Adapting to loss: A normative account of grief

Zack Dulberg (zdulberg@princeton.edu)
Princeton Neuroscience Institute, Princeton, NJ 08544

Rachit Dubey (rdubey@princeton.edu)
Department of Computer Science, Princeton University, Princeton NJ, 08544

Jonathan D. Cohen (jdc@Princeton.edu)
Princeton Neuroscience Institute, Princeton, NJ 08544

Abstract

Grief is a reaction to loss that is observed across human cultures and even in other species. While the particular expressions of grief vary significantly, universal aspects include experiences of emotional pain and frequent remembering of what was lost. Despite its prevalence, and its obvious nature, the normative value of grief is puzzling: *Why* do we grieve? Why is it *painful*? And why is it sometimes prolonged enough to be clinically impairing? Using the framework of reinforcement learning with memory replay, we offer answers to these questions and suggest that grief may have normative value with respect to reward maximization.

Keywords: grief; reinforcement learning; replay

Introduction

Grief is a common reaction to loss, typically including painful recollection of memories and depressed mood (Bonanno & Kaltman, 2001; Weiss, 2008). Improving our understanding of grief is important, given not only its prevalence, but also its potential for producing both mental (Nielsen, Christensen, Neergaard, Bidstrup, & Guldin, 2020) and physical (Fagundes & Wu, 2020) disability. The study of grief has spanned disciplines and even species (King, 2013). Classical theories of grief include: the "Five Stages" model, describing a sequence of emotional states (Kübler-Ross & Byock, 1969); attachment theory-based descriptions of painful bond breaking (Bowlby, 1979); and the "Dual Process" model, describing oscillation between loss processing and restoration (Schut, 1999). More recently, models informed by neurobiology and/or behavioural neuroscience have suggested that learning (O'connor & Seeley, 2022; Boddez, 2018), representational change (Shear & Shair, 2005), competitive (Békés, Roberts, & Németh, 2023) and cognitive-behavioral (Maccallum & Bryant, 2013) processes play central roles in grief, and may have identifiable neural correlates (Gündel, O'Connor, Littrell, Fort, & Lane, 2003).

Despite these efforts, formal computational models that capture the dynamics of grief remain scarce, making it challenging to quantify and predict individual variations in the grieving process. Here, we attempt to address this challenge by leveraging the framework of reinforcement learning (RL) (Sutton & Barto, 2018) to explain the cognitive and emotional shifts occurring during grief. We propose that the psychological pain experienced in grief functions as a form of 'inverse reward,' accelerating the update of value representations tied to the object of loss. This mechanism may explain why individuals painfully and repetitively focus on memories of the loss: such memory replay with re-labelled negative rewards serves to expedite the reconfiguration of the value landscape.

To explore this hypothesis, we construct an RL model in which an agent learns about several source of reward, after which some are removed. We explore how the agent can best respond to this loss. We identify the "grief response" with an operational definition of mood, and vary model parameters to examine how they trade-off with respect to the grief response and subsequent reward. Finally, we compare the behavior of the model to longitudinal data on individual differences in prolonged grief reactions, suggesting possible translational applications of our framework to clinical psychology and psychiatry.

Methods

Approach

Our hypothesis is that grieving reflects a rational learning process. To that end, our work adapts the computational framework of reinforcement learning (RL) to explore the possible learning advantages that grieving provides. RL considers learning in the context of a Markov decision process (MDP), such that at each time-step t, an agent perceives the state of the environment s_t and takes an action a_t, causing a transition to state s_{t+1} and receiving reward r_t. The agent tries to collect as much reward as it can, aiming to maximize the sum of discounted future rewards defined as G_t (Eqn. 1), where $\gamma \in [0,1]$ is the discount factor, a parameter reflecting the present value of future rewards.

$$G_t = r_t + \gamma r_{t+1} + \gamma^2 r_{t+2} ... = \sum_{t=0}^{\infty} \gamma^t r_t \tag{1}$$

The agent learns to estimate the expected return of action a in state s, defined as the action value $Q(s,a)$ (also referred to as the Q-value) as in Eqn. (2).

$$Q(s,a) = \mathbb{E}_\pi[G_t | s_t = s, a_t = a] \tag{2}$$

The temporal difference (TD) algorithm updates Q-values in state s after action a transitions to next state s' (having corresponding action values Q'), according to learning rate α, as in Eqn. 3 (arguments s, a dropped for clarity).

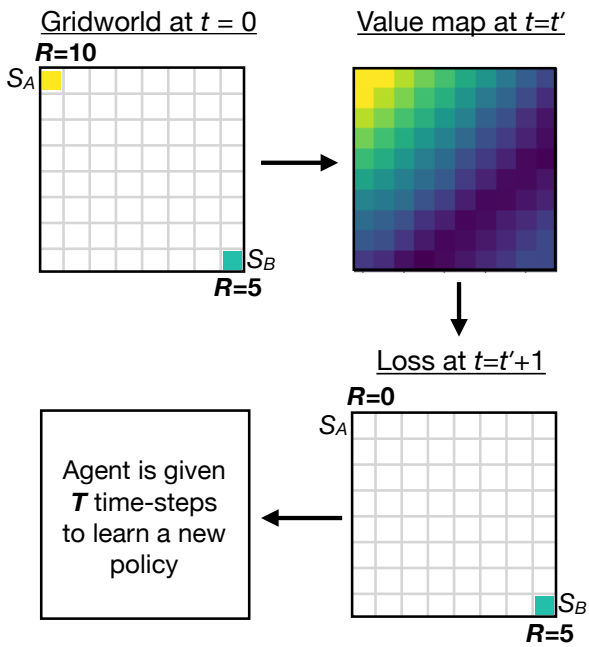

Figure 1: **Problem set-up.** Upper left: A 8x8 gridworld in which S_A provides a reward of 10 and S_B provides a reward of 5. Upper right: The learnt optimal value function. Lower right: Loss event where S_A is set to 0. Lower left: After the loss, the agent has T time-steps to learn a new policy and collect reward from S_B.

$$Q \leftarrow Q + \alpha(r + \gamma \max_{a'} Q' - Q) \tag{3}$$

Once the agent has learned accurate Q-values, the optimal behavioral policy selects the action with the highest Q-value in each state. Exploration can also be included, such that actions are selected randomly with probability ε, and otherwise 'greedily' as $\text{argmax}_a Q$.

TD-learning is known as a model-free strategy (Sutton & Barto, 2018), because value updates depend on current experiences and previously learned values rather than on a model of the environment used to explicitly project future values. Such gradual learning of cached values from experience is slow and inflexible. After a loss, it is this model-free value landscape we suggest grieving serves to re-map.

Problem Set-up

How might an adaptive agent respond to the loss of something valuable? As a first step toward answering this question, we consider a simple version of the problem, involving a NxN grid-world in which an agent can move in 4 cardinal directions. We label the grid location $(0,0)$ as S_A (state A), and the location (N,N) as S_B. These provide rewards of $r(S_A)$ and $r(S_B)$, such that $r(S_A)$ is more valuable than $r(S_B)$, with $r(s) = 0$ for all other states s (**Fig. 1**). We first allow the agent to learn the optimal policy in this environment, by taking random actions until its learned action values (Q-values) are stable. Then, $r(S_A)$ is set to 0, signifying a loss event. The agent

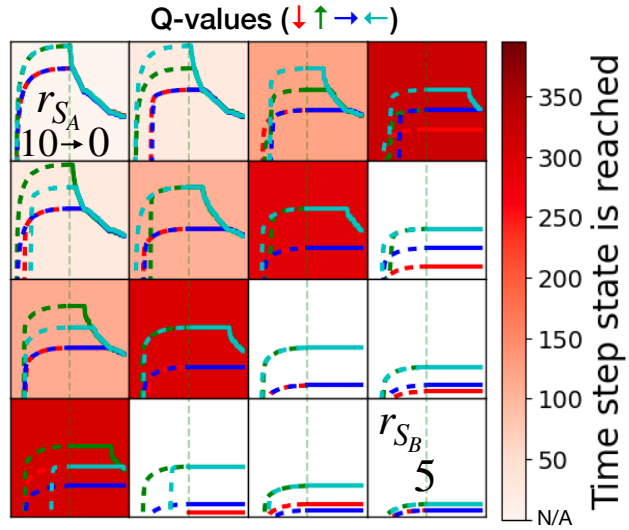

Figure 2: **TD-agent in simplified 4x4 grid.** Q-value evolution before and after loss event (vertical dashed line). The agent does not reach S_B within an allotted $T = 500$ time-steps.

is subsequently moved to S_A, and given T time-steps to maximize its reward. In analogy to the grieving process, the agent must 'get back to living,' by now seeking to collect reward from S_B after the loss of the opportunity for reward from S_A. In the sections that follow, we report simulations that explore various parameters relevant to grieving in the model.

Baseline agent: TD-learning

We begin with a baseline agent using TD-learning (TD-agent), and then explore ways in which it can improve its performance. Our default experiments use an 8x8 grid-world, starting with typical grid-world parameter settings: learning rate $\lambda = 0.1$, discount factor $\gamma = 0.95$, greedy action selection post-loss (i.e. $a = \text{argmax}_a Q$ with $\varepsilon = 0$), and $T = 5000$ steps to collect reward. One run of this agent is visualized in **Fig. 2** on a simplified 4x4 grid; it never reaches S_B, earning a total reward $R = \sum_{t=t'}^{T} r_t = 0$ (always calculated post-loss), because model-free TD-learning devalues states slowly as they are experienced, and is unable to adapt in the specified time horizon.

Grieving agent: Replaying re-labelled memories.

How can the learning process be accelerated? One obvious way is to increase the learning rate α, and another is to increase exploration (i.e. by increasing then annealing ε). These solutions are somewhat straightforward but face well known limitations (see Fig. 3a, top panel): increasing α is known to cause learning instability and interference (McCloskey & Cohen, 1989); similarly, increasing ε can actually worsen performance (since there is no 'short-cut' to unlearn the value of S_A, which continues to pull the agent away from S_B, unless ε reaches 0 precisely when the S_B neighborhood is entered).

A more sophisticated approach is to augment TD-learning

with a simple form of model-based learning referred to as DYNA (Sutton & Barto, 2018). DYNA can speed up model-free learning by performing TD-updates using past transitions sampled from a memory buffer, without having to re-experience them directly (called replay). But how should replay proceed after a loss? First, $r(S_A)$ should be altered in replayed transitions to reflect its loss (sometimes referred to as memory relabelling (Andrychowicz et al., 2017)). In the simplest case, old rewards in memory can be substituted for newly experienced ones (i.e. $r(S_A)^{new} = 0$). Here, we parameterize this choice with a variable called r_{grief}, such that $r(S_A)^{new} = r_{grief} r(S_A)^{old}$, aiming to accelerate the devaluation of S_A by, for example, setting $r_{grief} < 0$.

One problem with relabelling in the present setting is that values of $r_{grief} < 0$ may have little or no effect given the max operator in Eqn. 3, which assumes the best action is always taken, and therefore prevents the value of inferior actions from propagating. However, this is not a realistic assumption; worst-case scenarios matter when actions or state transitions are not deterministic. For instance, even if the edge of a cliff is attractive for the view it provides, its value should be mitigated by the risk of falling off it (e.g., from a gust of wind).

In order to model this, previous work introduced an optimism parameter w, which modifies TD-updates (Eqn. 4), maximizing optimism at $w = 1$ (reducing to Eqn. 3) and maximizing pessimism (i.e. sensitivity to worst-case scenarios) at $w = 0$ (Zorowitz, Momennejad, & Daw, 2020). Critically, w controls the propagation of low values, modulating the effect of r_{grief} in our simulations. We refer to this augmented agent as the **DYNA-Q_w-agent**, which performs updates using Eqn. 4 for both experienced and replayed transitions.

$$Q_w \leftarrow Q_w + \alpha(r + \gamma[w \max_{a'} Q'_w + (1-w) \min_{a'} Q'_w] - Q_w) \quad (4)$$

In order to capture when and how transitions are sampled from memory, we introduce two additional parameters. The first determines the number of transitions replayed per time-step in the environment, commonly called the replay ratio (*RR*) (Fedus et al., 2020). Insofar as replay takes time, a high *RR* can be considered a computational cost (Agrawal, Mattar, Cohen, & Daw, 2022). The second parameter determines the proportion of memories sampled for replay that pertain to loss-related transitions (i.e. propensity to dwell in grief), which we denote as p_{dwell}. During memory replay, we sample transitions in which S_A is the next-state with probability p_{dwell}, and sample transitions uniformly otherwise.

In sum, the DYNA-Q_w-agent, which we refer to as the **grieving-agent** for the rest of the paper, supplements TD-learning by using Eqn. 4 to update Q-values using 'painful' (i.e. re-labelled with r_{grief}) transitions sampled from a memory buffer, akin to an agent grieving about a loss.

Grief and mood disturbance

Since mood disturbance is a major factor in considerations of grief, we quantify it using reward prediction errors in the following way: when rewards are lower than expected, the agent is 'disappointed' and mood declines, and when they are higher than expected, it is 'pleasantly surprised' and mood improves. Eqn. 5 implements this process,

$$M \leftarrow M + \lambda(\delta - M), \quad (5)$$

where mood M is defined as the running accumulation (at rate λ) of prediction errors δ, as in prior work (Eldar, Rutledge, Dolan, & Niv, 2016), but with the addition of w in the δ term, shown in Eqn. 6 below:

$$\delta = \eta(r + \gamma[w \max_{a'} Q'_w + (1-w) \min_{a'} Q'_w] - Q_w) \quad (6)$$

Note that we include the term η, which reflects the differential effect on mood of remembered versus experienced events, by setting it to $\eta_{replay} \in (0,1)$ if the transition used to calculate the TD-error is replayed, and $(1-\eta_{replay})$ if the transition was currently experienced.

Finally, we quantify a performance metric C (for contentedness) that considers both total reward R and cumulative mood M (Eqn. 7), with a relative weighting factor $\eta_{mood} \in (0,1)$. Total reward and cumulative mood are normalized to comparable scales by dividing by their maximum values across simulations.

$$C = (1-\eta_{mood})R + \eta_{mood} \sum_t M_t \quad (7)$$

Experiments

We first calculate the total reward for a range of α, the number of time steps used to anneal ε from 1 to 0, and over a range of *RR* and r_{grief} with different settings of w. All agents use $\varepsilon = 0$, either directly after the loss event (grieving-agents) or after annealing (TD-agents). We also calculate the corresponding mood trajectories for a selected set of parameters.

Next, we consider how to rationally optimize the grieving process. For example, how painful should grief be? Given a computational budget (i.e. by fixing the replay ratio), is there a setting of r_{grief} that optimizes contentedness, balancing mood and reward? Similarly, how frequently should an agent dwell on their loss? In other words, what is the effect of p_{dwell} when considering different settings of η_{mood}? We then explore how identified optima change as other parameters are varied.

Finally, we explore the extent to which variation in r_{grief} may help explain clinical syndromes that can sometimes result from grief, such as prolonged grief disorder (PGD) (American Psychiatric Association, 2013).

Results

Behavior of learning agents

Fig. 3a displays performance of various agents as the total reward collected after a loss event. In the top panel, the TD-agent's performance (blue) is shown for different learning rates α, and as exploration is increased (shading). We

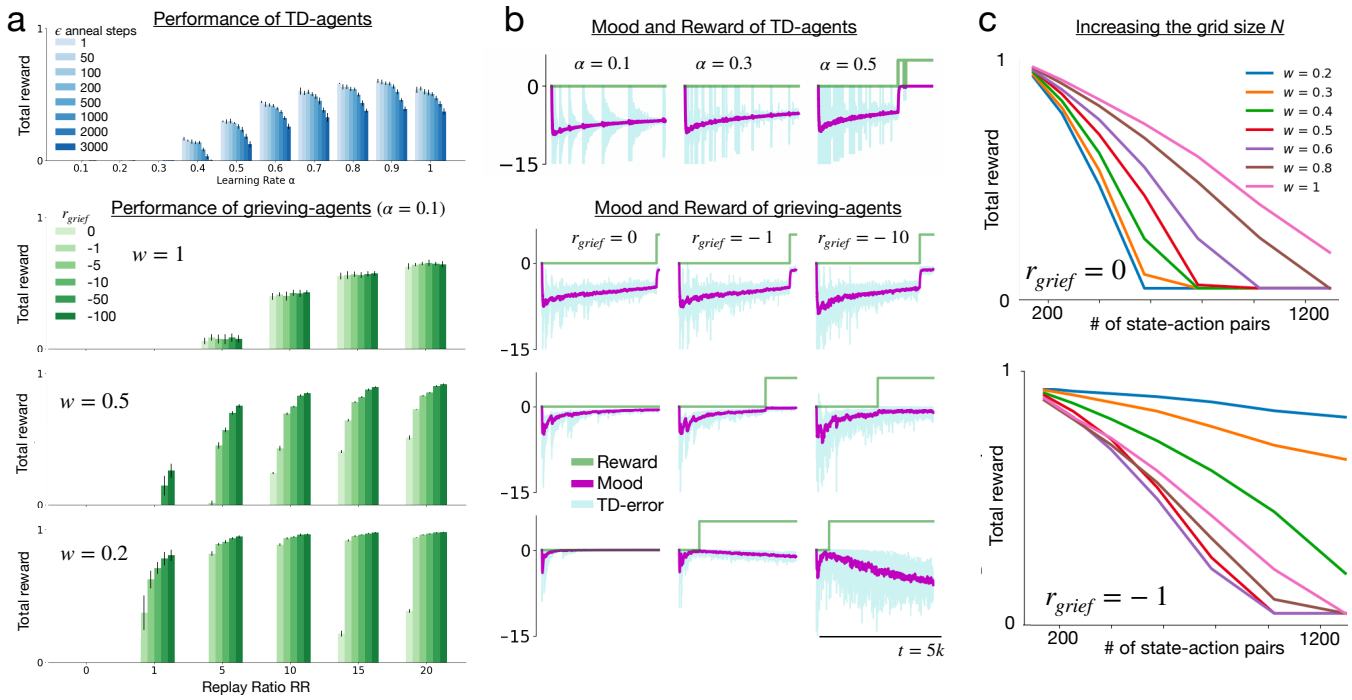

Figure 3: **Performance of the TD and grieving agents.** **a**: Total reward displayed for the TD-agent over a range of learning rates and annealing steps (blue). For the grieving-agents (green), we set $\alpha = 0.1$ and plot the total reward for $w \in [0.2, 0.5, 1]$ over a range of replay ratios and r_{grief}. Note that the bars show $\mu \pm \sigma$ over 10 runs. **b**: Post-loss trajectory of mood (purple) over $T = 5000$ time-steps ($RR = 5$, $\eta_{replay} = 0.5$, $\lambda = 0.1$) and reward (green) with w corresponding to adjacent panels in (a). Note that the replay ratio of the grieving-agents is set to 5. **c**: Total reward of the grieving-agents with increasing grid size N, for two r_{grief} settings and $RR = 10$. Total reward is normalized so that 1 is the highest achievable score (in both a and c).

observe that the total reward collected by the agent is 0 until α is increased to at least 0.4, growing gradually until $\alpha = 0.9$, whereas any amount of time spent exploring hurts performance. Thus, the TD-agent struggles to adapt to the loss in the environment.

Typically, RL uses $\alpha \leq 0.1$ to ensure stable learning. Therefore, in green, we display how the grieving-agent enhances performance with α fixed at 0.1. With $w = 1$ (top green panel), increasing RR improves total reward, but r_{grief} (shading) has no effect. As w decreases, lower r_{grief} increasingly improves total reward. For example, with $w = 0.5$, $r_{grief} = -10$, and $RR = 5$, total reward is approximately as high as with $\alpha = 0.9$. Thus, the grieving agent is better able to adapt to the loss event, approximating the optimal policy much faster.

Fig. 3b shows mood (purple), raw TD-error (light blue) and reward (green) trajectories after the loss event (arrows from panel **a** indicate corresponding bars). Interestingly, we note that there is an appearance of negative mood oscillations in the grieving agents. This is most prominent, for example, within the first $1k$ steps of the $w = 0.5, r_{grief} = -1$ trajectory (third row, middle column). We can speculate that these resemble 'waves of grief', which emerge in our simulations as states are successively re-encountered and devalued after loss. The overall extent of negative mood also depends on model parameters in a non-monotonic fashion, (i.e. both high and low w induce lower mood than $w = 0.5$ in the context of very low r_{grief}), which we will explore further in the next section.

Lastly, **Fig. 3c** displays performance of grieving-agents when the size of the grid-world is scaled up (measured by number of unique state-action pairs in the environment). While any amount of w hurts performance with $r_{grief} = 0$, the interaction of w with only $r_{grief} = -1$ strongly enhances robustness to grid scaling (for small w). This is an additional strength of the model, given real-world environments have many states.

Resource rational grieving

While exhaustive exploration of all model parameters is beyond the scope of this work, we now examine some informative sets of parameters that reveal optima with potential relevance to psychological phenomena.

Fig. 4a plots contentedness for different replay ratios and r_{grief}, with $w = 0.5$ and $\eta_{replay} = 0.5$ so that optimism and pessimism, and replays and experience are balanced *a priori*. We note that there is an optimal r_{grief} that balances reward and mood for each replay budget (the curved surface reaches a contentedness optimum for each replay ratio). **Fig. 4b** shows how the optimal r_{grief} changes as w and RR vary (considering $\eta_{mood} = 0.2$). A lower r_{grief} is needed with lower

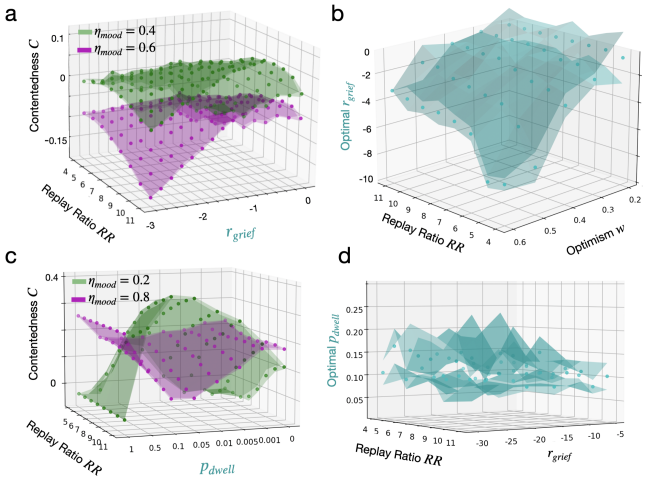

Figure 4: **Exploring parameter optima**. Solid points bounded by translucent surfaces indicate $\mu \pm \sigma$ over $N = 3$ runs. **a**: Contentedness C as a function of RR and r_{grief} for two different η_{mood}. **b**: r_{grief} that results in the maximum value of C (vertical axist) for different RR and w ($\eta_{mood} = 0.2$). **c**: Contentedness as a function of RR and p_{dwell} (semi-log scale) for two different η_{mood}. **b**: p_{dwell} that results in the maximum value of C (vertical axis) for different RR and w ($\eta_{mood} = 0$, i.e. reward component only).

RR, and with higher w. That is, with a lower replay budget, or lower sensitivity to negative rewards, a more negative r_{grief} is optimal.

Next, we investigate the parameter p_{dwell}, which reflects the fraction of loss memories replayed. **Fig. 4c** highlights an interesting property: the same value of p_{dwell} both maximizes reward and minimizes mood. This means that if reward maximization is the goal, then for an agent that is relatively insensitive to mood, replayed memories should be related to its loss around 10% of the time. However, for an agent that is sensitive to negative mood, this is the worst value of p_{dwell}. For the latter agent, avoiding (or persistent dwelling in) grief would both have a negative impact on life restoration, given that the p_{dwell} most effective for value re-mapping is also the most detrimental to mood. **Fig. 4b** shows how the p_{dwell} that maximizes total reward (i.e., considering $\eta_{mood} = 0$) remains around 0.1 even as r_{grief} and RR vary.

Accounting for prolonged grief

Prolonged grief disorder (PGD), defined as a condition in which recovery from grief is impaired, is typically associated with unabated emotional distress (Jordan & Litz, 2014). Here, we consider the extent to which the model introduced above may provide insight into how individual differences contribute to the development of PGD. Specifically, we consider how parameters of the model can be used to explore two dimensions of attachment style — 'compulsive care-seeking' (CCS), 'compulsive self-reliance' (CSR), and 'angry withdrawal' (AW) — that have been proposed to impact the development of PGD (Field & Sundin, 2001). This study reported that CCS-score is positively correlated with positive (+) loss-related thoughts (LRT) ($r = 0.34, p < 0.05$), and inversely correlated with negative (-) LRT ($r = -0.22$, n.s.). The opposite was observed for CSR-score, which was negatively correlated with +LRT ($r = -0.53, p < 0.001$) and positively correlated with -LRT ($r = 0.36, p < 0.05$), and for AW-score, which was less negatively correlated with +LRT ($r = -0.35, p < 0.05$) but more positively correlated with -LRT ($r = 0.42, p < 0.01$).

To model these different profiles, we sampled populations of CCS-agents with $r_{grief} \in [1,2]$ and $p_{dwell} = 0.02$, CSR-agents with $r_{grief} \in [0.1,-0.5]$and$p_{dwell} = 0.02$, and AW-agents with $r_{grief} \in [0.5,-1]$and$p_{dwell} = 0.05$. These reflected populations predominantly made up of individuals that replayed +LRT, -LRT, and more intense -LRT (in both intensity and frequency) respectively, while reflecting that the AW population should also have more intense +LRT than the CSR population.

Table 1 summarizes these data and parameters more compactly. **Fig. 5a** shows that this reproduces the empirical correlations measured between grief symptoms and CCS (blue), CSR (magenta) and AW (green) scores over time. We use negative mood $-M$ (eqn. 5) as a surrogate for intensity of grief symptoms, and plot corresponding correlations in **Fig. 5b**. While the correlation magnitudes are higher (presumably due to fewer sources of variance in our model compared to human data), the symptom trajectories are similar.

Table 1: Human data (Field & Sundin, 2001) and model parameters corresponding to different attachment styles

Style	Human data		Model parameters	
	+LRT cor.	-LRT cor.	r_{grief}	p_{dwell}
CCS	0.34^{*}	-0.22	$\in [1,2]$	0.02
CSR	-0.35^{*}	0.36^{*}	$\in [0.1,-0.5]$	0.02
AW	-0.53^{***}	0.42^{**}	$\in [0.5,-1]$	0.05

$^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$; two-tailed

The trajectories in **Fig. 5a** along with the data in **Table 1** suggest that replaying positive loss-related memories may prolong grief, while replaying negative ones may shorten it. Our model produced this effect using the following parameters: $RR = 10$, $w = 0.6$, $\alpha = 0.1$, $\lambda = 0.001$ and $\eta_{replay} = 0$. These were close to default parameters used in prior simulations, though it should be noted that it was experienced errors and not replayed errors that produced these particular trajectories (given $\eta_{replay} = 0$, i.e. r_{grief} affecting mood indirectly through value propagation rather than directly from replays). This is consistent with the possibility that individual differences exist in humans regarding the extent to which memories or experiences drive mood changes in grief.

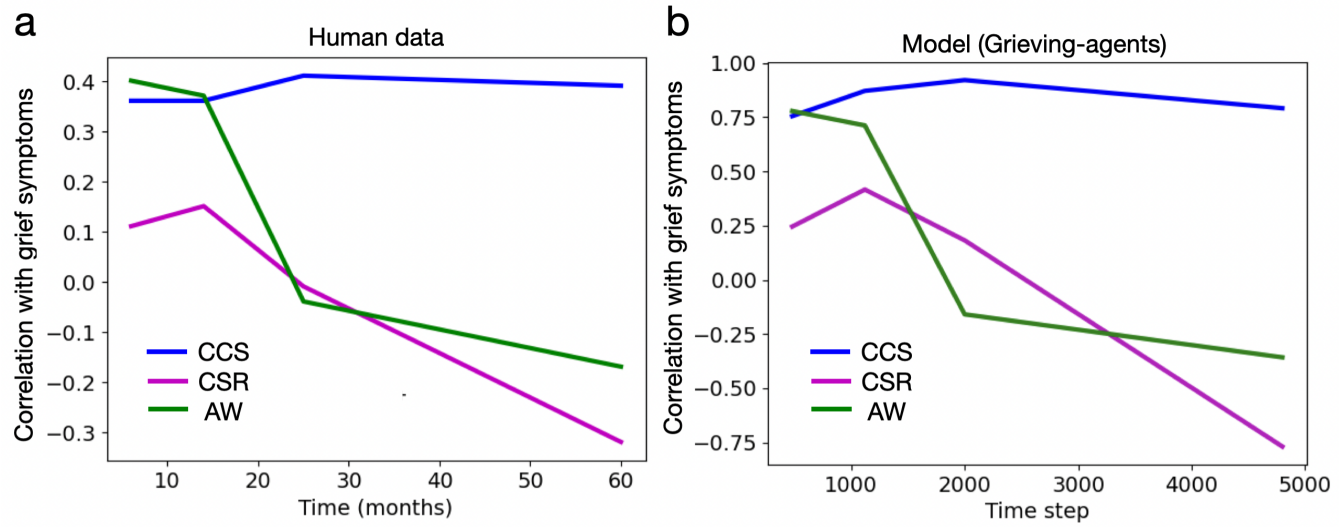

Figure 5: **Grief symptom trajectories. a**: Human data correlating scores on attachment-style dimensions (CCS: compulsive care-giving; CSR: compulsive self-reliance; AW: angry withdrawal) and grief symptom intensity over time (Field & Sundin, 2001). The blue trajectory (CCS) fits the pattern of a prolonged grief reaction, while magenta (CCR) and green (AW) adapt to the loss. **b**: Our model is capable of accounting for the same pattern, defining grief symptoms to be more intense when mood M (see eqn. 5) is more negative. The CCS agents replay positive memories of the loss ($r_{grief} \in [1,2]$, $p_{dwell} = 0.02$), the CSR agents replays slightly negative memories ($r_{grief} \in [0.1,-0.5]$, $p_{dwell} = 0.02$) while AW agents replay more negative memories, and to a lesser extent, more positive memories, more frequently ($r_{grief} \in [0.5,-1]$, $p_{dwell} = 0.05$). The correlation between $-M$ (symptoms) and r_{grief} is plotted at time steps with the same relative spacing as in **a**. We hypothesize that the valence, intensity and frequency of loss-related memories are the key factors through which attachment style modulates grieving (see main text).

Discussion

We have shown that augmenting a standard TD-agent with DYNA, Q_w-learning and memory re-labelling improves loss adaptation and may help explain features of human grieving. Specifically, this work suggests that grief may help to promote the learning of new model-free values after a loss. Grief is painful because negative relabelling of memories accelerates this learning. However, it can be prolonged if positive values are maintained for the object of loss. It has been previously suggested that internally generated negative rewards have computational benefits in accelerating the learning of TD-agents. (Dubey, Griffiths, & Dayan, 2022). Our work extends this idea to memory replay and relabelling, suggesting an additional function of generating negative rewards in the service of loss adaption.

Computational psychiatry aims to understand the shared principles underlying mental state dynamics cast in computationally explicit form, and to use this to quantify factors of variation that account for individual differences in them (Huys, Maia, & Frank, 2016). To that end, our results suggest ways in which measurable parameters might be expected to predict grief trajectories.

For example, we showed that the lower the time budget for recollection RR, and the lower the sensitivity to negative reward w, the more painful grieving should be. This might predict that those who are more easily cognitively fatigued (Agrawal et al., 2022), or those with a lower sensitivity to negative reward (Hundt et al., 2013) may actually experience more intense grief.

We also showed that there is a relatively robust setting of propensity to dwell in grief i.e., p_{dwell} $(= 0.1)$ that improves adaptation but worsens mood. This could potentially be measured as a treatment target, predicting that those with a value too low or too high would have a more prolonged course of grief. Indeed, one study found that the proportion of negative loss-related memories in prolonged grief disorder (PGD) patients was 0.39 ± 0.27 (Maccallum & Bryant, 2010), compared to 0.10 ± 0.13 in normal grievers (similar to the p_{dwell} optima we found).

As well, our simulations were motivated by the need to update model-free values after a loss. In principle, a fully model-based agent would not need to grieve its losses, and could instead simply delete them from its world-model. Given that individual differences in model-based vs. model-free decision making can be measured (Daw, Gershman, Seymour, Dayan, & Dolan, 2011), and have been shown to correlate with psychopathology (Voon et al., 2015), our work also predicts that individuals who rely less on model-free learning would experience less grief.

Lastly, we explored individual differences with respect to clinically-relevant grief symptomatology. We hypothesized that the previously reported modulation of grief trajectories by attachment style (Field & Sundin, 2001) is mediated by differences in the valence, intensity, and frequency of loss-

related memory replays. Our results comport with somewhat counter-intuitive observations that 'pleasurable reveries' can be prominent in PGD (O'Connor & Sussman, 2014), that more calming-feeling memories predict higher grief intensity (Boelen, Stroebe, Schut, & Zijerveld, 2006), and that yearning during grief (i.e., for high-value objects) is strongly connected to emotional pain in symptom networks (Malgaroli, Maccallum, & Bonanno, 2018). They are also consistent with neurobiological evidence that individuals with PGD have higher reward-related neural activity compared to typical grievers (O'Connor et al., 2008), as well as higher levels of oxytocin (Bui et al., 2019), which is known to increase attachment rewards (Scheele et al., 2013).

On this account, replaying positively valenced memories of the lost object maintains and increases its value, while the agent consequently experiences negative prediction errors in the post-loss environment - a process that could continue indefinitely. While this likely does not explain all cases of PGD, it may account for a particular sub-population, for whom novel pharmacological strategies could be suggested (for example, dopamine blocking agents that target reward circuitry directly). The ancient Greek tragedian Aeschylus provided a poetic summary of the present account: *'there is no pain so great as the memory of joy in present grief'*.

Future Work

This work can be extended in several directions. First, the parameter w, while interpretable, is a heuristic. A more realistic agent would take all Q-values into account, as well as specific uncertainties. For example, some actions or transitions may be more deterministic than others. Model-based RL with Bayesian updating or use of successor representations could allow for adaptation to these distributional properties of the environment. Next, memory replay was implemented as a fixed number of random (or loss-biased) samples on each step. Would similar optimal strategies for loss adaptation emerge if the memory replay policy itself was learned (Zha, Lai, Zhou, & Hu, 2019), and extended to accommodate memory relabelling?

Related to adaptive replay, one key component of grief we did not model was avoidance behavior, an additional major driver of PGD (Eisma & Stroebe, 2021). In our simulations, replayed rewards themselves did not contribute to the agent's experienced reward, except as an explanation for why grief is painful. However, interesting avoidance dynamics could emerge if a replay policy had to trade off present painful replays for future reward maximization. In that case, both avoiding painful replays and engaging in pleasurable replays (as previously discussed) could lead to prolonged grief, with their respective contributions accounting for individual differences in the syndrome.

Another direction would be to incorporate agents with prospective (i.e. planning) as well as retrospective replay, since positive counterfactual thoughts about the future is another known contributor to PGD (Golden & Dalgleish, 2012). Future work could also explore when partially retaining the value of lost objects would benefit generalization to related and still-accessible sources of reward, as achieving proper balance between retaining and relinquishing 'continuing-bonds' is another important function of grief (Hewson, Galbraith, Jones, & Heath, 2023).

Lastly, it will be important to scale our results up to more complex, ecologically valid environments. Modular deep Q-learning is a good candidate model for this (Dulberg, Dubey, Berwian, & Cohen, 2023), since modules maintain different memory buffers corresponding to separate reward components. This architecture could naturally accommodate memory re-labelling of distinct sources of value.

References

Agrawal, M., Mattar, M. G., Cohen, J. D., & Daw, N. D. (2022). The temporal dynamics of opportunity costs: A normative account of cognitive fatigue and boredom. *Psychological review*, *129*(3), 564.

American Psychiatric Association, f. (2013). *Diagnostic and statistical manual of mental disorders: Dsm-5* (Vol. 5) (No. 5). American psychiatric association Washington, DC.

Andrychowicz, M., Wolski, F., Ray, A., Schneider, J., Fong, R., Welinder, P., ... Zaremba, W. (2017). Hindsight experience replay. *Advances in neural information processing systems*, *30*.

Békés, V., Roberts, K., & Németh, D. (2023). Competitive neurocognitive processes following bereavement. *Brain Research Bulletin*, 110663.

Boddez, Y. (2018). The presence of your absence: A conditioning theory of grief. *Behaviour Research and Therapy*, *106*, 18–27.

Boelen, P. A., Stroebe, M. S., Schut, H. A., & Zijerveld, A. M. (2006). Continuing bonds and grief: A prospective analysis. *Death Studies*, *30*(8), 767–776.

Bonanno, G. A., & Kaltman, S. (2001). The varieties of grief experience. *Clinical psychology review*, *21*(5), 705–734.

Bowlby, J. (1979). The bowlby-ainsworth attachment theory. *Behavioral and Brain Sciences*, *2*(4), 637–638.

Bui, E., Hellberg, S. N., Hoeppner, S. S., Rosencrans, P., Young, A., Ross, R. A., ... Simon, N. M. (2019). Circulating levels of oxytocin may be elevated in complicated grief: a pilot study. *European Journal of Psychotraumatology*, *10*(1), 1646603.

Daw, N. D., Gershman, S. J., Seymour, B., Dayan, P., & Dolan, R. J. (2011). Model-based influences on humans' choices and striatal prediction errors. *Neuron*, *69*(6), 1204–1215.

Dubey, R., Griffiths, T. L., & Dayan, P. (2022). The pursuit of happiness: A reinforcement learning perspective on habituation and comparisons. *PLoS computational biology*, *18*(8), e1010316.

Dulberg, Z., Dubey, R., Berwian, I. M., & Cohen, J. D. (2023). Having multiple selves helps learning agents explore and adapt in complex changing worlds. *Pro-*

ceedings of the National Academy of Sciences, *120*(28), e2221180120.

Eisma, M. C., & Stroebe, M. S. (2021). Emotion regulatory strategies in complicated grief: A systematic review. *Behavior Therapy*, *52*(1), 234–249.

Eldar, E., Rutledge, R. B., Dolan, R. J., & Niv, Y. (2016). Mood as representation of momentum. *Trends in cognitive sciences*, *20*(1), 15–24.

Fagundes, C. P., & Wu, E. L. (2020). Matters of the heart: Grief, morbidity, and mortality. *Current directions in psychological science*, *29*(3), 235–241.

Fedus, W., Ramachandran, P., Agarwal, R., Bengio, Y., Larochelle, H., Rowland, M., & Dabney, W. (2020). Revisiting fundamentals of experience replay. In *International conference on machine learning* (pp. 3061–3071).

Field, N. P., & Sundin, E. C. (2001). Attachment style in adjustment to conjugal bereavement. *Journal of Social and Personal Relationships*, *18*(3), 347–361.

Golden, A.-M. J., & Dalgleish, T. (2012). Facets of pejorative self-processing in complicated grief. *Journal of consulting and clinical psychology*, *80*(3), 512.

Gündel, H., O'Connor, M.-F., Littrell, L., Fort, C., & Lane, R. D. (2003). Functional neuroanatomy of grief: an fmri study. *American Journal of Psychiatry*, *160*(11), 1946–1953.

Hewson, H., Galbraith, N., Jones, C., & Heath, G. (2023). The impact of continuing bonds following bereavement: A systematic review. *Death Studies*, 1–14.

Hundt, N. E., Brown, L. H., Kimbrel, N. A., Walsh, M. A., Nelson-Gray, R., & Kwapil, T. R. (2013). Reinforcement sensitivity theory predicts positive and negative affect in daily life. *Personality and Individual Differences*, *54*(3), 350–354.

Huys, Q. J., Maia, T. V., & Frank, M. J. (2016). Computational psychiatry as a bridge from neuroscience to clinical applications. *Nature neuroscience*, *19*(3), 404–413.

Jordan, A. H., & Litz, B. T. (2014). Prolonged grief disorder: diagnostic, assessment, and treatment considerations. *Professional Psychology: Research and Practice*, *45*(3), 180.

King, B. J. (2013). *How animals grieve*. University of Chicago Press.

Kübler-Ross, E., & Byock, I. (1969). On death and dying: What the dying have to teach doctors. *Nurses, Clergy and Their Own Families*, 133–150.

Maccallum, F., & Bryant, R. A. (2010). Impaired autobiographical memory in complicated grief. *Behaviour research and therapy*, *48*(4), 328–334.

Maccallum, F., & Bryant, R. A. (2013). A cognitive attachment model of prolonged grief: Integrating attachments, memory, and identity. *Clinical psychology review*, *33*(6), 713–727.

Malgaroli, M., Maccallum, F., & Bonanno, G. A. (2018). Symptoms of persistent complex bereavement disorder, depression, and ptsd in a conjugally bereaved sample: a network analysis. *Psychological medicine*, *48*(14), 2439–2448.

McCloskey, M., & Cohen, N. J. (1989). Catastrophic interference in connectionist networks: The sequential learning problem. In *Psychology of learning and motivation* (Vol. 24, pp. 109–165). Elsevier.

Nielsen, M. K., Christensen, K. S., Neergaard, M. A., Bidstrup, P. E., & Guldin, M.-B. (2020). Exploring functional impairment in light of prolonged grief disorder: a prospective, population-based cohort study. *Frontiers in Psychiatry*, *11*, 537674.

O'connor, M.-F., & Seeley, S. H. (2022). Grieving as a form of learning: Insights from neuroscience applied to grief and loss. *Current opinion in psychology*, *43*, 317–322.

O'Connor, M.-F., & Sussman, T. J. (2014). Developing the yearning in situations of loss scale: Convergent and discriminant validity for bereavement, romantic breakup, and homesickness. *Death Studies*, *38*(7), 450–458.

O'Connor, M.-F., Wellisch, D. K., Stanton, A. L., Eisenberger, N. I., Irwin, M. R., & Lieberman, M. D. (2008). Craving love? enduring grief activates brain's reward center. *Neuroimage*, *42*(2), 969–972.

Scheele, D., Wille, A., Kendrick, K. M., Stoffel-Wagner, B., Becker, B., Güntürkün, O., … Hurlemann, R. (2013). Oxytocin enhances brain reward system responses in men viewing the face of their female partner. *Proceedings of the National Academy of Sciences*, *110*(50), 20308–20313.

Schut, H., Margaret Stroebe. (1999). The dual process model of coping with bereavement: Rationale and description. *Death studies*, *23*(3), 197–224.

Shear, K., & Shair, H. (2005). Attachment, loss, and complicated grief. *Developmental Psychobiology: The Journal of the International Society for Developmental Psychobiology*, *47*(3), 253–267.

Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction*. MIT press.

Voon, V., Derbyshire, K., Rück, C., Irvine, M. A., Worbe, Y., Enander, J., … others (2015). Disorders of compulsivity: a common bias towards learning habits. *Molecular psychiatry*, *20*(3), 345–352.

Weiss, R. S. (2008). The nature and causes of grief.

Zha, D., Lai, K.-H., Zhou, K., & Hu, X. (2019). Experience replay optimization. *arXiv preprint arXiv:1906.08387*.

Zorowitz, S., Momennejad, I., & Daw, N. D. (2020). Anxiety, avoidance, and sequential evaluation. *Computational Psychiatry (Cambridge, Mass.)*, *4*.