsive disorder. *Compr Psychiatry*. 2012;53(2):187-194. doi:10.1016/j.comppsych.2011.03.002.