REVIEWS

Surviving threats: neural circuit and computational implications of a new taxonomy of defensive behaviour

Joseph LeDoux[1,2,3]* *and Nathaniel D. Daw*[4]

Abstract | Research on defensive behaviour in mammals has in recent years focused on elicited reactions; however, organisms also make active choices when responding to danger. We propose a hierarchical taxonomy of defensive behaviour on the basis of known psychological processes. Included are three categories of reactions (reflexes, fixed reactions and habits) and three categories of goal-directed actions (direct action–outcome behaviours and actions based on implicit or explicit forecasting of outcomes). We then use this taxonomy to guide a summary of findings regarding the underlying neural circuits.

As soon as there is life, there is danger.
Ralph Waldo Emerson[1]

As the eminent comparative psychologist T. C. Schneirla noted, behaviour is a decisive factor in natural selection[2]: life is a dangerous undertaking, and those organisms that are adept at surviving live to pass their genes on to their offspring. Predators are pervasive sources of harm to animals, and most predators are themselves prey to other animals. As a result, nervous systems are typically equipped with predatory defence systems. This is true of invertebrate and vertebrate species, and within mammals, defence circuits are highly conserved[3–5]. Although humans are only minimally affected by predatory attacks from other animals, our predatory defence systems have been co-opted to cope with social threats arising from members of our own species[6,7]. Thus, understanding how the human brain responds to threats is important for both well-being and mental health because psychological disorders centred on pathological threat processing are common[8,9]. Because of the limitations of studies of the human brain, investigations of conserved defensive networks in other mammals have provided a viable approach for acquiring information relevant to human defensive circuitry. However, our ability to understand the neural circuits underlying any class of behaviour is only as good as our understanding of the behaviour itself.

Organisms can respond to danger in a number of ways. In recent years, it has become apparent that similar behaviours can arise from distinct psychological processes that depend on different neural circuits[10] and embody distinct computational approaches to the problem of controlling action[11,12]. Although many of these distinctions have been studied mainly in appetitive behaviour, it is similarly necessary to go beyond superficial similarities and differences in order to understand the psychological processes, computations and neural mechanisms underlying defensive responses. In light of this, we here propose a hierarchical taxonomy of defensive behaviours on the basis of their known psychological processes. We use this framework to organize a review of the neural circuit and, where possible, the computational basis of specific behavioural examples of each of the various response modes in the defensive hierarchy and to identify gaps and hypotheses for future work.

A defensive taxonomy

Our taxonomy partly overlaps with and extends the tripartite division between Pavlovian, habitual and goal-directed responses proposed in the context of appetitive behaviour by Dickinson and Balleine[10,13,14]. In their scheme, Pavlovian responses are defined as innate behaviours that come under the control of novel stimuli through associative learning and arbitrary learnedinstrumental responses (such as lever-pressing) are divided into habits and goal-directed actions. Whereas goal-directed actions depend on their association with outcomes, habits are typically instrumental responses that have lost their relationship to the outcome over time. These psychological categories have also been linked to distinct approaches to the computational problem of evaluating and selecting favourable actions[12]: specifically, goal-directed actions are thought to correspond to ‘model-based’ algorithms for evaluating actions on the basis of their outcomes, whereas habits are ‘model-free’ without such computations. Although this taxonomy has been both influential and useful, we recommend

Innate behaviours
Behaviours, such as reflexes and fixed responses, that all members of a species share as part of their heritage and that make minimal demands on learning.

[1]*Center for Neural Science and Department of Psychology, New York University, New York, NY, USA*
[2]*Department of Psychiatry and Department of Child and Adolescent Psychiatry, New York University Langone Medical School, New York, NY, USA.*
[3]*Nathan Kline Institute for Psychiatry Research, Orangeburg, NY, USA.*
[4]*Princeton Neuroscience Institute and Department of Psychology, Princeton University, Princeton, NJ, USA.*
*e-mail: ledoux@cns.nyu.edu

doi:10.1038/nrn.2018.22
Published online 29 Mar 2018

Instrumental responses
Responses that are learned because of their relationship with some consequent outcome (such as safety or food) and include both actions and habits.

Habits
Learned behaviours that are acquired as a result of their initial relation to an outcome but do not depend on the value of the outcome.

Actions
Behaviours that result in an expected outcome as a consequence of a previously learned contingency between the behaviour and its outcome (action–outcome behaviours) or of deliberative cognitive forecasting (implicit or explicit) of a possible outcome.

Reactions
Behaviours — such as reflexes, fixed responses and habits — that are directly elicited by innate or learned stimuli.

Startle
A flinch-like behavioural reflex often studied in the laboratory by using sudden, loud acoustic stimuli.

that its focus be modified and extended to broaden the scope of responses covered in both the appetitive and aversive domains.

Rather than focusing on learned associations, our taxonomy treats unlearned or innate responses as foundational. In particular, we distinguish between two categories of innate reactions: reflexes and fixed reaction patterns. These responses are automatically elicited by external stimuli. Both of these types of reaction can also come under the control of Pavlovian stimuli; thus, together they subsume Dickinson and Balleine's Pavlovian category. We also adopt and extend their instrumental categories: alongside habits, we thus subdivide goal-directed behaviours into three categories of actions (behaviours that result in some expected outcome) according to how their outcome is forecast. Specifically, in addition to those actions that use learned, direct action–outcome associations, we also consider those in which goals are forecast indirectly through deliberative cognitive processes. Computationally, this corresponds to the recognition that there are multiple types of model-based controllers, using different types of model. We include both implicit (nonconscious) and explicit (conscious) forecasts, which helps to account for dissociations in neural mechanisms and provides a basis for considering the role of states such as fear in these behaviours. The result is a hierarchical taxonomy consisting of six categories of behaviours: innate species-typical reflexes and fixed reactions, learned instrumental responses and actions based on nonconscious or conscious deliberation (TABLE 1). Importantly, although we use this taxonomy to discuss defensive behaviours, it may apply equally well to the appetitive domain. The defensive behaviours in these six categories occur to differing degrees in different organisms in the animal kingdom; however, we here focus on these responses in mammals (some of which exhibit all six kinds of behaviour, especially humans).

Innate responses

Defensive reflexes. If you step on a sharp object barefooted, your leg reflexively withdraws. If an insect flies close to your eye, you reflexively blink to protect the underlying tissues. Reflexes such as these are more or less hardwired stimulus–response connections that are automatically and rapidly triggered by an innately programmed stimulus (known as an unconditioned stimulus (US)). They are part of an organism's species heritage and are common to all members of the species. Reflexes are typically graded (that is, the more intense the stimulus, the stronger the response (up to a limit)) and usually involve one muscle or a limited set of muscles. From a computational perspective, reflexes (and the fixed reaction patterns described below) constitute largely preprogrammed responses to different types of event[11]. They are adapted to the organism's ethological niche and, within the organism, to the particular context[5,15]. Reflexes occur throughout invertebrate and vertebrate phyla as part of their innate survival repertoire.

A defensive behavioural reflex that has been studied extensively in mammals is startle[16,17]. This flinch-like response involves muscles up and down the body. The neck and back of mammals are outside of their visual field, and startle reflexes protect the organism in the case of a predatory attack to these areas by shortening and stiffening muscles, thus reducing exposure and making penetration more difficult. The reflex can be elicited by touch, acoustic stimuli or vestibular stimuli, and combinations of these produce more robust responses. In laboratory settings, startle is often studied by using sudden, loud acoustic stimuli.

Defensive reflexes such as startle, although innate, can be modulated by learning. For example, an innocuous stimulus (such as a light that comes on and goes off) normally has minimal effects on startle reflexes. However, after being paired with an aversive US through

Table 1 | **A defensive taxonomy**

Reactions and behaviours	Elicited or emitted	Instrumental	Goal-directed	Basis of goal-directedness	Implicit or explicit	Example
Species-typical reactions						
Reflexes	Elicited	No	No	N/A	Implicit	Startle
Fixed reaction patterns	Elicited	No	No	N/A	Implicit	Freezing
Instrumental behaviours						
Habits	Elicited	Yes	No	N/A	Implicit	Avoidance responses that persist despite lack of evidence that harm will come if they are not performed
Action–outcome behaviours	Emitted	Yes	Yes	Action–outcome contingency	Implicit	Avoidance responses based on a history of harm
Deliberative actions, implicit	Emitted	Yes	Yes	Nonconscious deliberation	Implicit	Avoidance of possible harm by implicitly anticipating a potentially dangerous event
Deliberative actions, explicit	Emitted	Yes	Yes	Conscious deliberation	Explicit	A conscious feeling of fear that motivates a plan to mitigate or escape from present harm or to avoid future harm

N/A, not applicable.

Pavlovian conditioning, it becomes a conditioned stimulus (CS) and the magnitude of the startle reflex increases if tested while the CS is present[18,19].

Defensive fixed reaction patterns. Similar to reflexes, fixed reaction patterns are more or less innate behavioural responses that are automatically elicited by particular stimuli and are common to all members of a species. Departing from the more traditional ethological designation of 'fixed action patterns' (REF. 20), we here substitute 'reaction' for 'action' to convey their automatic nature. This substitution also allows us to respect the recent view that the term 'actions' should be reserved for emitted, flexible behaviours rather than elicited, fixed ones[14].

In contrast to reflexes, fixed reaction patterns are less directly related to stimulus intensity, are slower in onset and typically involve complex patterns of coordinated responses. They sometimes involve the whole body (for example, in freezing, flight and defensive fighting[21–25]) but in other cases involve more restricted groups of muscles (for example, in facial expressions or vocalizations[26–32]). Which reaction or reflex occurs depends on the context. One crucial dimension is the imminence of the threat, which controls a shift from behaviours aimed at orienting and obtaining information to those that aim to avoid detection (such as freezing) and finally to escape or fight[5,8,33].

Like reflexes, defensive fixed reaction patterns such as freezing[21,22,34,35] and flight[36] can, through Pavlovian conditioning, come under the control of CS that are associated with aversive outcomes. Although this is traditionally called 'Pavlovian fear conditioning', we prefer to use the expression 'Pavlovian threat conditioning' to avoid the implication that subjective 'fear' underlies the process[8]. By freezing or fleeing in the presence of a learned warning signal, the organism may prevent being detected or captured. Species-typical innate defensive fixed reaction patterns, like reflexes, are also widespread in the animal kingdom[37–41].

Defensive instrumental behaviours

Organisms can learn, through experience, to produce arbitrary (rather than fixed, species-specific) responses[10,42–45]. Pressing a lever[46] or shuttling in a runway[47,48] to escape or avoid harm are laboratory examples in the defensive realm, and comparable behavioural paradigms have been developed for human studies[49–54].

Behaviours such as these, which depend on past experience, are often referred to as instrumental responses. Traditionally, two types of such instrumental response are distinguished: actions, which are goal-directed, and habits, which are not. These responses depend on previous experiences in different ways: computationally, they derive their action preferences by using, or respectively not using, a model of the environment. The traditional taxonomy emphasizes actions as being acquired via trial-and-error experiences with an action's outcomes. However, we here add a second type of goal-directed action — one that is based on cognitive deliberation (or forecasting) of possible future outcomes and requires a more elaborate world model than trial-and-error learning. We thus distinguish two kinds of instrumental behaviour (habits and actions) and within actions distinguish action–outcome behaviours and deliberative actions, with the latter involving either implicit or explicit cognitive forecasting.

Defensive habits. Instrumental responses can, in some cases, be acquired through habit learning[13,55–61]. With so-called 'model-free' learning, actions that achieve some reinforcing outcome are repeated when similar stimuli arise again. The resulting action tendencies, or habits, are thus akin to reflexes and fixed reaction patterns because they are automatic, stimulus-triggered responses; however, the elicited response is arbitrary and learned rather than preprogrammed. These responses are known as habits because, once acquired, they are stimulus-elicited regardless of the actual value of the outcome and are thus difficult to break.

The existence of habits means that not all instrumental behaviours are goal-directed; that is, not all such behaviours are chosen in a way that depends on obtaining a particular, valued outcome. Distinguishing habits from goal-directed instrumental behaviours is difficult; however, recent work on appetitive conditioning has identified strategies for doing so[10]. These strategies may be applicable to aversive instrumental responses (that is, avoidance responses) and thus may be used to determine under what circumstances they are goal-directed versus habitual (BOX 1).

In children and adolescents, habit learning seems to be the dominant means by which instrumental responses are acquired[62]. In adults, habits commonly emerge when a stimulus–response association is strengthened by being repeatedly reinforced and are typically observed following extensive training[57,58]. However, in adults, habits can also develop alongside goal-directed actions[63–65] and contribute to behaviour even early in training, as verified by neural manipulations that inactivate outcome-dependent behaviour[66]. Habitual avoidance is a way to defend against harm quickly and efficiently, but habits can also be pathological. Alterations in the balance between habitual and goal-directed behaviour have been implicated in conditions such as obsessive–compulsive disorder and substance abuse in humans[50,67]. For example, handwashing is a useful way to avoid bacterial infections; however, when it becomes habitual and independent of any actual benefit, it can become pathological.

Defensive action–outcome behaviours. When the performance of an instrumental behaviour is dependent upon its learned consequences, it is said to be goal-directed. This definition may seem tautological, but (as mentioned above) other instrumental behaviours (notably habits) that may appear goal-directed are not (BOX 1). Truly goal-directed actions depend on two factors[10,14,57]: the recognition of a contingency between action and outcome and the current status of the outcome as a valued goal. Such goal-directed actions are typically said to be emitted (rather than elicited or triggered) in the presence of relevant stimuli, reflecting the active engagement of the organism in deciding what to do when.

Pavlovian conditioning
The process through which animals learn to associate initially arbitrary stimuli with biologically important stimuli such as threats.

World model
An internal representation of the contingencies of the environment, such as a spatial map or the steps in a task.

Appetitive conditioning
Learning based on the prediction of rewards.

Box 1 | Actions versus habits in avoidance

Whether a behaviour is goal-directed or habitual cannot be determined by simply observing its performance: additional tests must be used[10,14,57]. In the appetitive domain, a standard approach is to screen for the sensitivity of the behaviour to either of the two defining features of goal-directed behaviour: the action–outcome contingency (the predictive relationship between the action and outcome) or the outcome value. If these are degraded, performance of the instrumental response will diminish if it depends on goal-directed learning but will be unaffected if it is a habit. For example, if a sated animal reduces lever-pressing for food, this demonstrates the role of the outcome value in the decision to act. Persistent lever pressing in the presence of satiation is, conversely, consistent with outcome-independent stimulus–response habits.

The distinction between goal-directed and habitual behaviour is also relevant to the aversive domain, in which the nature of instrumental avoidance has been the subject of long and contentious debate[47,69,173,209–215]. Mower's classic theory argued for a two-factor (Pavlovian plus instrumental) learning account of avoidance[68]. However, lacking the modern understanding of habit, instrumental here meant goal-directed. Evidence that learned avoidance responses tend to continue after extinction of the Pavlovian conditioned stimulus was taken to mean that the response could not be instrumental and must therefore be Pavlovian in nature. For example, Bolles argued that avoidance was nothing more than a Pavlovian conditioned flight response[210,211]. With no way out of this conundrum, behavioural work on avoidance (and the nascent effort to pursue its neural basis) fell out of favour[173,216].

However, it now appears that the argument against the instrumentality of avoidance was ill-founded[173]. Much of the evidence against instrumentality was not based on goal-directed behaviours and instead involved animals that had been overtrained[48,180]. Overtraining is now recognized as a way to shift animals from goal-directed actions to habits[66,160,173]. However, it is also unlikely that avoidance is simply a flight response. The conditioned escape and avoidance responses that occur in typical laboratory procedures are physically quite different from innate flight, which is a stereotyped burst of activity. Indeed, recent studies have shown that the circuits underlying Pavlovian conditioning of innate flight reactions[36] overlap with the Pavlovian freezing circuit and that both are distinct from the avoidance learning circuit[168,173].

Modern concepts and recent findings thus go a long way towards correcting the misconception of avoidance as nothing more than a Pavlovian response. The failings of the two-factor theory are thus overcome by adding habit as a third factor[173]. Armed with the conceptual advances that have been made in appetitive conditioning, it should now be possible to determine, using contingency degradation and/or outcome devaluation studies, under what conditions avoidance is a goal-directed instrumental response as opposed to a habit.

Outcomes such as these are reinforcing, as they increase the likelihood of the action taking place again in the future, but they are sometimes called 'goals' to distinguish them from other, more conventional reinforcers and to capture the prospective focus of the action.

The classic case of goal-directed behaviour studied in the laboratory involves directly learning the action–outcome contingency (for example, by allowing an organism to observe that a lever press is followed by food or that a shuttling response is associated with safety). We use the expression 'action–outcome behaviours' to describe actions based on contingencies learned in this way. Computationally, an action–outcome contingency is the simplest case of what computer scientists know as a world model — a representation of the predictive contingencies governing a task. Goal-directed action is thus an instance of model-based evaluation — choosing an action on the basis of a (simple) model predicting its consequences.

The goal-directedness of behaviour has been studied extensively in the appetitive domain. As we argue in BOX 1, certain forms of avoidance behaviour can also be considered essentially goal-directed. These allow an organism to select actions when in danger with the goal of escaping from or avoiding exposure to harm on the basis of past experience with the action–outcome contingency. Such learning is often said to involve two processes or factors: an initial stage of Pavlovian conditioning, leading to freezing, followed by a stage of instrumental learning[47,68,69]. In contrast to this active avoidance, circumvention of harm can also occur through inaction[70]. In this so-called passive avoidance, harm is avoided by remaining still. We focus on active avoidance in this Review because in passive avoidance it is harder to separate Pavlovian and instrumental influences on learning and to distinguish outcome-dependent and habitual responses. Below, 'avoidance' therefore refers to the active form unless otherwise indicated.

Active avoidance
A type of experimental task in which organisms must produce a particular response to avoid harm.

Defensive deliberative actions. Traditionally, a defining characteristic of outcome-dependent actions is that their performance depends on learning the action–outcome contingency; that is, the animal must learn that lever pressing produces food or shuttling avoids harm[10,14]. We here propose that, in some cases, contingency may go beyond an animal's direct experience of the association between a particular action and outcome and that goal-directed decisions can also be guided by constructive or extrapolative planning, in which the relationship between an action and its ultimate outcome is forecasted, indirectly, by drawing on diverse mnemonic representations, such as spatial maps or world models[12,63], episodes[71,72] or schemas[73]. We refer to these sorts of actions as 'deliberative actions', highlighting the necessity of actively constructing or forecasting the outcome.

Computationally, model-based control generalizes beyond single-step action–outcome models to more elaborate representations, the consequences of which for a candidate action's outcome must be iteratively computed via some process such as a tree search. The computational concept of model-based control probably corresponds to a number of neurally and psychologically distinct systems that work with different representations.

The need to reason about the consequences of an action arises in many circumstances, including when past experience does not provide a relevant action–outcome contingency, when relationships are indirect (involving multiple steps or more than one action and outcome; for example, during the planning of spatial trajectories) and when a previously learned action–outcome contingency must be generalized to wholly novel situations. For example, if you were previously successful in escaping from harm by running away but are trapped on a riverbank with no clear escape route, you might generalize from your experience with running and choose swimming as the next option.

The suggestion that multiple levels of instrumental control are layered atop simpler, reflex-based and reaction-based defence circuitry also raises the question of the circumstances in which these different strategies are deployed. One intriguing suggestion, from the recently published 'survival optimization system theory' (REF. [5]), is that the hierarchy of computational complexity at least partly corresponds to and extends classic ideas about the hierarchy of responses to threat imminence. Specifically, it is suggested that slow and computationally laborious deliberation is best suited to situations requiring prevention and avoidance when the agent is fairly safe.

We suggest that such extrapolative planning involves implicit (nonconscious) cognitive processing and/or conscious deliberation. This proposal sharpens an earlier suggestion by Balleine and Dickinson[74], who associated consciousness with goal-directed behaviour generally but did not distinguish between action–outcome behaviour and more deliberative forms of behaviour that depend on forecasting rather than directly learned contingencies.

That many cognitive actions can be planned and executed without requiring explicit conscious deliberation is now known. For example, recent work has shown that flexible decision phenomena can function implicitly[75,76]. Furthermore, although deliberation is often associated with working memory and working memory is often linked with consciousness[77–82], recent work has shown that working memory can function independently of consciousness[83–88].

We argue that, in most animals, implicit cognitive processing is likely to dominate[8,89]. Evidence shows that birds have elements of episodic memory[90]; however, evidence that these episodic-like representations are actually consciously experienced is lacking as the studies in this area do not clearly distinguish conscious from nonconscious metacognitions[89,91–93]. Heyes argues that, just as planetary motion follows rules, so does animal behaviour[89]. However, neither planets nor animals know the rules or consciously choose to follow them. She argues that nonhuman animals primarily depend on implicit, domain-general rules, such as those underlying complex associative learning, to guide behaviour, whereas humans (especially in social situations) are able to use specific, conscious and verbally reportable rules[89]. We consider that both human and nonhuman animals undertake elaborate, nonconscious forms of cognitive deliberation. For instance, planning novel routes in spatial navigation or deciding to cache food to meet anticipated future needs requires some form of iterative traversal of a mental map or model[90,94]. However, the complexity of this forecasting does not in itself necessarily imply that it occurs consciously[89,95–97]. Thus, conscious deliberation, especially when involving linguistic reasoning, may be specific to humans. However, even humans are not aware of all cognitive deliberation in real time. As Lashley[98] and many others since have pointed out, we have limited access to the cognitive processes underlying the introspective conscious content that is sometimes generated as a by-product of such processing.

In the defensive situation, one reason to consider the distinction between nonconscious and conscious cognitive actions is to question the folk-psychological expectation (reified in scientific terminology) that defensive responses of all kinds are causally motivated by conscious emotional states such as fear. Fear is relevant, but only in some circumstances in some species. The contribution (or lack of contribution) of fear to each of the defensive behavioural categories described above is discussed in BOX 2.

Neural circuits for defence

The circuitry underlying defensive behaviours changes as reactions give way to actions, as actions become habits and as circuits underlying forecasted outcomes take over when past instrumental learning has not provided a suitable action or habit (FIG. 1). In addition, distinct circuits are engaged when implicit or explicit forecasting is used to guide actions that allow escape or avoidance of harm. Throughout much of the discussion below, the details we describe have been most extensively documented in rodents; however, they are consistent in the other mammals, including humans, in which they have been studied to a lesser extent. Nevertheless, in the case of explicit, conscious deliberation, data exist only in humans.

Circuits for innate responses

Reflex circuits. The acoustic startle reflex has been studied extensively both behaviourally and neurobiologically[99–102]. Studies in rodents show that reflex depends on connections between the brainstem auditory neurons that process the startle-eliciting stimulus and the motor circuits that control neck, face, eye, back and leg muscles. Specifically, the cochlear nerve transmits auditory signals to the cochlear nucleus in the brainstem. The cochlear nucleus connects to the pontine reticular nucleus, the outputs of which innervate facial, cranial and spinal motor neurons. These connections are made up of short pathways that allow the reflex response to be executed within 6–10 ms of the onset of the acoustic stimulus.

We focus on the startle reflex over other reflexes because of its close relation to the circuitry underlying Pavlovian threat conditioning (described below). Specifically, a Pavlovian CS modulates startle through output connections of the central nucleus of the amygdala (CeA) to the pontine reticular nucleus[19]. The CeA is also involved in controlling other innate and conditioned reactions and in the learning of instrumental actions, as described in the following sections.

Fixed reaction pattern circuits. Freezing behaviour, a defensive fixed reaction, has been studied in rodents and nonhuman primates and can be elicited by both innate threat stimuli[23,103–105] and stimuli that have become threats through Pavlovian threat conditioning[21,22,106,107]. The amygdala is a central hub in the circuitry of freezing in response to both categories of threat stimuli[8,23,37,107–109]. In both cases, the amygdala receives sensory inputs about the threat and connects with downstream targets, especially the periaqueductal grey (PAG) region, to control the expression of freezing responses. However, the amygdala subregions and downstream targets differ somewhat in the cases of innate and learned threats.

Innate threats include the sights, sounds and odours of predators. Auditory and visual information about predators is received by the lateral amygdala (LA), which projects to the accessory basal amygdala (ABA), whereas olfactory predator cues enter the amygdala via the medial nucleus[23,110,111]. Both the ABA and the medial amygdala project to the ventromedial hypothalamus,

Box 2 | Conscious emotional experiences and explicit action forecasting

Darwin argued that emotions, such as fear, are innate states of mind that cause innate behaviours[26]. Indeed, it is common today to assume that fearful feelings are the causes of defensive fixed reaction patterns and defensive responses[217–219] and, similarly, that a reduction in fearful feelings reinforces the acquisition of instrumental avoidance responses[47,68,173,209,212–215]. However, several lines of evidence question the validity of this conclusion[8,77,173,195,207]. First, in dangerous situations, subjectively experienced fear is not reliably correlated with behavioural defence responses. Second, when subjective awareness of the threat stimulus is prevented experimentally, the stimulus still elicits body responses despite the person not being able to report what the stimulus is, and despite not reporting any feelings of fear, even when prompted. Third, brain damage that disrupts behavioural responses elicited by threats does not necessarily eliminate conscious feelings of fear. Fourth, pharmaceutical treatments do not affect defensive responses and feelings of fear equally.

Some investigators use ‘fear’ to denote a nonsubjective state that underlies defensive behaviour[19,21,69,220–222]. However, the neural states and brain circuits that control innate defensive responses, actions learned by their outcomes and cognitively forecasted actions clearly differ. Thus, many such states are currently labelled by the undifferentiated term ‘fear’. This problem could be remedied if we assume that there are different varieties of nonsubjective fear states. However, a more fundamental problem[223] is that the use of a subjective-state term to describe nonsubjective states (such as fear, hunger or pleasure) means that our concept of the neural circuits in question becomes conflated with the subjective properties that we are trying to circumvent[3,4,8,77,195,207,208,224]. Thus, most people who read or hear about such ‘fears’ think the author is describing subjective experiences.

To avoid confusion, the neural state that controls freezing and supporting physiological responses has been called a neural state of a ‘defensive survival circuit’ (REFS [3,4,8]). The defensive survival circuit state initiates a more general state of brain and body arousal that has been called a ‘defensive global organismic state’ (REFS [3,4,8]). This state then invigorates and directs behaviours on the basis of an action–outcome contingency. This conception leaves the term ‘fear’ to unambiguously denote the subjective experience of fear itself. In our taxonomy, we thus exclude a necessary role of conscious fear in defensive reflexes, fixed reactions, habits, action–outcome behaviours and implicit cognitive forecasting but give it a considerable role in explicit forecasting in threatening situations.

An emerging view treats fear as a cognitive state constructed from nonemotional ingredients[8,77,225]. Specifically, this suggests that fear arises from the coalescence in the working memory of sensory signals, memory signals and the various physiological consequences instantiated in the global organismic state that follows survival circuit activation[4,8,77,195,207,224]. What makes fear different from other emotional states, and emotional states different from nonemotional states of consciousness, is thus the set of underlying signals being processed in working memory (a related view has also been expressed by Barrett[225]).

Working memory is now recognized to have conscious and nonconscious facets[83–88]. Therefore, it is proposed that the cognitive state that initially results in working memory is a nonconscious representation and that re-representation of this state results in a conscious state of fear[8,77] (FIG. 3). Once present as a conscious state in working memory networks, fear is proposed to become a major factor in informing decisions about how to act in a given situation[8]. We propose that conscious fear contributes to explicit deliberative forecasting. For example, if you are walking down a dark street and notice a group of shady characters ahead, this may arouse fear and prompt you to decide to change your route. On the other hand, it is also possible that you first decide nonconsciously and then only consciously rationalize the choice after the fact[226]. Distinguishing which form of deliberation underlies decisions is thus an important challenge for the field.

which connects with the PAG. The PAG is the interface with motor systems controlling freezing and other innate defensive behavioural responses.

For learned threats, the circuits overlap with, but also diverge from, those responsible for processing innate threats[8,23]. Information concerning the CS again arrives via the LA, where it acquires threatening properties during Pavlovian threat conditioning by way of synaptic plasticity that is induced by the convergence of the CS and US pathways onto single LA cells. After conditioning, information concerning the CS is therefore able to flow through the LA to the other amygdala regions that control freezing. The cellular and molecular events in the LA that transform the neutral stimulus into an aversive CS are well characterized[108,109,112–119]. The LA connects to the CeA through direct and indirect circuits within the amygdala[110,120]. Within the CeA, different classes of identified cell types interact to process the inputs and control further outputs[36,118,121–125]. Through connections from the CeA (especially its medial division) to the PAG, CS-evoked freezing is triggered[36,122,126–130]. The outputs of the CeA to hypothalamic and brainstem targets control not only freezing behaviour but also autonomic and endocrine reactions elicited by the CS. CeA outputs also activate neuromodulatory systems, which release noradrenaline, dopamine, serotonin, acetylcholine and other modulators across the brain. Areas of the amygdala are among the targets of these modulators, creating a feedforward process that helps to sustain vigilant processing of and responding to the threat[8]. Furthermore, CeA outputs to the brainstem startle circuitry (described above) allow conditioned threats to enhance startle reflexes[19].

The processing of innate and learned threats in the amygdala is regulated by other brain areas[8,131–134]. The hippocampus provides information about the context in which the danger is occurring, whereas the infralimbic and prelimbic regions of the medial prefrontal cortex facilitate the adjustment of amygdala output activity in response to changes in the threat potential of the environment. The paraventricular thalamus regulates CeA-mediated consolidation of threat memories and their behavioural expression[123,135].

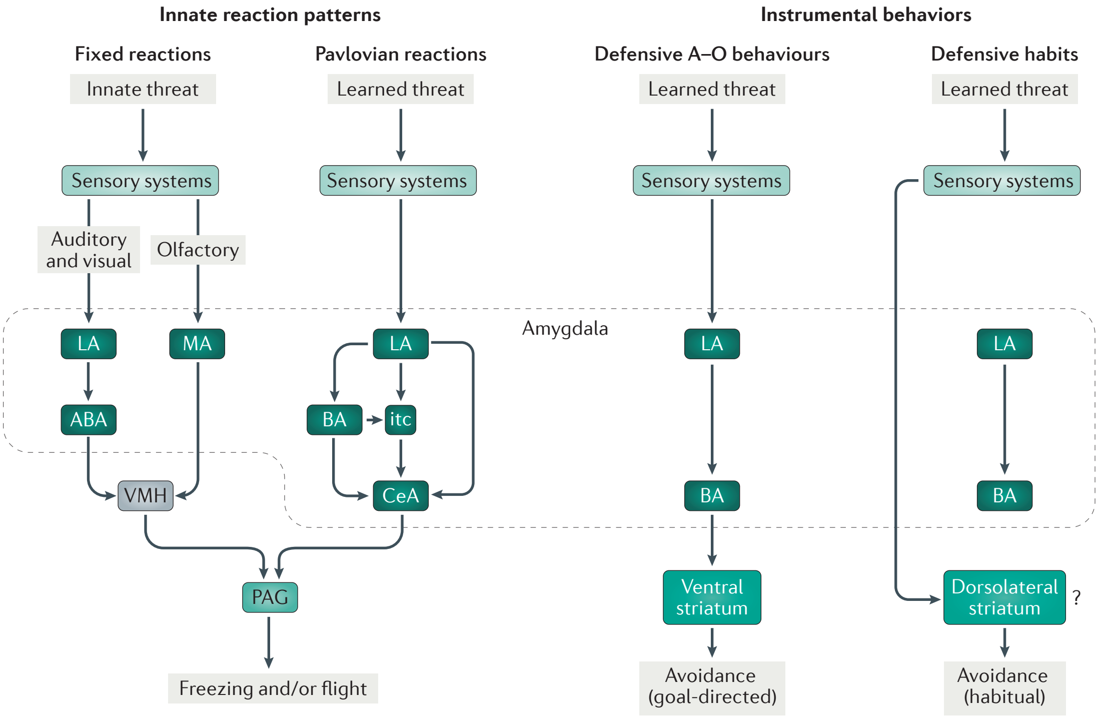

Figure 1 | **Neural circuits underlying innate reactions and instrumental actions and habits.** Schematic images illustrate the circuits proposed to underlie defensive fixed reactions, action–outcome (A–O) behaviours and habits. Defensive fixed reactions can be elicited by species-typical (known as innate or prepared) stimuli (innate threats) or by previously neutral stimuli that have become associated with innate threats (learned threats) via Pavlovian conditioning[21–25]. Learned threats can also initiate instrumental behaviours[47,48,68,69,171,173]. Those that are controlled by their learned outcomes are A–O behaviours, whereas habits are instrumental behaviours that are performed independent of learned outcomes[10,13]. The circuits that mediate innate fixed[23–25] and Pavlovian[108,109] reactions and A–O behaviours[171,173,216] are well-established pathways that convey sensory information through a series of amygdala nuclei to descending pathways via the periaqueductal grey (PAG). By contrast, little is known about the habit circuitry except that the amygdala does not participate[160,161]. Extrapolation from appetitive conditioning research suggests that the dorsolateral striatum plays a role in defensive habits[12,66,174]; however, this has not been tested (indicated by a question mark). ABA, accessory basal amygdala; BA, basal amygdala; CeA, central amygdala; itc, intercalated nuclei of the amygdala; LA, lateral amygdala; MA, medial amygdala; VMH, ventromedial hypothalamus.

When there is an element of uncertainty about the threat, the role of the amygdala diminishes and the bed nucleus of the stria terminalis (BNST) plays a greater role[136–140]. Although this role is often said to be that of creating a state of anxiety, a more objective description states that the BNST processes future, uncertain threats[8]. In doing so, it receives connections from the hippocampus rather than from the sensory pathways that feed into the amygdala; however, its outputs overlap with those of the amygdala.

The results described above arise from studies in rodents and, to some extent, nonhuman primates. The human brain cannot be explored with the same precision. However, research on patient populations, as well as brain imaging studies of healthy participants, confirm the basic findings of animal research[141–143]. Thus, lesions of the human amygdala disrupt Pavlovian threat conditioning[144,145]. Following Pavlovian threat conditioning, the CS elicits increased neural activity in the amygdala (as measured by functional MRI studies)[146,147]. Depth electrode recordings also support a role for the human LA in the rapid processing of threat-related stimuli[148,149]. Furthermore, imaging studies implicate the BNST in processing uncertain, future threats[150].

In humans, it is considerably easier to make distinctions between conscious and nonconscious processing than it is in animals. On the basis of procedures such as masking or bistable stimulation (both of which manipulate the conscious availability of perceptual stimuli), studies have suggested that the amygdala supports implicit or nonconscious processing of Pavlovian conditioned threats[151–156]. This means that the contribution of the amygdala to threat processing can be studied similarly in humans and other mammals without making it difficult to defend assumptions about animal consciousness.

Circuits for instrumental responses

Avoidance circuits. Avoidance is the prototypical defensive action. Compared with appetitive behaviour, for which the goal is typically concrete and affirmative, the goal of defence is to avoid a threatened aversive outcome. Therefore, success at this goal is signalled negatively by omission of the threat. This negative signalling means that learning to avoid is predicated on first learning that some threat stimulus predicts the aversive outcome.

Active avoidance learning has long been viewed as a two-stage learning process[47,68]. First, a Pavlovian CS–US association between the warning signal and a shock is made. Then, over time, instrumental behaviours are acquired because of their success in avoiding the harm produced by the expected shock. These two processes are particularly separable in signalled active avoidance tasks (the paradigm principally discussed here), in which a CS signals impending threat but can be terminated (and the threat avoided) by a particular response.

Although early efforts to identify the neural circuits of avoidance used crude lesion methods and produced confusing results[157], more recent research has built on the success of circuit studies of Pavlovian conditioning to implicate a specific set of circuits in active avoidance. Specifically, this work has shown that the LA and the basal amygdala (BA) are required for active avoidance and (unlike for reactive freezing) the CeA is not[158–164]. This finding suggests that reactive freezing and active avoidance depend on different intra-amygdala circuits that emanate from the LA: connections from the LA to the CeA for reactive freezing and connections from the LA to the BA for active avoidance. The involvement of the LA in both responses suggests that a CS–US association encoded within the LA circuits during Pavlovian conditioning is used for both reactions and actions evoked by the CS, presumably reflecting the aforementioned dependence of the avoidance response on the Pavlovian association.

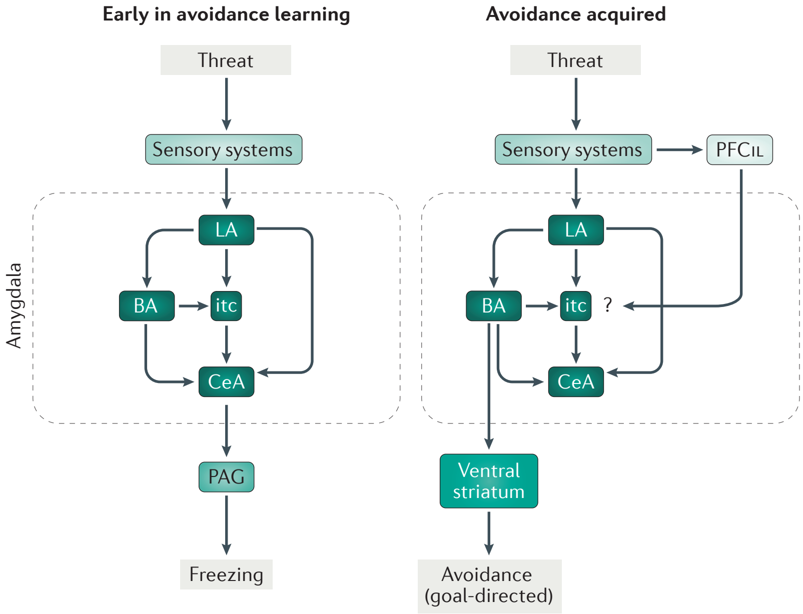

Figure 2 | **Switching between reactive and active coping during active avoidance learning.** Freezing and avoidance are competing defensive behaviours: actions such as avoidance cannot be taken while a freezing reaction is being expressed. Active avoidance is acquired in stages[68,69,215]. For example, in a task in which a tone warning signal precedes an electric shock, the tone initially becomes a threatening Pavlovian conditioned stimulus that elicits freezing (a form of passive coping). Over multiple trials, the animals learn to actively perform responses that turn off the tone and thus prevent the shock from being delivered (a form of active coping). Early in training, the tone controls freezing through a pathway involving the lateral amygdala (LA), the central amygdala (CeA) and the periaqueductal grey (PAG)[108,109] (see also the Pavlovian reaction in FIG. 1). Later in training, LA outputs are switched from the CeA to the basal amygdala (BA), which then connects with the ventral striatum to control the active avoidance response (see also FIG. 1). The switch appears to be controlled by the infralimbic region of the medial prefrontal cortex (PFCIL)[159,171,173]. However, the exact effect on the amygdala is not known (indicated by the question mark). Under certain conditions, this goal-directed instrumental action–outcome behaviour can lose its relationship to the outcome and become habitual (not shown, but see FIG. 1). itc, intercalated nuclei of the amygdala.

Unlike freezing, which is driven by CeA connections to the brainstem, avoidance requires connections from the BA to the ventral striatum (specifically, the nucleus accumbens (NAcc))[161,162,165–168]. This circuitry is also similar to that required for another variant of active avoidance conditioning, known as 'escape from threat', in which rats learn to perform actions that are reinforced solely by CS termination[169]. The amygdala circuits underlying escape from threat mirror those of signalled active avoidance: LA and BA are required, but CeA is not[170].

Not only is the CeA unnecessary for active avoidance learning, the reactive freezing responses that it subserves actually appear to compete or interfere with avoidance: lesions of the CeA eliminate freezing and enhance the acquisition and expression of avoidance[160,161,171]. CeA lesions also rescue performance in animals that fail to express avoidance behaviour owing to excessive freezing, allowing these 'poor performers' to start to avoid normally[172]. Thus, CeA lesions do not prevent avoidance but do constrain expression indirectly by eliminating freezing, which is a competing response (active avoidance cannot take place in a freezing organism)[173]. CeA lesions also modulate instrumental appetitive actions by altering competing responses[174–176]. However, the CeA is known to be required for the acquisition of appetitive instrumental habits — this is not the case in avoidance[174].

A distinction between aversive actions and reactions can also be observed at the level of the NAcc. A successful avoidance response to the presentation of a warning signal was preceded by an increase in NAcc dopamine. If no such dopamine increase was observed, subjects failed to avoid[166]. Conversely, presentation of an aversive Pavlovian CS caused a decrease in NAcc dopamine release, suggesting that defensive actions and reactions also have distinct neurochemistry in NAcc.

How does the transition from reactive freezing to active avoidance occur? Recent findings indicate that the medial prefrontal cortex plays a key role[159,171,173,177] by switching behavioural control from the dominant LA–CeA–PAG circuit to the LA–BA–NAcc circuit (FIG. 2). In particular, the infralimbic region of the medial prefrontal cortex (PFCIL) suppresses defensive reactions and facilitates defensive actions by toggling between different amygdala output pathways, inhibiting reactions controlled by the CeA–PAG projection (freezing) and enhancing actions controlled by the BA–NAcc projection (avoidance)[171,173].

Few studies have examined the brain mechanisms of active avoidance in humans. However, the results of those that have are broadly consistent with the

Reward devaluation
An experimental procedure in which the value of an action's outcome is reduced; used to verify that the action is goal-directed as opposed to habitual.

animal literature. Thus, studies using functional imaging have implicated the amygdala, NAcc and medial prefrontal cortex in active avoidance in humans[49,52–54,178].

Habits versus action–outcome behaviours. Although evidence from manipulations designed to assess habitual behaviour, such as reward devaluation, is not yet available for aversive behaviour, the avoidance circuits described above strongly parallel those that support goal-directed, rather than habitual, behaviour in appetitive conditioning (BOX 3). This similarity supports the idea that these avoidance actions are likely to also be goal-directed.

However, in addition to the sequential acquisition of defensive reactions and actions, we believe that a third mechanism — defensive habit — is also involved in avoidance learning[173]. For appetitive behaviour, habits depend on the dorsolateral striatum in both animals and humans[50,55,56,61,179]. Studies in rodents show that the transition of appetitive instrumental behaviour from goal-directed actions to goal-independent habits reflects a shift from circuits that control goal-directed actions to the dorsolateral striatum habit circuitry[66].

There is increasing (although still indirect) evidence that an analogous progression occurs for avoidance responses. Thus, with long-term training, avoidance responses tend to resist extinction of the CS[48,180] and the avoidance response becomes independent of the amygdala[160,164]. A crucial aim for future work will be to investigate whether the amygdala-independent responses are indeed habits that depend on the dorsolateral striatum.

Avoidance learning, although traditionally described as two-factor learning[47,68,69], thus appears to proceed through three distinct phases, each associated with its own neural circuitry[173]. The first is Pavlovian, involving defensive reactions underpinned by the LA–CeA–PAG pathway. The second is instrumental and involves defensive actions that require the LA–BA–NAcc pathway. In order to transition from reaction to action, the PFCIL is recruited to suppress freezing and facilitate avoidance. The third and final phase involves defensive habits, which are independent of the amygdala. Extrapolation from appetitive findings suggests that the dorsolateral striatum is involved.

Deliberative actions. Above, we have suggested that a pair of circuits involving different nuclei of the amygdala and striatum support two pathways for instrumental avoidance: goal-directed actions that result from action–outcome associations and stimulus–response habits. However, the behavioural taxonomy that we describe (TABLE 1) posits additional categories of goal-directed behaviour that support more flexible, constructive forecasting of action outcomes and draw on knowledge beyond unitary action–outcome associations. In some cases, such forecasting involves explicit, conscious deliberation. As used here, however, deliberative does not necessarily imply conscious deliberation because much cognitive processing is known to occur nonconsciously (including the integration of sensation and memory in perception, the planning of sentence structure and the predispositions, attitudes and biases that shape behaviour and decision-making itself[181–185]).

A key reason to differentiate these additional categories of goal-directed actions from those described above is that their neural substrates are distinct. In particular,

Box 3 | Circuit parallels between appetitive and aversive instrumental behaviour

Comparison of the circuitry for avoidance with that for appetitive instrumental behaviour reveals impressive parallels. For example, acquisition of active avoidance depends on the lateral amygdala (LA) and basal amygdala (BA); however, once avoidance is overtrained (presumably becoming habitual), these areas are not needed[159,160]. Similarly, the LA and BA are required for appetitive instrumental goal-directed behaviours but not for habitual instrumental responses[227,228].

Evidence from manipulations designed to test for habitual behaviours, such as reward devaluations, is not yet available for avoidance behaviour. However, the parallels between the circuits involved in avoidance and appetitive behaviours suggest that the LA–BA circuit contributes to active avoidance by enabling a goal-directed instrumental action rather than a habit or Pavlovian reaction. If correct, this suggestion would resolve the controversy that arose from Bolles's argument that avoidance is a Pavlovian conditioned flight response[210,211] (BOX 1). Positioning avoidance instead as an example of goal-directed instrumental choice is consistent with the clear anatomical dissociation between an LA–central amygdala (CeA) pathway for Pavlovian freezing responses and an LA–BA pathway for avoidance[173,216].

This same reasoning may also help to resolve one key difference between appetitive and aversive behaviour: the LA and BA are required to acquire avoidance behaviour (although performance later becomes independent of these structures, which we suggest provides evidence of a transition to habitual control)[173]. For appetitive tasks, however, animals can acquire instrumental responses for reward even with pre-training lesions to these structures[227]. However, without the LA and BA, these appetitive behaviours are evidently directly acquired as habits: they are insensitive to reward devaluation even before overtraining[227]. This concept suggests that animals can acquire appetitive habits, but not avoidance, without the LA and BA.

This seeming inconsistency may reflect a crucial conceptual difference between appetitive and aversive situations. An action can be directly reinforced by the desirable outcome it produces: this allows appetitive habits to be learned by reinforcement (without passing through a goal-directed stage). However, successful avoidance has no direct reinforcer: success in avoidance is signalled, instead, by the absence of an aversive outcome. The ability to recognize when avoidance succeeds (and learning to repeat that response, as a habit) depends on having previously learned to expect the aversive outcome. That is, avoidance learning rests on prior Pavlovian learning (and hence the LA), whereas appetitive learning does not. Indeed, in a special case of appetitive conditioning known as conditioned reinforcement, in which Pavlovian conditioned stimuli (rather than primary reinforcers) reinforce instrumental behaviour, the LA–BA complex is necessary[229,230].

unlike simple action–outcome instrumental learning[186], these more flexible outcome-directed behaviours are likely to require long-term memory systems (especially the hippocampus) and prefrontal areas supporting working memory (FIG. 3). Because these sorts of behaviour have not been studied to any considerable degree in avoidance, we view this as an important area for future work. Nevertheless, results from appetitive tasks provide hints.

Perhaps the most famous example of goal-directed deliberative behaviour is route planning in spatial navigation, such as that required to find novel shortcuts. Planning a spatial route, drawing on information about obstacle and goal locations (the 'cognitive map'), illustrates a goal-directed behaviour that is constructed using richer information than just the association between a single action and outcome. Such spatial planning is widely believed to depend on the hippocampus, although (as with simpler instrumental behaviours) it is known that this behaviour can transition to a simpler striatum-dependent habit stage with overtraining (under some conditions the two are acquired in parallel)[50,64–66].

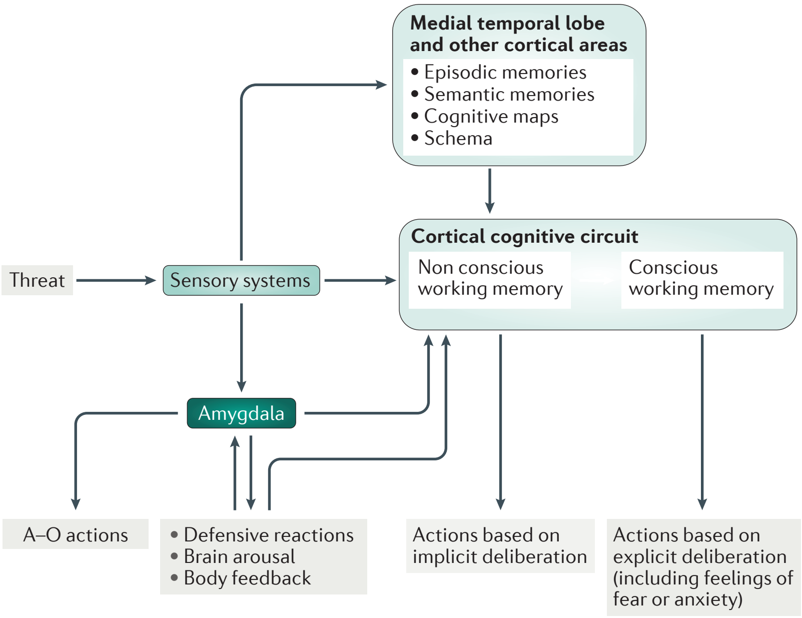
Medial temporal lobe and other cortical areas
• Episodic memories
• Semantic memories
• Cognitive maps
• Schema
Cortical cognitive circuit
Non conscious working memory
Conscious working memory
Threat
Sensory systems
Amygdala
A–O actions
• Defensive reactions
• Brain arousal
• Body feedback
Actions based on implicit deliberation
Actions based on explicit deliberation (including feelings of fear or anxiety)

Figure 3 | **Proposed circuits for deliberative defensive actions.** In addition to direct action–outcome (A–O) learning, actions can be guided by more constructive or extrapolative deliberation. Sensory systems deliver threat information to the amygdala[110,111], which controls defensive reactions by way of connections to the periaqueductal grey[108,109] and A–O actions by way of connections to the ventral striatum[171,173,216] (see text and FIGS. 1,2). Sensory systems separately deliver threat information to medial temporal lobe systems that form and store various kinds of long-term memory representations and the cognitive control circuits that underlie temporary or working memory[86,111,190,192]. Included in the cognitive control networks are the lateral and medial prefrontal and parietal cortex, the insula cortex and the interactions between these areas. Nonconscious working memory[83–88] is proposed to integrate sensory, memory, amygdala, brain arousal and body signals[8,77]. The resulting representation can be used by cognitive control processes to initiate defensive actions implicitly (nonconscious deliberation). Re-representation of nonconscious working memory states is proposed to result in a conscious state (a thought or an emotional feeling) that can contribute to explicit action choices (conscious deliberation)[8,77].

Planning in a multistep decision task widely used to index model-based control is also impaired by hippocampal inactivation[187]. However, hippocampal lesions do not affect goal-directed (that is, devaluation-sensitive) instrumental lever-pressing for reward, a behaviour that is presumably based on directly learned action–outcome associations[186]. We suggest that this reflects a distinction between the neural substrates supporting direct action–outcome learning and those supporting more deliberative forms of behaviour.

The hippocampus also supports flexible, constructive decision-making in a variety of nonspatial settings, including tasks requiring piecing together relations among multiple stimuli[75,76,188], 'model-based' planning over multiple steps of actions[187] and decisions based on episodic memories or schemas[72,73]. These are examples of goal-directed choice behaviour that depend on different sorts of long-term memory supported by the hippocampus; therefore, it is likely that analogous results would be seen for avoidance versions of these tasks, although this remains to be tested.

What then, would be the role of conscious awareness, including conscious fear, in these behaviours? The answer to this question reflects the final subdivision of our taxonomy. Hippocampal-dependent memory is classically known as 'explicit' or 'declarative' memory because humans can in many circumstances verbally declare its contents[189]. However, although hippocampal memories can be readily made conscious, the access and use of hippocampal memories, which may require working memory (see below), are not necessarily conscious[8,190]. Thus, it is conceivable that flexible decision behaviours informed by hippocampal memories can reflect nonconscious processing despite the fact that we can also consciously access those memories. Effects such as acquired equivalence and sensory preconditioning are canonical examples of deliberative (but not necessarily conscious[191]) forecasting of value and are rooted in relational memories stored by the hippocampus and overlying cortices in both humans and nonhuman animals[75,76,192–194]. Thus, humans show these effects (expressed, for instance, as preferences between stimuli that have differential, indirectly learned relationships to reward) without being able to consciously report the underlying chain of reasoning[75,76].

Thus, we argue that, regardless of the conscious status of other animals, human conscious experiences have a decisive role in behaviour that is conceptually, psychologically and neurally distinct from the processes that control reflexes, fixed reactions, habits, contingency-dependent learned behaviours and even behaviours on the basis of implicit forecasting. Although we intend this as a general statement about conscious experiences, we are particularly interested in how subjective experiences of fear might affect defensive behaviour. Thus, although fear is not necessarily the cause of fixed reactions such as freezing, instrumental action–outcome behaviours such as avoidance or even certain flexible deliberative behaviours controlled implicitly, it can still be instrumental in controlling behaviour. To explain this, we must define what we mean by fear.

In our conceptual scheme, fear results in part from inferences that arise from self-monitoring, generating a complex higher-order state that is assembled in working memory through the integration of various nonconscious lower-order representations[8,77] (BOX 2). Once fear has been assembled in working memory, we propose that it operates in the same manner as any other kind of working memory representation (for example, biasing attention and further processing of relevant information). Importantly, we suggest that subjectively experienced fear serves as a basis for deliberative explicit decision-making and behavioural control so as to guide avoidance of or escape from an existing or anticipated threatening situation[8,77,195]. Although fairly little work has been done on the neural basis of conscious fear itself, prefrontal working memory circuits are, at a minimum, engaged[196], suggesting that conscious fear emerges via working memory in a manner that is similar to nonemotional conscious experiences[77]. Given that working memory has both nonconscious and conscious aspects, additional work is needed to specify the nature of the conscious and unconscious behavioural control that is related to the experience of fear (BOX 2).

All this suggests that deliberative actions should depend on working memory (see earlier discussion) and its underlying neural substrates, such as dopaminergic actions in the prefrontal cortex. Indeed, a surprising range of human decision behaviours studied in the laboratory does depend on these substrates; however, experiments manipulating interference or load also verify that these behaviours are accompanied by other influences on choice that are not necessarily dependent on conscious working memory[197,198]. Again, more work is needed to understand the dynamics of interactions between conscious and nonconscious processing.

Conclusions

Even simple behaviours can arise from a multiplicity of causes, a point that has often been more appreciated in the study of appetitive than aversive behaviour. We have proposed a number of distinct systems underlying defensive behaviours — some well documented, some extrapolated from research on appetitive behaviour and some inferred. Such a fine-grained taxonomy helps in several ways: first, it provides a framework for untangling the neural circuits underlying these behaviours; second, it helps to address some long-standing puzzles and paradoxes in the field; and third, it helps clarify and delineate the role of emotions (notably, feelings such as fear) in behaviour.

The taxonomy of healthy behaviour may also clarify the dimensions underlying pathological behaviour[199]. In particular, the notion that the brain has many routes for responding to threat raises a higher-level problem, which has also arisen repeatedly in our review: how do responses compete, and, more generally, how is the appropriate response for a given situation selected and adaptively deployed? Imbalanced deployment or switching of defensive behaviours may underlie several psychopathologies. For instance, there is now substantial evidence that the inflexible nature of habitual behaviour underlies the seemingly compulsive symptoms observed across several psychiatric disorders[50,67]. The aversive setting suggests several promising extensions to that general theme. For instance, it enables us to differentiate cognitively controlled implicit and explicit (conscious) goal-directed avoidance behaviours from reactions, the execution of which minimally engages cognitive capacities. This distinction may offer a way to better integrate behavioural theories that root disorders such as anxiety in basic processes such as Pavlovian learning[200–203] with cognitive theories that emphasize more abstract conceptual dysfunction such as maladaptive schemas and beliefs[204,205]. Excessive or uncontrolled cognitive forecasting may also underlie rumination, cognitive paralysis and overthinking in numerous mood disorders[206]. Similarly, idle passivity and worry, which are present in many anxiety disorders, and anergia and avolition in depression, might also be understood in terms of imbalance in the competition between active and passive modes of responding to threat (FIG. 2). More generally, much as Pavlovian learning has provided a useful animal model relevant to anxiety disorders, the range and sophistication of instrumental avoidance suggest even greater promise in this area. Finally, whereas subjective experience has been sidelined by both psychopharmaceutical and cognitive and/or behavioural approaches to the treatment of psychological adjustment problems, our inclusion of explicit conscious deliberation and fear as part of the human defensive repertoire recognizes the need to give greater emphasis to subjective experience in the evaluation and treatment of psychiatric disorders[8,195,207,208].

1. Emerson, R. W. in *The Selected Lectures of Ralph Waldo Emerson* (eds Bosco, R. A. & Myerson, J.) 301 (Univ. of Georgia Press, 1863).
2. Schneirla, T. C. in *Nebraska Symposium on Motivation* (ed. Jones, M. R.) 1–42 (Univ. of Nebraska Press, 1959).
3. LeDoux, J. Rethinking the emotional brain. *Neuron* **73**, 653–676 (2012).
4. LeDoux, J. E. Coming to terms with fear. *Proc. Natl Acad. Sci. USA* **111**, 2871–2878 (2014).
This article provides a summary of the issues related to describing defensive behaviour in terms of fear versus threat processing.
5. Mobbs, D., Hagan, C. C., Dalgleish, T., Silston, B. & Prevost, C. The ecology of human fear: survival optimization and the nervous system. *Front. Neurosci.* **9**, 55 (2015).
6. Emery, N. J. & Amaral, D. G. in in *Cognitive Neuroscience of Emotion Series in Affective Science.* (eds Lane, R. D. & Nadel, L.) 156–191 (Oxford Univ. Press, 2000).
7. Adolphs, R. The social brain: neural basis of social knowledge. *Annu. Rev. Psychol.* **60**, 693–716 (2009).
8. LeDoux, J. E. *Anxious: Using the Brain to Understand and Treat Fear and Anxiety.* (Viking, 2015).
9. Grupe, D. W. & Nitschke, J. B. Uncertainty and anticipation in anxiety: an integrated neurobiological and psychological perspective. *Nat. Rev. Neurosci.* **14**, 488–501 (2013).
10. Balleine, B. W. & Dickinson, A. Goal-directed instrumental action: contingency and incentive learning and their cortical substrates. *Neuropharmacology* **37**, 407–419 (1998).
This paper presents a clear statement of the difference between action and habit, and their neural substrates, in appetitive conditioning.
11. Bach, D. R. & Dayan, P. Algorithms for survival: a comparative perspective on emotions. *Nat. Rev. Neurosci.* **18**, 311–319 (2017).
12. Daw, N. D., Niv, Y. & Dayan, P. Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nat. Neurosci.* **8**, 1704–1711 (2005).
13. Dickinson, A. Action and habits: the development of behavioural autonomy. *Phil. Trans. R. Soc. B Biol Sci.* **308**, 67–78 (1985).
14. Dickinson, A. in *Animal Learning and Cognition* (ed. Mackintosh, N. J.) 45–79 (Academic Press, 1994).
15. Mobbs, D. & Kim, J. J. Neuroethological studies of fear, anxiety, and risky decision-making in rodents and humans. *Curr. Opin. Behav. Sci.* **5**, 8–15 (2015).
This paper describes defensive behaviour in a neuroethological context.
16. Yeomans, J. S., Li, L., Scott, B. W. & Frankland, P. W. Tactile, acoustic and vestibular systems sum to elicit the startle reflex. *Neurosci. Biobehav Rev.* **26**, 1–11 (2002).

17. Leaton, R. N. & Cranney, J. Potentiation of the acoustic startle response by a conditioned stimulus paired with acoustic startle stimulus in rats. *J. Exp. Psychol. Anim. Behav. Process* **16**, 279–287 (1990).
18. Brown, J. S., Kalish, H. I. & Farber, I. E. Conditioned fear as revealed by magnitude of startle response to an auditory stimulus. *J. Exp. Psych.* **41**, 317–328 (1951).
19. Davis, M. in in *The Amygdala: Neurobiological Aspects of Emotion, Memory, and Mental Dysfunction* (ed. Aggleton, J. P.) 255–306 (Wiley-Liss, 1992).
20. Tinbergen, N. *The Study of Instinct* (Oxford Univ. Press, 1951).
21. Bolles, R. C. & Fanselow, M. S. A perceptual-defensive-recuperative model of fear and pain. *Behav. Brain Sci.* **3**, 291–323 (1980).
22. Blanchard, R. J. & Blanchard, D. C. Crouching as an index of fear. *J. Comp. Physiol. Psych* **67**, 370–375 (1969).
23. Gross, C. T. & Canteras, N. S. The many paths to fear. *Nat. Rev. Neurosci.* **13**, 651–658 (2012).
This is a cogent review comparing circuits underlying unlearned and learned defensive responses.
24. Silva, B. A., Gross, C. T. & Graff, J. The neural circuits of innate fear: detection, integration, action, and memorization. *Learn. Mem.* **23**, 544–555 (2016).
25. Rosen, J. B., Asok, A. & Chakraborty, T. The smell of fear: innate threat of 2,5-dihydro-2,4,5-trimethylthiazoline, a single molecule component of a predator odor. *Front. Neurosci.* **9**, 292 (2015).
26. Darwin, C. *The Expression of the Emotions in Man and Animals.* (Fontana Press, 1872).
27. Eibl-Eibesfeldt, I. & Sutterlin, C. in *Fear and Defense* (eds Brain, P. F., Parmigiani, S., Blanchard, R. & Mainardi, D.) 381–408 (Harwood, 1990).
28. Ekman, P. Facial expression and emotion. *Am. Psychol.* **48**, 384–392 (1993).
29. Ploog, D. Neurobiology of primate audio-vocal behavior. *Brain Res.* **228**, 35–61 (1981).
30. Hofer, M. A. Multiple regulators of ultrasonic vocalization in the infant rat. *Psychoneuroendocrinology* **21**, 203–217 (1996).
31. Blanchard, D. C., Griebel, G. & Blanchard, R. J. Mouse defensive behaviors: pharmacological and behavioral assays for anxiety and panic. *Neurosci. Biobehav. Rev.* **25**, 205–218 (2001).
32. Owings, D. H., Rowe, M. P. & Rundus, A. S. The rattling sound of rattlesnakes (Crotalus viridis) as a communicative resource for ground squirrels (Spermophilus beecheyi) and burrowing owls (Athene cunicularia). *J. Comp. Psychol.* **116**, 197–205 (2002).
33. Fanselow, M. S. & Lester, L. S. in *Evolution and Learning* (eds Bolles, R. C. & Beecher, M. D.) 185–211 (Erlbaum, 1988).
34. Bouton, M. E. & Bolles, R. C. Contextual control of the extinction of conditioned fear. *Learn. Motiv.* **10**, 445–466 (1979).
35. LeDoux, J. E. in *Handbook of Cognitive Neuroscience* (ed. Gazzaniga, M. S.) 357–368 (Plenum Publishing Corp., 1984).
36. Fadok, J. P. et al. A competitive inhibitory circuit for selection of active and passive fear responses. *Nature* **542**, 96–100 (2017).
37. LeDoux, J. E. *The Emotional Brain* (Simon and Schuster, 1996).
38. Janak, P. H. & Tye, K. M. From circuits to behaviour in the amygdala. *Nature* **517**, 284–292 (2015).
39. Hawkins, R. D. & Byrne, J. H. Associative learning in invertebrates. *Cold Spring Harb. Perspect. Biol.* **7**, a021709 (2015).
40. Anderson, D. J. & Adolphs, R. A. Framework for studying emotions across species. *Cell* **157**, 187–200 (2014).
41. Anderson, D. J. Circuit modules linking internal states and social behaviour in flies and mice. *Nat. Rev. Neurosci.* **17**, 692–704 (2016).
42. Thorndike, E. L. *The Elements of Psychology.* (The Mason-Henry Press, 1905).
43. Skinner, B. F. *The Behavior of Organisms: An Experimental Analysis.* (Appleton-Century-Crofts, 1938).
44. Hull, C. L. *Principles of Behavior.* (Appleton-Century-Crofts, 1943).
45. Dickinson, A. Associative learning and animal cognition. *Phil. Trans. R. Soc. B Biol Sci.* **367**, 2733–2742 (2012).
46. Herrnstein, R. J. & Hineline, P. N. Negative reinforcement as shock-frequency reduction. *J. Exp. Anal. Behav.* **9**, 421–430 (1966).
47. Miller, N. E. in *Handbook of Experimental Psychology* (ed. Stevens, S. S.) 435–472 (Wiley, 1951).
This article provides a classic description of the two-factor theory of fear in avoidance.
48. Overmier, J. B. & Brackbill, R. M. On the independence of stimulus evocation of fear and fear evocation of responses. *Behav. Res. Ther.* **15**, 51–56 (1977).
49. Boeke, E. A., Moscarello, J. M., LeDoux, J. E., Phelps, E. A. & Hartley, C. A. Active avoidance: neural mechanisms and attenuation of Pavlovian conditioned responding. *J. Neurosci.* **37**, 4808–4818 (2017).
50. Gillan, C. M. et al. Disruption in the balance between goal-directed behavior and habit learning in obsessive-compulsive disorder. *Am. J. Psychiatry* **168**, 718–726 (2011).
51. Dymond, S. & Roche, B. A contemporary behavior analysis of anxiety and avoidance. *Behav. Analyst* **32**, 7–27 (2009).
52. Delgado, M. R., Jou, R. L., LeDoux, J. E. & Phelps, E. A. Avoiding negative outcomes: tracking the mechanisms of avoidance learning in humans during fear conditioning. *Front. Behav. Neurosci.* **3**, 33 (2009).
53. Schlund, M. W., Hudgins, C. D., Magee, S. & Dymond, S. Neuroimaging the temporal dynamics of human avoidance to sustained threat. *Behav. Brain Res.* **257**, 148–155 (2013).
54. Collins, K. A., Mendelsohn, A., Cain, C. K. & Schiller, D. Taking action in the face of threat: neural synchronization predicts adaptive coping. *J. Neurosci.* **34**, 14733–14738 (2014).
This study showed that synchronization between the amygdala, striatum and medial prefrontal cortex predicts successful active coping with threats in humans.
55. Burguiere, E., Monteiro, P., Mallet, L., Feng, G. & Graybiel, A. M. Striatal circuits, habits, and implications for obsessive-compulsive disorder. *Curr. Opin. Neurobiol.* **30**, 59–65 (2015).
56. Packard, M. G. & Knowlton, B. J. Learning and memory functions of the basal ganglia. *Annu. Rev. Neurosci.* **25**, 563–593 (2002).
57. Adams, C. D. & Dickinson, A. Instrumental responding following reinforcer devaluation. *Quart. J. Exp. Psychol. Section B* **33**, 109–121 (1981).
58. Adams, C. D. Variations in the sensitivity of instrumental responding to reinforcer devaluation. *Quart. J. Exp. Psychol. Section B* **34**, 77–98 (1982).
59. Dezfouli, A. & Balleine, B. W. Habits, action sequences and reinforcement learning. *Eur. J. Neurosci.* **35**, 1036–1051 (2012).
60. Everitt, B. J. & Robbins, T. W. Neural systems of reinforcement for drug addiction: from actions to habits to compulsion. *Nat. Neurosci.* **8**, 1481–1489 (2005).
61. Everitt, B. J. & Robbins, T. W. From the ventral to the dorsal striatum: devolving views of their roles in drug addiction. *Neurosci. Biobehav. Rev.* **37**, 1946–1954 (2013).
62. Decker, J. H., Otto, A. R., Daw, N. D. & Hartley, C. A. From creatures of habit to goal-directed learners: tracking the developmental emergence of model-based reinforcement learning. *Psychol. Sci.* **27**, 848–858 (2016).
This paper presents a summary of the model-based versus model-free computational approach to actions and habits in humans.
63. Tolman, E. C. Cognitive maps in rats and men. *Psychol. Rev.* **55**, 189–208 (1948).
64. Packard, M. G. & McGaugh, J. L. Inactivation of hippocampus or caudate nucleus with lidocaine differentially affects expression of place and response learning. *Neurobiol. Learn. Mem.* **65**, 65–72 (1996).
65. Gibson, B. M. & Shettleworth, S. J. Place versus response learning revisited: tests of blocking on the radial maze. *Behav. Neurosci.* **119**, 567–586 (2005).
66. Yin, H. H., Knowlton, B. J. & Balleine, B. W. Inactivation of dorsolateral striatum enhances sensitivity to changes in the action-outcome contingency in instrumental conditioning. *Behav. Brain Res.* **166**, 189–196 (2006).
67. Gillan, C. M., Kosinski, M., Whelan, R., Phelps, E. A. & Daw, N. D. Characterizing a psychiatric symptom dimension related to deficits in goal-directed control. *eLife* **5**, e11305 (2016).
68. Mowrer, O. H. Two-factor learning theory: summary and comment. *Psychol. Rev.* **58**, 350–354 (1951).
69. Rescorla, R. A. & Solomon, R. L. Two process learning theory: relationships between Pavlovian conditioning and instrumental learning. *Psych. Rev.* **74**, 151–182 (1967).
70. Krypotos, A. M., Effting, M., Kindt, M. & Beckers, T. Avoidance learning: a review of theoretical models and recent developments. *Front. Behav. Neurosci.* **9**, 189 (2015).
71. Lengyel, M. & Dayan, P. in in *Advances in Neural Information Processing Systems 20* (eds Platt, J. C., Koller, D., Singer, Y. & Roweis, S.) 889–896 (MIT Press, 2007).
72. Gershman, S. J. & Daw, N. D. Reinforcement learning and episodic memory in humans and animals: an integrative framework. *Annu. Rev. Psychol.* **68**, 101–128 (2017).
73. Kumaran, D., Summerfield, J. J., Hassabis, D. & Maguire, E. A. Tracking the emergence of conceptual knowledge during human decision making. *Neuron* **63**, 889–901 (2009).
74. Balleine, B. W. & Dickinson, A. in *Consciousness and Human Identity* (ed. Cornwall, J.) 57–85 (Oxford Univ. Press, 1998).
75. Wimmer, G. E. & Shohamy, D. Preference by association: how memory mechanisms in the hippocampus bias decisions. *Science* **338**, 270–273 (2012).
76. Shohamy, D. & Wagner, A. D. Integrating memories in the human brain: hippocampal-midbrain encoding of overlapping events. *Neuron* **60**, 378–389 (2008).
77. LeDoux, J. E. & Brown, R. A higher-order theory of emotional consciousness. *Proc. Natl Acad. Sci. USA* **114**, E2016–E2025 (2017).
This paper proposes an extension of the higher-order theory of consciousness to emotional consciousness.
78. Baars, B. J. Global workspace theory of consciousness: toward a cognitive neuroscience of human experience. *Prog. Brain Res.* **150**, 45–53 (2005).
79. Baddeley, A. D. *Working memory, thought and action.* (Oxford Univ. Press, 2007).
80. Frith, C., Perry, R. & Lumer, E. The neural correlates of conscious experience: an experimental framework. *Trends Cogn. Sci.* **3**, 105–114 (1999).
81. Shallice, T. in in *Consciousness in contemporary science* (eds Marcel, A. & Bisiach, E.) 305–333 (Oxford Univ. Press, 1988).
82. Maia, T. V. & Cleeremans, A. Consciousness: converging insights from connectionist modeling and neuroscience. *Trends Cogn. Sci.* **9**, 397–404 (2005).
83. Soto, D. & Silvanto, J. Reappraising the relationship between working memory and conscious awareness. *Trends Cogn. Sci.* **18**, 520–525 (2014).
84. Bergstrom, F. & Eriksson, J. Maintenance of non-consciously presented information engages the prefrontal cortex. *Front. Hum. Neurosci.* **8**, 938 (2014).
This article presents evidence demonstrating nonconscious aspects of working memory.
85. Pan, Y., Lin, B., Zhao, Y. & Soto, D. Working memory biasing of visual perception without awareness. *Atten. Percept. Psychophys.* **76**, 2051–2062 (2014).
86. Eriksson, J., Vogel, E. K., Lansner, A., Bergstrom, F. & Nyberg, L. Neurocognitive architecture of working memory. *Neuron* **88**, 33–46 (2015).
87. Jacob, J., Jacobs, C. & Silvanto, J. Attention, working memory, and phenomenal experience of WM content: memory levels determined by different types of top-down modulation. *Front. Psychol.* **6**, 1603 (2015).
88. Trubutschek, D. et al. A theory of working memory without consciousness or sustained activity. *eLife* **6**, e23871 (2017).
89. Heyes, C. Blackboxing: social learning strategies and cultural evolution. *Phil. Trans. R. Soc. B Biol Sci.* **371**, 20150369 (2016).
90. Clayton, N. S., Griffiths, D. P., Emery, N. J. & Dickinson, A. Elements of episodic-like memory in animals. *Phil. Trans. R. Soc. B Biol Sci.* **356**, 1483–1491 (2001).
91. Kornell, N. Where is the "meta" in animal metacognition? *J. Comp. Psychol.* **128**, 143–149 (2014).
92. Smith, J. D., Couchman, J. J. & Beran, M. J. The highs and lows of theoretical interpretation in animal-metacognition research. *Phil. Trans. R. Soc. B Biol Sci.* **367**, 1297–1309 (2012).
93. Smith, J. D., Couchman, J. J. & Beran, M. J. Animal metacognition: a tale of two comparative psychologies. *J. Comp. Psychol.* **128**, 115–131 (2014).
94. Raby, C. R., Alexis, D. M., Dickinson, A. & Clayton, N. S. Planning for the future by western scrub-jays. *Nature* **445**, 919–921 (2007).
95. Shettleworth, S. J. Clever animals and killjoy explanations in comparative psychology. *Trends Cogn. Sci.* **14**, 477–481 (2010).

96. Suddendorf, T. & Corballis, M. C. Behavioural evidence for mental time travel in nonhuman animals. *Behav. Brain Res.* **215**, 292–298 (2010).
97. Heyes, C. Animal mindreading: what's the problem? *Psychon Bull. Rev.* **22**, 313–327 (2015).
98. Lashley, K. in *Cerebral Mechanisms in Behavior* (ed. Jeffers, L. A.) (Wiley, 1950).
99. Gomez-Nieto, R. et al. Origin and function of short-latency inputs to the neural substrates underlying the acoustic startle reflex. *Front. Neurosci.* **8**, 216 (2014).
100. Davis, M., Gendelman, D. S., Tischler, M. D. & Gendelman, P. M. A primary acoustic startle circuit: lesion and stimulation studies. *J. Neurosci.* **2**, 791–805 (1982).
101. Yeomans, J. S. & Frankland, P. W. The acoustic startle reflex: neurons and connections. *Brain Res. Brain Res. Rev.* **21**, 301–314 (1995).
102. Jordan, W. P. & Leaton, R. N. Startle habituation in rats after lesions in the brachium of the inferior colliculus. *Physiol. Behav.* **28**, 253–258 (1982).
103. Blanchard, D. C. & Blanchard, R. J. Innate and conditioned reactions to threat in rats with amygdaloid lesions. *J. Comp. Physiol. Psych.* **81**, 281–290 (1972).
104. Blanchard, R. J., Flannelly, K. J. & Blanchard, D. C. Defensive behavior of laboratory and wild Rattus norvegicus. *J. Comp. Psychol.* **100**, 101–107 (1986).
105. Rosen, J. B., Pagani, J. H., Rolla, K. L. & Davis, C. Analysis of behavioral constraints and the neuroanatomy of fear to the predator odor trimethylthiazoline: a model for animal phobias. *Neurosci. Biobehav. Rev.* **32**, 1267–1276 (2008).
106. Bouton, M. E. & Bolles, R. C. Conditioned fear assessed by freezing and by the suppression of three different baselines. *Animal Learn. Behav.* **8**, 429–434 (1980).
107. Kalin, N. H., Shelton, S. E. & Davidson, R. J. The role of the central nucleus of the amygdala in mediating fear and anxiety in the primate. *J. Neurosci.* **24**, 5506–5515 (2004).
108. Fanselow, M. S. & Poulos, A. M. The neuroscience of mammalian associative learning. *Annu. Rev. Psychol.* **56**, 207–234 (2005).
109. Johansen, J. P., Cain, C. K., Ostroff, L. E. & LeDoux, J. E. Molecular mechanisms of fear learning and memory. *Cell* **147**, 509–524 (2011).
This paper is a summary of the circuit, cellular and molecular mechanisms of Pavlovian aversive conditioning.
110. Pitkänen, A., Savander, V. & LeDoux, J. E. Organization of intra-amygdaloid circuitries in the rat: an emerging framework for understanding functions of the amygdala. *Trends Neurosci.* **20**, 517–523 (1997).
111. Amaral, D. G., Price, J. L., Pitkänen, A. & Carmichael, S. T. in in *The Amygdala: Neurobiological Aspects of Emotion, Memory, and Mental Dysfunction* (ed. Aggleton, J. P.) 1–66 (Wiley-Liss, 1992).
112. Sah, P., Westbrook, R. F. & Luthi, A. Fear conditioning and long-term potentiation in the amygdala: what really is the connection? *Ann. NY Acad. Sci.* **1129**, 88–95 (2008).
113. Sweatt, J. D. Neural plasticity and behavior — sixty years of conceptual advances. *J. Neurochem.* **139** (Suppl. 2), 179–199 (2016).
114. Keifer, O. P. Jr., Hurt, R. C., Ressler, K. J. & Marvar, P. J. The physiology of fear: reconceptualizing the role of the central amygdala in fear learning. *Physiology* **30**, 389–4014 (2015).
115. Bocchio, M., Nabavi, S. & Capogna, M. Synaptic plasticity, engrams, and network oscillations in amygdala circuits for storage and retrieval of emotional memories. *Neuron* **94**, 731–743 (2017).
116. Maren, S. Synaptic mechanisms of associative memory in the amygdala. *Neuron* **47**, 783–786 (2005).
117. Ciocchi, S. et al. Encoding of conditioned fear in central amygdala inhibitory circuits. *Nature* **468**, 277–282 (2010).
118. Haubensak, W. et al. Genetic dissection of an amygdala microcircuit that gates conditioned fear. *Nature* **468**, 270–276 (2010).
119. Grundemann, J. & Luthi, A. Ensemble coding in amygdala circuits for associative learning. *Curr. Opin. Neurobiol.* **35**, 200–206 (2015).
120. Smith, Y. & Pare, D. Intra-amygdaloid projections of the lateral nucleus in the cat: PHA-L anterograde labeling combined with postembedding GABA and glutamate immunocytochemistry. *J. Comp. Neurol.* **342**, 232–248 (1994).
121. Li, H. et al. Experience-dependent modification of a central amygdala fear circuit. *Nat. Neurosci.* **16**, 332–339 (2013).
122. Penzo, M. A., Robert, V. & Li, B. Fear conditioning potentiates synaptic transmission onto long-range projection neurons in the lateral subdivision of central amygdala. *J. Neurosci.* **34**, 2432–2437 (2014).
123. Penzo, M. A. et al. The paraventricular thalamus controls a central amygdala fear circuit. *Nature* **519**, 455–459 (2015).
124. Yu, K., Garcia da Silva, P., Albeanu, D. F. & Li, B. Central amygdala somatostatin neurons gate passive and active defensive behaviors. *J. Neurosci.* **36**, 6488–6496 (2016).
125. Sanford, C. A. et al. A central amygdala CRF circuit facilitates learning about weak threats. *Neuron* **93**, 164–178 (2017).
126. LeDoux, J. E., Iwata, J., Cicchetti, P. & Reis, D. J. Different projections of the central amygdaloid nucleus mediate autonomic and behavioral correlates of conditioned fear. *J. Neurosci.* **8**, 2517–2529 (1988).
127. Fanselow, M. S., DeCola, J. P., De Oca, B. M. & Landeira-Fernandes, J. Ventral and dorsolateral regions of the midbrain periacqueductal gray (PAG) control different stages of defensive behavior: dorsolateral PAG lesions enhance the defensive freezing produced by massed and immediate shock. *Aggressive Behav.* **21**, 63–77 (1995).
128. Yu, K. et al. The central amygdala controls learning in the lateral amygdala. *Nat. Neurosci.* **20**, 1680–1685 (2017).
129. Shackman, A. J. & Fox, A. S. Contributions of the central extended amygdala to fear and anxiety. *J. Neurosci.* **36**, 8050–8063 (2016).
130. Fox, A. S., Oler, J. A., Tromp, D. P., Fudge, J. L. & Kalin, N. H. Extending the amygdala in theories of threat processing. *Trends Neurosci.* **38**, 319–329 (2015).
131. Maren, S., Phan, K. L. & Liberzon, I. The contextual brain: implications for fear conditioning, extinction and psychopathology. *Nat. Rev. Neurosci.* **14**, 417–428 (2013).
132. Giustino, T. F. & Maren, S. The role of the medial prefrontal cortex in the conditioning and extinction of fear. *Front. Behav. Neurosci.* **9**, 298 (2015).
133. Do Monte, F. H., Quirk, G. J., Li, B. & Penzo, M. A. Retrieving fear memories, as time goes by. *Mol. Psychiatry* **21**, 1027–1036 (2016).
134. Sotres-Bayon, F. & Quirk, G. J. Prefrontal control of fear: more than just extinction. *Curr. Opin. Neurobiol.* **20**, 231–235 (2010).
135. Do-Monte, F. H., Quinones-Laracuente, K. & Quirk, G. J. A temporal shift in the circuits mediating retrieval of fear memory. *Nature* **519**, 460–463 (2015).
136. Walker, D. L. & Davis, M. Role of the extended amygdala in short-duration versus sustained fear: a tribute to Dr. Lennart Heimer. *Brain Struct. Funct.* **213**, 29–42 (2008).
137. Hammack, S. E., Todd, T. P., Kocho-Schellenberg, M. & Bouton, M. E. Role of the bed nucleus of the stria terminalis in the acquisition of contextual fear at long or short context-shock intervals. *Behav. Neurosci.* **129**, 673–678 (2015).
138. Kim, S. Y. et al. Diverging neural pathways assemble a behavioural state from separable features in anxiety. *Nature* **496**, 219–223 (2013).
139. Duvarci, S., Bauer, E. P. & Pare, D. The bed nucleus of the stria terminalis mediates inter-individual variations in anxiety and fear. *J. Neurosci.* **29**, 10357–10361 (2009).
140. Waddell, J., Morris, R. W. & Bouton, M. E. Effects of bed nucleus of the stria terminalis lesions on conditioned anxiety: aversive conditioning with long-duration conditional stimuli and reinstatement of extinguished fear. *Behav. Neurosci.* **120**, 324–336 (2006).
141. Hartley, C. A. & Phelps, E. A. Changing fear: the neurocircuitry of emotion regulation. *Neuropsychopharmacology* **35**, 136–146 (2010).
142. Phelps, E. A. & LeDoux, J. E. Contributions of the amygdala to emotion processing: from animal models to human behavior. *Neuron* **48**, 175–187 (2005).
143. Buchel, C. & Dolan, R. J. Classical fear conditioning in functional neuroimaging. *Curr. Opin. Neurobiol.* **10**, 219–223 (2000).
144. LaBar, K. S., LeDoux, J. E., Spencer, D. D. & Phelps, E. A. Impaired fear conditioning following unilateral temporal lobectomy in humans. *J. Neurosci.* **15**, 6846–6855 (1995).
145. Bechara, A. et al. Double dissociation of conditioning and declarative knowledge relative to the amygdala and hippocampus in humans. *Science* **269**, 1115–1118 (1995).
146. LaBar, K. S., Gatenby, J. C., Gore, J. C., LeDoux, J. E. & Phelps, E. A. Human amygdala activation during conditioned fear acquisition and extinction: a mixed-trial fMRI study. *Neuron* **20**, 937–945 (1998).
147. Morris, J. S., Ohman, A. & Dolan, R. J. Conscious and unconscious emotional learning in the human amygdala. *Nature* **393**, 467–470 (1998).
148. Mendez-Bertolo, C. et al. A fast pathway for fear in human amygdala. *Nat. Neurosci.* **19**, 1041–1049 (2016).
149. Luo, Q. et al. Emotional automaticity is a matter of timing. *J. Neurosci.* **30**, 5825–5829 (2010).
150. Lebow, M. A. & Chen, A. Overshadowed by the amygdala: the bed nucleus of the stria terminalis emerges as key to psychiatric disorders. *Mol. Psychiatry* **21**, 450–463 (2016).
151. Dolan, R. J. & Vuilleumier, P. Amygdala automaticity in emotional processing. *Ann. NY Acad. Sci.* **985**, 348–355 (2003).
152. Pourtois, G., Schettino, A. & Vuilleumier, P. Brain mechanisms for emotional influences on perception and attention: What is magic and what is not. *Biol. Psychol.* **92**, 492–512 (2013).
153. Ohman, A., Carlsson, K., Lundqvist, D. & Ingvar, M. On the unconscious subcortical origin of human fear. *Physiol. Behav.* **92**, 180–185 (2007).
154. Whalen, P. J. et al. Human amygdala responsivity to masked fearful eye whites. *Science* **306**, 2061 (2004).
155. Morris, J. S., Ohman, A. & Dolan, R. J. A subcortical pathway to the right amygdala mediating "unseen" fear. *Proc. Natl Acad. Sci. USA* **96**, 1680–1685 (1999).
156. Liddell, B. J. et al. A direct brainstem-amygdala-cortical 'alarm' system for subliminal signals of fear. *Neuroimage* **24**, 235–243 (2005).
157. Sarter, M. F. & Markowitsch, H. J. Involvement of the amygdala in learning and memory: a critical review, with emphasis on anatomical relations. *Behav. Neurosci.* **99**, 342–380 (1985).
158. Nader, K., Majidishad, P., Amorapanth, P. & LeDoux, J. E. Damage to the lateral and central, but not other, amygdaloid nuclei prevents the acquisition of auditory fear conditioning. *Learn. Mem.* **8**, 156–163 (2001).
159. Moscarello, J. M. & LeDoux, J. E. Active avoidance learning requires prefrontal suppression of amygdala-mediated defensive reactions. *J. Neurosci.* **33**, 3815–3823 (2013).
This paper demonstrates the contribution of the medial prefrontal cortex in switching from freezing to active avoidance.
160. Lazaro-Munoz, G., LeDoux, J. E. & Cain, C. K. Sidman instrumental avoidance initially depends on lateral and basal amygdala and is constrained by central amygdala-mediated Pavlovian processes. *Biol. Psychiatry* **67**, 1120–1127 (2010).
This paper demonstrates that with overtraining active avoidance comes to be amygdala-independent.
161. Choi, J. S., Cain, C. K. & LeDoux, J. E. The role of amygdala nuclei in the expression of auditory signaled two-way active avoidance in rats. *Learn. Mem.* **17**, 139–147 (2010).
162. Bravo-Rivera, C., Roman-Ortiz, C., Brignoni-Perez, E., Sotres-Bayon, F. & Quirk, G. J. Neural structures mediating expression and extinction of platform-mediated avoidance. *J. Neurosci.* **34**, 9736–9742 (2014).
This article demonstrates the role of the amygdala and ventral striatum in a novel active avoidance paradigm.
163. Maren, S., Poremba, A. & Gabriel, M. Basolateral amygdaloid multi-unit neuronal correlates of discriminative avoidance learning in rabbits. *Brain Res.* **549**, 311–316 (1991).
164. Poremba, A. & Gabriel, M. Amygdala neurons mediate acquisition but not maintenance of instrumental avoidance behavior in rabbits. *J. Neurosci.* **19**, 9635–9641 (1999).
165. Killcross, S., Robbins, T. W. & Everitt, B. J. Different types of fear-conditioned behaviour mediated by separate nuclei within amygdala. *Nature* **388**, 377–380 (1997).
166. Oleson, E. B., Gentry, R. N., Chioma, V. C. & Cheer, J. F. Subsecond dopamine release in the nucleus accumbens predicts conditioned punishment and its successful avoidance. *J. Neurosci.* **32**, 14804–14808 (2012).
167. Bravo-Rivera, C., Roman-Ortiz, C., Montesinos-Cartagena, M. & Quirk, G. J. Persistent active avoidance correlates with activity in prelimbic cortex and ventral striatum. *Front. Behav. Neurosci.* **9**, 184 (2015).

168. Ramirez, F., Moscarello, J. M., LeDoux, J. E. & Sears, R. M. Active avoidance requires a serial basal amygdala to nucleus accumbens shell circuit. *J. Neurosci.* **35**, 3470–3477 (2015).
This is an investigation of the role of connections from the BA to the NAcc in active avoidance using a disconnection approach.
169. Cain, C. K. & LeDoux, J. E. Escape from fear: a detailed behavioral analysis of two atypical responses reinforced by CS termination. *J. Exp. Psychol. Anim. Behav. Process* **33**, 451–463 (2007).
170. Amorapanth, P., LeDoux, J. E. & Nader, K. Different lateral amygdala outputs mediate reactions and actions elicited by a fear-arousing stimulus. *Nat. Neurosci.* **3**, 74–79 (2000).
This early study shows the differing contributions of the CeA and the BA to Pavlovian reactions and avoidance actions, respectively.
171. Moscarello, J. M. & LeDoux, J. Diverse effects of conditioned threat stimuli on behavior. *Cold Spring Harb. Symp. Quant. Biol.* **79**, 11–19 (2014).
172. Martinez, R. C. et al. Active versus reactive threat responding is associated with differential c-Fos expression in specific regions of amygdala and prefrontal cortex. *Learn. Mem.* **20**, 446–452 (2013).
173. LeDoux, J. E., Moscarello, J., Sears, R. & Campese, V. The birth, death and resurrection of avoidance: a reconceptualization of a troubled paradigm. *Mol. Psychiatry* **22**, 24–36 (2017).
This article summarizes how the concept of habit helps to solve controversies about avoidance that have plagued the field since the 1970s.
174. Lingawi, N. W. & Balleine, B. W. Amygdala central nucleus interacts with dorsolateral striatum to regulate the acquisition of habits. *J. Neurosci.* **32**, 1073–1081 (2012).
This paper presents evidence for interactions between the amygdala and the dorsolateral striatum in appetitive habit learning.
175. McDannald, M., Kerfoot, E., Gallagher, M. & Holland, P. C. Amygdala central nucleus function is necessary for learning but not expression of conditioned visual orienting. *Eur. J. Neurosci.* **20**, 240–248 (2004).
176. Corbit, L. H. & Balleine, B. W. Double dissociation of basolateral and central amygdala lesions on the general and outcome-specific forms of pavlovian-instrumental transfer. *J. Neurosci.* **25**, 962–970 (2005).
177. Mobbs, D. et al. When fear is near: threat imminence elicits prefrontal-periaqueductal gray shifts in humans. *Science* **317**, 1079–1083 (2007).
178. Seymour, B., Daw, N. D., Roiser, J. P., Dayan, P. & Dolan, R. Serotonin selectively modulates reward value in human decision-making. *J. Neurosci.* **32**, 5833–5842 (2012).
179. Balleine, B. W. & O'Doherty, J. P. Human and rodent homologies in action control: corticostriatal determinants of goal-directed and habitual action. *Neuropsychopharmacology* **35**, 48–69 (2010).
180. Solomon, R. L. & Wynne, L. C. Traumatic avoidance learning: the principles of anxiety conservation and partial irreversibility. *Psychol. Rev.* **61**, 353–385 (1954).
181. Kihlstrom, J. F. The cognitive unconscious. *Science* **237**, 1445–1452 (1987).
182. Radman, Z. *Before Consciousness: In Search of the Fundamentals of Mind.* (Imprint Academic, 2017).
183. Hassin, R. R., Uleman, J. S. & Bargh, J. A. *The New Unconscious.* (Oxford Univ. Press, 2005).
184. Wilson, T. D. *Strangers to Ourselves: Self-Insight and the Adaptive Unconscious.* (Harvard Univ. Press, 2002).
185. Banaji, M. R. & Greenwald, A. G. *Blind Spot: Hidden Biases of Good People.* (Bantam Books, 2016).
186. Corbit, L. H., Ostlund, S. B. & Balleine, B. W. Sensitivity to instrumental contingency degradation is mediated by the entorhinal cortex and its efferents via the dorsal hippocampus. *J. Neurosci.* **22**, 10976–10984 (2002).
187. Miller, K. J., Botvinick, M. M. & Brody, C. D. Dorsal hippocampus contributes to model-based planning. *Nat. Neurosci.* **20**, 1269–1276 (2017).
188. Dusek, J. A. & Eichenbaum, H. The hippocampus and memory for orderly stimulus relations. *Proc. Natl Acad. Sci. USA* **94**, 7109–7114 (1997).
189. Squire, L. *Memory and Brain* (Oxford, 1987).
190. Murray, E. A., Wise, S. P. & Graham, K. S. *The Evolution of Memory Systems: Ancestors, Anatomy, and Adaptations.* (Oxford Univ. Press, 2017).
191. Henke, K. A model for memory systems based on processing modes rather than consciousness. *Nat. Rev. Neurosci.* **11**, 523–532 (2010).
192. Eichenbaum, H. A cortical-hippocampal system for declarative memory. *Nat. Rev. Neurosci.* **1**, 41–50 (2000).
193. Myers, C. E. et al. Dissociating hippocampal versus basal ganglia contributions to learning and transfer. *J. Cogn. Neurosci.* **15**, 185–193 (2003).
194. Talk, A. C., Gandhi, C. C. & Matzel, L. D. Hippocampal function during behaviorally silent associative learning: dissociation of memory storage and expression. *Hippocampus* **12**, 648–656 (2002).
195. LeDoux, J. E. & Hofmann, S. G. The subjective experience of emotion: a fearful view. *Curr. Opin. Behav. Sci.* **19**, 67–72 (2018).
196. Koizumi, A., Mobbs, D. & Lau, H. Is fear perception special? Evidence at the decision-making and subjective confidence. *Soc. Cogn. Affect Neurosci.* **11**, 1772–1782 (2016).
197. Collins, A. G. & Frank, M. J. How much of reinforcement learning is working memory, not reinforcement learning? A behavioral, computational, and neurogenetic analysis. *Eur. J. Neurosci.* **35**, 1024–1035 (2012).
198. Otto, A. R., Gershman, S. J., Markman, A. B. & Daw, N. D. The curse of planning: dissecting multiple reinforcement-learning systems by taxing the central executive. *Psychol. Sci.* **24**, 751–761 (2013).
199. Insel, T. et al. Research domain criteria (RDoC): toward a new classification framework for research on mental disorders. *Am. J. Psychiatry* **167**, 748–751 (2010).
200. Mineka, S. & Zinbarg, R. A contemporary learning theory perspective on the etiology of anxiety disorders: it's not what you thought it was. *Am. Psychol.* **61**, 10–26 (2006).
201. Bouton, M. E., Mineka, S. & Barlow, D. H. A modern learning theory perspective on the etiology of panic disorder. *Psychol. Rev.* **108**, 4–32 (2001).
202. Craske, M. G. et al. Optimizing inhibitory learning during exposure therapy. *Behav. Res. Ther.* **46**, 5–27 (2008).
203. Craske, M. G., Treanor, M., Conway, C. C., Zbozinek, T. & Vervliet, B. Maximizing exposure therapy: an inhibitory learning approach. *Behav. Res. Ther.* **58**, 10–23 (2014).
204. Beck, A. T., Emery, G. & Greenberg, R. L. *Anxiety Disorders and Phobias: A Cognitive Perspective.* (Basic Books, 1985).
205. Huys, Q. J., Daw, N. D. & Dayan, P. Depression: a decision-theoretic analysis. *Annu. Rev. Neurosci.* **38**, 1–23 (2015).
206. Huys, Q. J. et al. Bonsai trees in your head: how the pavlovian system sculpts goal-directed choices by pruning decision trees. *PLoS Comput. Biol.* **8**, e1002410 (2012).
207. LeDoux, J. E. & Pine, D. S. Using Neuroscience to Help Understand Fear and Anxiety: A Two-System Framework. *Am. J. Psychiatry* **173**, 1083–1093 (2016).
208. LeDoux, J., Brown, R., Pine, D. S. & Hofmann, S. G. Know thyself: well-being and subjective experience. *Cerebrum* http://www.dana.org/Cerebrum/2018/Know_Thyself__Well_Being_and_Subjective_Experience/ (2018).
This article makes the case for a more generous view of the importance of subjective experience in the clinical treatment of anxiety disorders.
209. Mowrer, O. H. & Lamoreaux, R. R. Fear as an intervening variable in avoidance conditioning. *J. Comp. Psych.* **39**, 29–50 (1946).
This paper presents Mowrer's influential two-factor theory of avoidance, which greatly influenced the negative view of avoidance in clinical practice.
210. Bolles, R. C. Avoidance and escape learning: simultaneous acquisition of different responses. *J. Comp. Physiol. Psychol.* **68**, 355–358 (1969).
211. Bolles, R. C. in *The Psychology of Learning and Motivation* Vol. 6 (ed. Bower, G. H.) 97–145 (Academic Press, 1972).
This article is one of Bolles's critiques of the avoidance paradigm, which suppressed interest in this form of behaviour for decades.
212. Masterson, F. A. & Crawford, M. The defense motivation system: a theory of avoidance behavior. *Behav. Brain Sci.* **5**, 661–696 (1982).
213. Levis, D. J. in *Contemporary Learning Theories: Pavlovian Conditioning and the Status of Traditional Learning Theory* (eds Klein, S. B. & Mowrer, R. R.) 227–277 (Lawrence Erlbaum Assn, 1989).
214. McAllister, D. E. & McAllister, W. R. in *Fear, Avoidance, and Phobias: A Fundamental Analysis* (ed. Denny, M. R.) (Erlbaum, 1991).
215. McAllister, W. R. & McAllister, D. E. in *Aversive Conditioning and Learning* (ed. Brush, F. R.) 105–179 (Academic Press, 1971).
216. Cain, C. K. & LeDoux, J. E. in in *Handbook of Anxiety and Fear* (eds Blanchard, R. J., Blanchard, D. C., Griebel, G. & Nutt, D.) 103–124 (Academic Press, 2008).
217. Panksepp, J. *Affective Neuroscience.* (Oxford Univ. Press, 1998).
218. Panksepp, J., Fuchs, T. & Iacabucci, P. The basic neuroscience of emotional experiences in mammals: The case of subcortical FEAR circuitry and implications for clinical anxiety. *Appl. Animal Behav. Sci.* **129**, 1–17 (2011).
219. Adolphs, R. The biology of fear. *Curr. Biol.* **23**, R79–93 (2013).
220. Brown, J. S. & Farber, I. E. Emotions conceptualized as intervening variables — with suggestions toward a theory of frustration. *Psychol. Bull.* **48**, 465–495 (1951).
221. Rosen, J. B. & Schulkin, J. From normal fear to pathological anxiety. *Psychol. Rev.* **105**, 325–350 (1998).
222. Perusini, J. N. & Fanselow, M. S. Neurobehavioral perspectives on the distinction between fear and anxiety. *Learn. Mem.* **22**, 417–425 (2015).
223. Marx, M. H. Intervening variable or hypothetical construct? *Psychol. Rev.* **58**, 235–247 (1951).
224. LeDoux, J. E. Semantics, surplus meaning, and the science of fear. *Trends Cogn. Sci.* **21**, 303–306 (2017).
225. Barrett, L. F. *How Emotions are Made.* (Houghton Mifflin Harcourt, 2017).
226. Wegner, D. *The Illusion of Conscious Will.* (MIT Press, 2002).
227. Balleine, B. W., Killcross, A. S. & Dickinson, A. The effect of lesions of the basolateral amygdala on instrumental conditioning. *J. Neurosci.* **23**, 666–675 (2003).
228. Johnson, A. W., Gallagher, M. & Holland, P. C. The basolateral amygdala is critical to the expression of pavlovian and instrumental outcome-specific reinforcer devaluation effects. *J. Neurosci.* **29**, 696–704 (2009).
229. Gewirtz, J. C. & Davis, M. Second-order fear conditioning prevented by blocking NMDA receptors in amygdala. *Nature* **388**, 471–474 (1997).
230. Burns, L. H., Everitt, B. J. & Robbins, T. W. Effects of excitotoxic lesions of the basolateral amygdala on conditional discrimination learning with primary and conditioned reinforcement. *Behav. Brain Res.* **100**, 123–133 (1999).

Acknowledgements

The authors were supported by the US National Institutes of Health (grants R56DA029053 and R01MH38774 to J.L.) and the US Army Research Office (grant W911NF-16-1-0474 to N.D.D.).

Author contributions

J.L. and N.D.D. made substantial contributions to discussions of the content of the article, wrote the article and reviewed and/or edited the manuscript before submission.

Competing interests

The authors declare that there are no competing interests.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Reviewer information

Nature Reviews Neuroscience thanks C. Gross, D. Mobbs and the other anonymous reviewer(s), for their contribution to the peer review of this work.