Regret from cognition to code

Alan Dix[1 [0000-0002-5242-7693]] and Genovefa Kefalidou[2 [0000-0002-2889-7564]]

[1] Computational Foundry, Swansea University, Swansea, Wales, UK
alan@hcibook.com
[2] School of Computing and Mathematical Sciences, University of Leicester, Leicester, UK
gk169@leicester.ac.uk
https://alandix.com/academic/papers/regret-2021/

Abstract. Regret seems like a very negative emotion, sometimes even debilitating. However, emotions usually have a purpose, in the case of regret to help us learn from past mistakes. In this paper we first present an informal cognitive account of the way regret is built from a wide range of both primitive and more sophisticated mental abilities. The story includes Skinner-level learning, imagination, emotion, and counter-factual reasoning. When it works well this system focuses attention on aspects of past events where a small difference in behaviour would have made a big difference in outcome – precisely the most important lessons to learn. The paper then takes elements of this cognitive account and creates a computational model, which can be applied in simple learning situations. We find that even this simplified model boosts machine learning reducing the number of required training samples by a factor of 3–10. This has theoretical implications in terms of understanding emotion and mechanisms that may cast light on related phenomena such as creativity and serendipity. It also has potential practical applications in improving machine leaning and maybe even alleviating dysfunctional regret.

Keywords: regret · cognitive model · emotion · machine learning.

1 Introduction

Regret seems such a negative emotion, worrying about what might have been rather than about what could be. It seems so maladaptive and at best some redundant extension of feelings that are worthwhile. However looking at it more deeply it turns out to not only be a well-adapted feeling, but one that demonstrates the rich interactions between different levels of cognition: rational thought, vivid imagination and basic animal conditioning. Particularly interesting is the role that quite complex assessments of probability plays in regret – the closer you were to averting a disaster but failed, the worse it seems! Furthermore, regret is associated with risk identification, risk-taking and prediction of potential outcomes. For example, making a life decision, which is quite risky (with high levels of uncertainty), can induce potential regret from not making this life decision, which can be greater than from having made it.

The origins of the models presented in this paper date back to an exploratory essay [14] that led to the cognitive model described in section 2. This was followed by an early version of the computational model in section 3, which suggested potential positive learning effects but was only reported in informal talks. The current work formulates these models more thoroughly and systematically explores and evaluates the full range of parameters and options within the model allowing more rigorous and reliable analysis. Before proceeding to this, we will examine some of the psychological literature on regret.

1.1 The psychology of regret

Rationality and agency The ability to perform hypothetical comparisons (i.e. between an imagined state and a factual state) necessitate rational thinking and the capability to mentally represent these (e.g. *counterfactual thinking* – [33]) in defining *anticipated regret*), intention and risk-taking. Epstude and Roese [16] defined two different pathways as responsible for experiencing and perceiving regret. Information-based pathways directly affect intentions and consequently behaviour. On the other hand, content-neutral pathways can facilitate indirect effects from one's mind-set, positive and negative affect and motivational factors. Both pathways act as functional regulatory platforms for managing goals and 'controlling' behaviours within a socio-cognitive context. The foundational concept of the functional theory of counterfactual thinking is goal-setting and the comparison between current state and desired state.

At the same time, responsibility and agency appears to be critical in defining and experiencing regretful emotions. For example, Zeelenberg et al. [41], suggested that regret manifests -as an emotion- primarily to those that account themselves as responsible for a regretful action or interaction. A theory developed that reflects this is the Decision Justification Theory [13], according to which, decision-related regret is associated to comparative evaluation outcomes and 'self-blame'. Although self-blame can cause distress and negative emotions, experiencing regret has the potential to lead to better decision-making in the future.

Theories of regret A number of theories have been developed in the past aiming to define or model regret ranging from economic theories to socio-cognitive and interactional, including prospect theory [22] and regret theory [21]. According to Prospect theory, the losses and gains someone perceives are different depending on how these are formulated and on what types of affect generate. For example, if an investor is presented with two 'equal' options for investment opportunities, of which one is presented as associated to potential gains while the other one as associated to potential losses, the investor will tend to choose the former as an attempt to avoid the emotional impact that losses could cause (also known as "loss-aversion" theory). Two fundamental components of Prospect theory are the certainty (linked to probabilities) and the isolation effect (when outcomes are the same with the same probabilities, but where there are different routes or pathways to achieve these). In such cases, investors tend to follow ‘the known’ path or the one of ‘least resistance’ in an attempt to minimise their cognitive load. It is important to note that Prospect theory refers to pairs of choices; when the number of choice options increase, the complexity increases as well, due to additional interplaying factors. This is something that ‘regret theory’ [21] aimed to address. In regret theory, a key aspect is the so-called ‘choiceless utility function’, which represents the consequential state one experiences if no specific choice is made. In effect, utility, in this context, is associated with the psychological and human–computer interaction (HCI) notions of pleasurability and desirability. In regret theory, an individual chooses amongst multiple options, aiming to maximise the so-called “expected modified utility”. Loomes and Sugden [21] posit that independently of whether someone experiences regret or not, they will attempt to maximise the “expected modified utility” when making decisions under uncertainty.

Recapitulation and reflection Prospect and regret theories focus on prospects and probabilities estimations. In contrast, Norm theory [23] includes forward-looking recapitulation of past events based on prior (or even currently experienced) ‘norms’. Such ‘norms’ can be very personal and specific for different people suggesting individual differences in perception. Regret in such cases is dependent on memories and the capability to recall and process these (e.g. through mental simulation that can involve decision-making heuristics and biases).

The notion of mental simulation (and its role in perceiving regret) is further linked with the concept of ‘mental models’ whereby mental representations (or ‘mappings’) of the world facilitate a network of interconnected pieces of information that provide the ground-basis for reasoning and inferences generation [9–11]. Other ‘retrospective’ accounts of understanding regret includes the Reflection Evaluation Model that supports that reflection and evaluation mechanisms intertwine to promote comparative judgements that involve social, counterfactual and temporal aspects [26, 27]. Reflections tend to be inherently experiential and in support of a quasi-realistic simulated state or scenario that could be considered as ‘true’ at a given moment. On the other hand, evaluations tend to be based on factual (and not fictional) past incidents that are then compared and assessed on the basis of the goals fulfilment.

Regret and human–computer interaction Regret has an emerging role within the HCI community. Recent research suggests that regretful behaviours (i.e. in the form of interactions) can encourage ‘remediation strategies’ such as deleting unwanted or regretful messages, an action that can in turn become a source of confusion, uncertainty and information gaps. This is something prominent within certain demographic groups (e.g. teenagers), often associated with perceptions of trust and privacy. In all cases, feelings of regret are operationalised on the basis of psychological developmental theories for learning such as Pavlovian classical conditioning (stimulus-based associations that support learning) [29, 12, 5], Skinnerian operant (or instrumental) conditioning (consequential-based behavioural learning) [35, 37], and Bandura's Social Learning Theory which posits that learning occurs by observation (Bobo doll experiment [4]). Indeed, psychological learning theories have been applied in the past in the design and evaluation of technologies, including robotic and automated systems (see e.g. Touretzky and Saksida's operant conditioning Skinnerbots [38] and [25]). Developmental research has also found that children's decision-making can be improved if regret is experienced [31, 28].

2 A Cognitive Model of Regret

This section presents a cognitive model of regret as an *adaptive learning mechanism* summarised from [14]. As regret is a complex emotion, the model builds incrementally from simpler leaning mechanisms each step of which has benefit in itself, as is necessary for plausible evolutionary development. The basic steps are: (a) an unpleasant effect is experienced; (b) potential actions that might have been causes are brought to mind; (c) a counter-factual assessment is made of how close the key actions were to averting or reducing the bad effect; (d) this modifies the emotional feeling of regret; (e) the image in memory of the past action is available simultaneously with the current (modified) emotion; (f) simple associative memories are then formed. As is evident this includes very basic associative learning, with emotion and imagination and even counterfactual reasoning. We will now look at these building blocks in more detail and see how they come together to form the emotion we call regret.

2.1 Underlying Cognitive Systems that Enable Regret

Associative learning. We begin with basic Pavlovian and Skinnerian behavioural conditioning (see Section 1.1) as it is present in all but the most basic animals. In simple associative learning, if you perform an action resulting in negatively perceived outcomes, then you learn not to repeat this action again. In effect, associative learning is helping to identify relationships between two or more stimuli (Pavlovian classical conditioning [5]). When associations are already learnt, if a condition repeats (e.g. in the form of stimuli-trigger), then feelings and experiences associated before to this stimuli re-emerge and give rise to aversion or compliance to perform a task or make a decision [32]. Figure 1 (left) illustrates this with the example of touching a sharp thorn and learning it is painful.

From reactions to foresight – proto-imagination. Basic associative learning requires near simultaneous presentation of action and consequence, for more complex learning some form of memory imagery is required. This is also needed for momentary forward planning, which may be its origin. Complex planning allows both re-active and pro-active behaviours for action and decision-making. More advanced conscious planning behaviours are part of meta-cognitive abilities related to socio-cultural and socio-cognitive activity [2] and are affected by

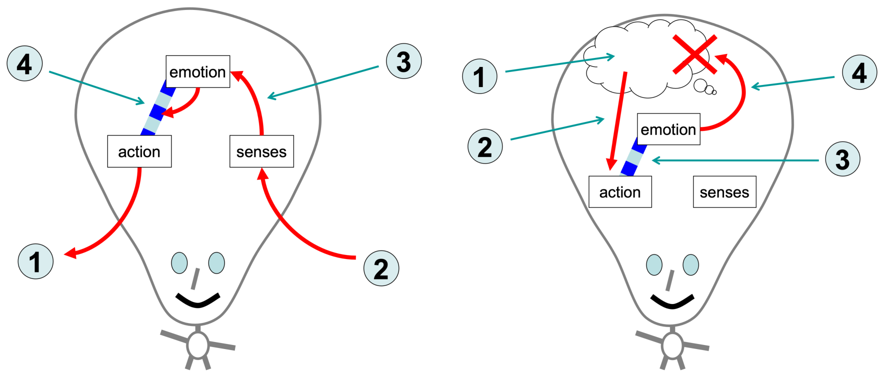

Fig. 1. (left) Simple associative memory: (1) touch thorn; (2) thorn pricks finger; (3) evaluation – Ow! it hurts! (4) learnt association – touching thorn is bad.
(right) Momentary planning: (1) prospective imagination of planned action; (2) causes similar brain activity to actually doing it! (3) learnt association fires; (4) veto of planned action.

attitudes, beliefs, motivations and goals, all of which contribute to imagination and forward-planning [19]. Indeed, adaptive and dynamic self-regulatory mechanisms are necessary to support scaffolding and planning both within short-term and long-term contexts [1]. However, there is a lower and more primitive level of foresight when, for example, one half-imagines what is going to happen as one reaches for a door handle and hence surprised if the door is jammed. At multiple levels we have predictive abilities that enable us to prepare to act even before sensing the world [8]. However, this momentary foresight still needs a level of proto-imagination as illustrated in Figure 1 (right), which leverages the way intention and action cause similar neural activity.

Dealing with delays. As noted, simple associative learning is attenuated by any delay between action and consequence (for a detailed review of the delay literature see [20]); such effects can also be associated with feedback and reinforcement loops and time retention [18, 36]. Indeed, effects of delays and impact on neural dynamics and learning have been explored before within the context of neural network simulations [3]. In reality incidents never happen absolutely simultaneously, but if brain activation decays slowly enough by the time a consequence occurs the areas associated with the last action may still be active enough to cause learning.

These learning effects reduce significantly once the delay is more than a few seconds. For simple creatures this means that learning of delayed consequences is all but impossible. More cognitively complex animals do appear to be able to learn delayed consequences using some form of recall, even without full human memory. Figure 2 (left) shows how this can occur bringing past related events into one's imagination and hence making past events and present consequences available for low-level associative learning. This then forms the basis for more complex learning in the interplay of memory, imagination and procedural skills [15] and later still to organising knowledge and structuring learning [40].

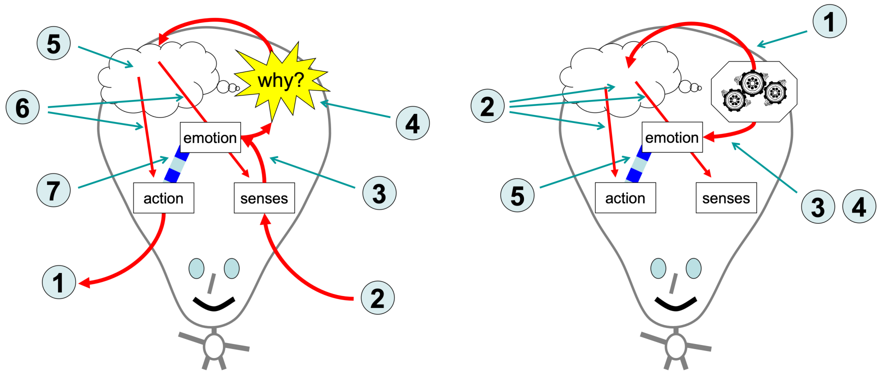

Fig. 2. (left) Dealing with delayed feedback: (1) touch unusual plant; (2) some hours later finger is sore; (3) evaluation – Ow that hurts! (4) desire to make sense; (5) recent salient events brought to mind – retrospective imagination; (6) causes simultaneous activation in relevant areas; (7) learnt association – don't touch that plant again!
(right) Regret reinforcing critical learning: (1) logical deduction of what mattered determines what is brought to mind; (2) imagination causes simultaneous activation in relevant areas; (3) causes negative emotion "if only I hadn't" ... regret; (4) counter factual deduction of how much it matters influences strength of emotion; (5) learnt association stronger or weaker depending on strength of emotion.

2.2 Regret as a Learning Mechanism

Attention in memory and counter-factual reasoning. A crucial element of association-making so far is the near simultaneity of imagined events. For example, recalling high school memories can evoke imagination processes whereby multiple separate events from that time intertwine and fuse in one go. For a sequence of events, because we focus on just a few highly salient events (e.g. touching a plant and feeling), we can replay potentially lengthy sequences in fast-forward and hence bring events close enough for learning to occur. However, if all past events are *equally* present, there cannot be effective learning. It is crucial that the most appropriate past events are remembered.

If through this more rational consideration of events we decide that particular actions or stimuli were not relevant to the good or bad consequences, these 'drop out' of the story of the events so that those that are recalled most strongly as we replay the events in our mind are precisely those which were part of the casual chain leading to the effect. In particular, in the case of negative effects, it is the things that if we 'only had not done' that are recalled – regret. Because of this regret we are able to associate negative emotions with the right actions, those that we would wish to avoid in future. This is shown in steps 1-3 in Figure 2 (right).

Note that this association is different from declarative knowledge that one might remember and think about logically later. Whilst the potential ill effects of a night on the town are something one can consider at leisure, in situations requiring fight or flight responses the difference between rapid emotional reactions and more rational consideration is crucial [24].

Boosting near misses. Regret has another adaptive mechanism – the tuning of the strength of emotional response depending on the probability that our actions were the principal cause (*credit assignment* in AI terms). Some of the actions we perform may have contributed to a bad effect, but we'd have had to do something very unusual or perhaps some other additional actions to avoid the effect. However, other actions are ones where we feel they could very nearly have been different and changed things. Steps 4 and 5 in Figure 2 illustrate this, as well as bringing relevant events to mind, the counter-factual "*what if*" reasoning amplifies the emotional response where a small change in behaviour could have made a large difference to outcomes.

For example, if you had bought a lottery ticket with the right number on it you would have won a million pounds, but you could easily have not got the right one – you don't feel too bad. However, imagine you almost chose to buy the lottery ticket with your birth date on, but decide not; later you find that the number came up – you are likely to feel more regret. Indeed research has suggested that regret aversion can partially be responsible for not exchanging lottery tickets even when there is a possibility of high material gain [39].

This lesser feeling of regret when things you did were less significant in the result and greater when what you did almost tipped the scales and made a difference is perfectly sensible. Higher emotional intensity can lead to higher levels of learning and stronger negative feelings (including jealousy and envy [42] as well as regret) attached to the action can affect actions and decisions to be made in future incidents.

Recapitulation. The final part of the learning armoury of regret is also the aspect that is most problematic in day-to-day life. Some events cause immediate regret, such as burning the toast, but are soon forgotten (perhaps due to the severity level of the incident or the impact value). However others, that have had especially large impact, become a repetitive rumination, which can be psychologically crippling, but, when not pathological, is also a learning mechanism. In machine learning it is common to use several copies of examples that are both rare and significant in order to improve learning. In a similar way the repeated exposure to the imagined events means that their learning impact is increased. In a situation, such as a near miss from being eaten by a wild animal, this is clearly critical for survival as we do not want to have to repeat such experiences in order to learn.

2.3 Cognitive Model Summary

In summary, regret is a subtle and well-adapted mechanism that enables us to learn effectively from the past recruiting deep (evolutionarily old) mechanisms. In particular, although regret allows us to manage "*if only*" statements, the mechanisms do not deal with more complex modalities such as "*if only but I couldn't have known*", or "*if only, but it will never happen again*". In effect, certain contextual parameters are not considered when processing (or experiencing) regret. Whilst we may be able to do the reasoning for these (although the former seems to elude many), these are not able to mollify the emotional reactions of regret.

It is interesting to note that regret is often considered purely as a negative emotion. Indeed there no single word for the positive equivalent of regret "it worked but only because"? Empirical studies in economics and psychology show that humans have a tendency to weigh negative results more strongly than positive ones, perhaps because 'in the wild' not learning to avoid bad things may kill you whereas missing good things simply means you have to try another time. As an adaptive mechanism regret shows that not only are negative effects stronger, but that we have additional mechanisms for negative emotion that may not exist for their positive counterparts.

3 A Computational Model of Regret

Aspects of the cognitive model have been built into a computational model. This was initially intended solely as a means of exploring and validating the cognitive model. For this, the core question was "*does regret aid learning*", and this will be addressed by looking at two metrics:

- ***asymptotic score*** – does it do better in the long run
- ***rate of learning*** – does it get to the same score with less exposures

The former is the obvious quality metric, but the second is critical in real-life situations where experience is precious: for early humankind experiments could be fatal. The computational model is not expected to mimic real human data as human learning typically involves multiple simultaneous mechanisms. The intention instead is to explore the plausibility of the cognitive model, by evaluating the efficacy of the cognitively inspired computational model.

While these initial aims are about cognitive understanding, we will see that the results also show promise as a technique to boost machine learning efficiency, especially in contexts when obtaining learning examples is costly. Computational experiments may not risk immediate death, they still consume time and energy, contributing to global carbon emissions and ultimately cataclysmic climate change.

3.1 Overall Architecture

Figure 3 shows the overall architecture of the model. The machine learning module interacts with an environment, which in our experiments is a simplified variant of the card game Pontoon. The machine learning module has two main parts a basic learning component and the regret module. The architecture is designed so that the regret module is loosely coupled and can be added to different forms of underlying learner and environment.

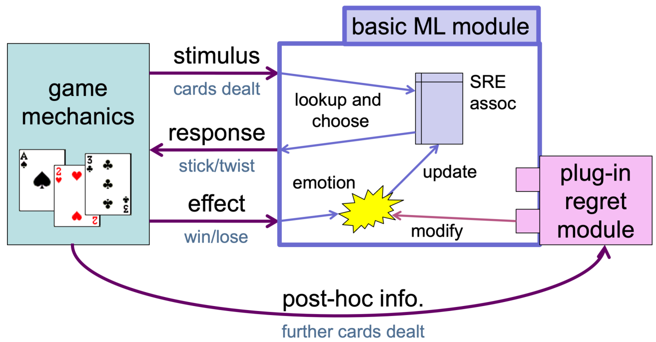

Fig. 3. Computational model architecture.

3.2 Example Game – Simple Pontoon

A simple version of Pontoon was chosen for these experiments as it is stochastic, offering challenge to the learner, but also simple enough to have perfect posterior knowledge, allowing very simple counter-factual evaluation.

Pontoon, or *vingt-un* is played with a normal pack of 52 cards. The player(s) and banker receive two cards each. Players may choose to 'twist' (receive an extra card) or 'stick' (keep the cards they have). The banker has a fixed rule for doing this: twist if the total is less than 17, stick otherwise. The aim is to have a hand with a higher total than the banker without going over 21 (called 'going bust'). The simplified version used in the experiments only has cards with values 1, 2 and 3 and the limit is 4 rather than 21 (so that a total of 5 or 6 is'bust'). The player and banker initially are dealt one card each and can have only one extra card. Furthermore, the extra cards are 'dealt' even if the player or banker stick to enable perfect posterior knowledge, but the additional cards are only counted in the score if the player/banker had chosen to twist before 'seeing' their card. Table 1 shows the rules for when the player wins or looses. This is then translated into a score of 10 for a win or -10 for a loss, but this may be modified by the regret module.

Because the game is so simple it is possible to exhaustively calculate all possible games (81 in total) and the expected winnings for each play. The best possible play strategy leads to winning just over one third of the time. Scoring a win as 10 and loss as -10, this gives an expected best possible score of -3.09 and corresponding worst possible score of -7.53. (Note that in the long-term the banker always wins, hence even the optimal strategy yields a negative expected score.) The actual scores from learning can be compared with these to see how close to optimal the learner becomes or normalised as a percentage of optimal gain by transforming a raw score of x into $100 \times (x + 7.53)/3.44$, so that the worst possible strategy corresponds to a normalised score of 0 and the optimal strategy is 100%. This normalised value is used in subsequent discussion and graphs.

Table 1. Simple Pontoon Rules (cards 1-3 only). Rules apply in order..

condition	outcome
player bust (> 4)	lose
banker bust (> 4)	win
player $>$ banker	win
player $=$ banker	lose
player $<$ banker	lose

3.3 The Environment

Although we use a specific game, this is presented to the machine learning component as a generic stimulus–response–effect environment (Fig. 4). When requested, the environment generates a new stimulus (`generateBefore` – the 'before' state); in the case of the Pontoon game, this is two cards (in the range 1–3), one each for the player and banker. It also provides a set of potential 'plays' (`getPlays`), possible actions that the player can choose take; in the case of Pontoon just 'stick' (keep the current card only) or 'twist' (have an additional card). The machine learning component chooses a play and the environment returns an 'after' state (`generateAfter`), which may depend on the chosen play, but also may involve stochastic elements; in the case of Pontoon a second card each for the player and banker. In some games, there is a clear arc of progress from start to finish, hence the `before` and `after` states are potentially of different types, but in other kinds of ongoing situations, these may be the same. Finally, an evaluation function (`evaluate`) gives a feedback score based on the before and after states and chosen play.

3.4 The basic learning component

The abstract interface of the basic learning module (Fig. 5) has two principal methods. The first, `getResponse`, takes a stimulus and set of possible responses

```
Before generateBefore();
Play[] getPlays( Before before );
public After generateAfter( Before before, Play play );
double evaluate( Before before, Play play, After after);
```

Fig. 4. Abstract interface of the game /environment component.

and returns a chosen response. The second, `condition`, takes an evaluation of the response (e.g. win or lose in Pontoon) and uses this to update its internal state.

The basic learning module used in experiments is a Skinner-like stimulus–response engine. This has an exhaustive table of all previous stimulus–response pairs and evaluation weight for each. For the simple learner the stimulus and response are completely opaque, simply used to look up or set relevant values; however a more complex learner, such as a neural network, might need a more detailed representation in order to generate generalised strategies.

When asked to provide a response the Skinner module consults its table to find the evaluation of each matching stimulus–response pair. Previously unseen pairs are given a default weight. Variations in this default weight alter the extent of novelty seeking vs. risk aversion of the learner. The weights of the possible responses are used to generate a probability and the module then stochastically selects a response. The probability distribution of the response is parametrised by a power value: a power of 1 giving a linear probability (response with weight 2 twice as likely as weight 1), a power of two using squared weights and a nominal 11 acts as a 'winner takes all' where the response with the highest weight is always chosen.

The update function simply adds the effect to the current weight of the stimulus–response pair, with non-linear Sigmoid applied to keep it within a +/-100 range.

```
Response getResponse( Stimulus stimulus, Response[] responseSet );
void     condition( Stimulus stimulus, Response response, double effect );
```

Fig. 5. Abstract interface of the learning component.

3.5 The Regret Module

The regret module is surprisingly simple, perhaps underlining how it builds on previous aspects of cognition. Figure 6 is the core function that implements regret. Without regret, the basic learning algorithm (`learner.condition`) would use the raw effect to modify the conditioning feedback. The counter-factual reason is embodied in the function `findBestResponse`. This uses posterior knowledge to predict what would have happened if any other play had been done and then returning the best possible result. In simple situations we may have perfect posterior knowledge, but in complex ones this may involve some form of uncertain or probabilistic inference. For example if you slip whilst rock climbing and are saved by the rope, you may factor in an element of "*but the rope might not have held*" even though you survived this time.

```
learn( stimulus, response, afterwards, effect):
    best = findBestResponse( stimulus, afterwards)
    regret = best - effect
    emotion = effect    //  equal to effect without regret
    IF ( effect >= 0 ) {
    THEN   emotion = effect * POS_NO_REGRET_FACTOR - regret * POS_REGRET_FACTOR
    ELSE  emotion = effect * NEG_NO_REGRET_FACTOR - regret * NEG_REGRET_FACTOR
    learner.condition(stimulus,response,emotion)
```

Fig. 6. Pseudocode for regret thinker. The only adaptation to the basic learning function is to modify the strength of positive or negative feedback.

The regret thinker has several adjustable parameters: two each for positive and negative outcomes. The `NO_REGRET` factors are about modifying the effect when the outcome was as good as it could be. The `REGRET` factors about modifying things when there was a better option (`regret` > `0`). These are separate factors because human emotional responses tend to be different to otherwise 'equal' positive and negative situations. For example, we might expect a 'no regret' situation to potential reduce the negative feelings for a negative outcome "*well I did the best I could*", but boost the positive feelings for a positive outcome.

The earliest code only had 'negative' regret, that is only the second arm of the '`if` statement when the initial effect was less than zero (a loss), as this corresponded to the day-to-day meaning of the term. However, the code looked 'messy' and hence, we experimented with adding the alternative and found that this boosted learning. Essentially this is a '*grass is greener*' emotion, for example if eating and enjoying a meal in a restaurant you might see someone else eating a different meal that look very tasty and then feel less happy about your own meal!

4 Experimental results

The underlying learner and regret engine have a significant number of parameters. We ran experiments over a wide range of configurations to avoid chance conclusions. We present typical examples, and summary views; the full data is available online at https://alandix.com/academic/papers/regret-2021/.

4.1 Obtaining Learning Saturation and Reducing Stochastic Noise

Both the basic Skinner-like learning and regret-enhanced learning make rapid initial gains in average game scores with the majority of learning gains over the first 1000 exposures. Learning slows but continues at a slower pace, so we have run all experiments to 10,000 iterations to examine asymptotic saturated learning. This approach to saturation is evident in the example run in Figure 7.

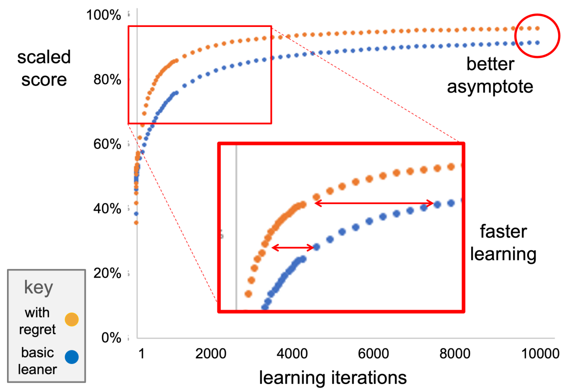

Fig. 7. Learning scaled to min/max possible scores showing better asymptotic values and faster learning. (linear weight, mild risk aversion, negative regret only).

Each run is stochastic, both in terms of the cards played and also the response of the low-level learning algorithm. Running over large numbers of iterations means that much of the randomness averages out as learning progresses, but substantial variation remains. The standard deviation of scores in the Pontoon game is approximately 1, 0.3 and 0.1 at 100, 1000 and 10000 iterations respectively. We therefore run each configuration 10000 times and created average learning traces with distribution statistics. This replication and averaging therefore means that the standard deviations of reported means are about 0.01, 0.003 and 0.001 at 100, 1000 and 10000 iterations respectively, or approx. 5%, 1.5% and 0.5% for normalised scores such as Figure 7. Note that the variation is still quite large early in learning until about 100 iterations, and this is evident in some of the results (e.g. Fig. 8). Detailed studies of early learning would need more replications, but we will confine ourselves to further along the learning process when a greater level of learning has been achieved.

4.2 Observed Behaviour

Figure 7 shows a typical run, in this case parameterised for linear weighting of the Skinner learner, mild risk aversion and negative regret only. The graph highlights two aspects. First is the *improvement in learning*: the scores at the end of this particular rule are 90.8% for the basic learner rising to 95.3% when regret is added. Both are still slowly rising with further learning even at 10000 iterations, so it is possible that the basic learning will eventually reach similar levels, but clearly only after vast numbers of learning steps. Second is *faster learning*: the regret learner obtains the same level of learning after fewer iterations. We look at both in more detail below.

While the exact numbers vary for different parameter configurations, the overall pattern is similar. For example, in the winner takes all, high risk averse, positive regret configuration, the asymptotic learning is closer to 100%, but there is still substantial improvement (97.3% for basic learning vs. 98.4% for regret). There is also consistent speedup of around 2.5 times faster (regret reaches 97.3% after only 3600 iterations).

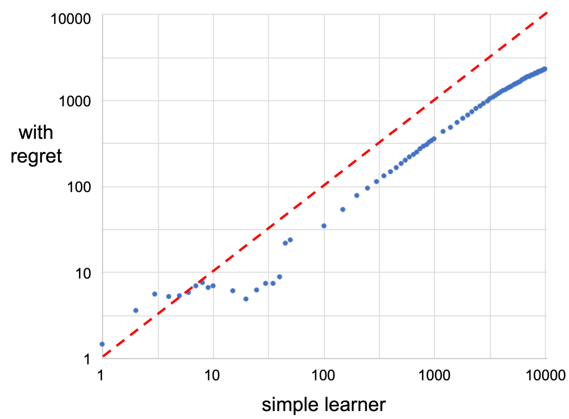

Fig. 8. Speed up, log-scale – number of regret exposures to reach same quality of learning as simple learner. Note higher variation for smaller numbers of trials. (linear weight, mild risk aversion, negative regret only).

Faster learning . Although this is evident in Figure 7, it is hard to assess the precise gain. Figure 8 represents the same data by plotting how long it takes the regret learner to obtain the same level of learning as the simple learner without regret. The lower part of the graph is quite noisy (even with 10,000 replications!), but the data is stable after about 100 iterations and clearly shows that adding regret substantially reduces the number of iterations required. In this case, the difference is about 0.5 on the log scale for much of the range, corresponding to a speed-up ratio (as measured by number of iterations) of just over 3. This is typical over the range of configurations, with speed-up ratios between 2.5 and 10 times in the central part of the range. Note too that the speed up increases towards the higher values in Figure 8. This is possibly because the saturation value is higher for regret (better asymptotic learning), so that an extended version of the graph would see the lower curve level off with the simple learner never reaching the same levels of learning.

Better learning. As noted, the example run in Figure 7 appears to show an improvement in asymptotic learning. This is observed across all configurations of parameters tested. Figure 9 compares learning of the basic learner compared to those with regret for a wide variety of parameter configurations for the Skinner-like learner and regret. The graph on the left shows performance after 1000 learning iterations and the right shows 10,000 learning iterations. Each dot shows the normalised average score over all 10,000 replications of the same parameter configuration. The x and y axes show the percentage of maximum score obtained by the simple learner and the same leaner with regret added. Note that the axes show different ranges on the left and right graphs as for both simple and regret learners the performance continues to improve over this range. The six vertical lines of dots on each graph are because there are six configurations of the simple learner and for each of these six alternative configurations of regret were added.

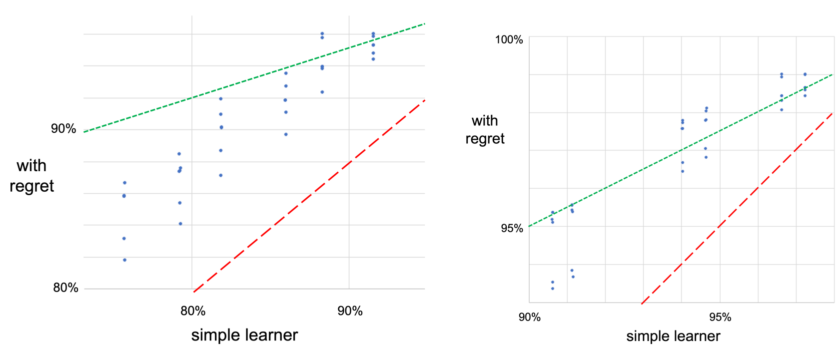

Fig. 9. Comparative scores after (left) 1000 iterations (right) 10,000 iterations.

The dashed line on each graph denotes equal learning. As is evident in every configuration of the simple learner regret improves performance. Some of the regret configurations improve it more than others, but all make substantial improvements. The dotted line shows the half-way point between the simple learner performance and perfect performance. As is evident, adding regret is close to or exceeds this mark, especially at higher levels of learning. That is, in many cases, regret halves the gap between the basic learning performance and perfect learning.

If one examines the cases in more detail, there are patterns amongst the various regret parametrisations. Positive regret on its own is not as effective as negative regret or both combined. However, we will not explore these in detail here as these differences may relate to the specific example game, where there was only win or lose, hence no way for an alternative action to have been better than a win. One point that is promising is that if anything the proportionate improvements (in the sense of how much they make the outcome nearer to optimal) are better with the more effective Skinner-learner configurations. Although we would be cautious in generalising this, it does suggest the boosting effect of regret is not limited to poor learners.

4.3 Computational Model Discussion

Our initial reason for creating a computational model of regret was to validate and explore properties of the cognitive model. However, we have seen that it also shows promise as a way to enhance other machine learning algorithms. Regret is often used as a *metric* within machine learning for both practical algorithms and as part of theoretical analysis (e.g. [6, 34, 43]). In addition, some algorithms such as Counterfactual Regret Minimization [44] and Deep Counterfactual Regret Minimization [7] work explicitly to minimise regret. However, the ability to 'bolt on' regret to existing algorithms does not appear to have been exploited previously.

Over recent years there has been growing interest in what has been termed 'human-like computation' [17, 30], not least in order to emulate the single-shot learning of higher-cognition compared to the vast number of exposures needed by sub-symbolic learning. Regret sits somewhere between the two achieving *fewer shot* learning and offering insights into the way higher- and lower-level cognition can work together. This is important as forms of hybrid learning are likely to be essential for next generation AI.

The computational model has also yielded some promising insights into regret itself, not least the importance of what we have called 'positive regret', that is the '*grass is greener*' effect after making even a decision with positive consequences. This can of course be problematic if one does not fully appreciate the positive things that happen, but it does improve learning, especially to help escape local maxima.

Finally, although not a new insight, the experimental results emphasise the methodological importance of (a) exploring the space of free-parameters and (b) running sufficient replications.

On the first of these, it is common in the machine learning literature to see papers that quote many fine-tuning parameters, such as network sizes or relaxation constants without any explanation of why the values were chosen or whether they are critical to the results. In early explorations of the regret model it appeared that positive regret was actually substantially more important than negative regret in terms of faster learning, however when a wider parameter-space was explored this effect was not found to be consistent, and restricted to 'winner takes all' low-level learning where local maxima are harder to escape. It would have been easy to publish these early results, which would have not only been misleading, but not exposed the underlying properties of positive regret.

On the second methodological point, many machine learning methods include stochastic or pseudo-random elements. The results are therefore also likely to have variability. The models used in this paper have very small numbers of learned weights (a few dozen) compared to many millions or billions in deep neural networks (DNNs). However, even so, many replications were needed in order to obtain statistically reliable results. The sizes of many DNNs makes this level of replication all but impossible, creating significant challenges in assessing sensitivity and reliability for safety critical applications and in interpreting results theoretically.

5 Conclusions and Future Work

The models presented offer both theoretical insight into regret as a human emotion and practical potential as a way to enhance machine leaning. Future work will explore both of these directions and also the way that understanding of regret and associated cognitive functions can be used as part of interactive applications that can help users to enhance the positive aspects and control the negative ones.

Future regret modelling may include mapping associative emotions to regret to create an ecological context whereby a richer (perhaps more enhanced and realistic) model of regret can be generated. In particular, we would like to explore the recapitulation aspect of regret that is not included in the current computational model and is critical in pathological regret. Given the dynamic nature of emotions (positive or negative), a next step would also include to explore more dynamic (or interactive) models of emotions that can be calibrated on-the-fly through different technological means. In that way platforms of 'in-the-wild' modelling that supports different layers of 'human-in-the-loop' interactions can be further designed. Indeed, there is an ongoing research on interactive Machine Learning approaches, utilised to provide 'verification' mechanisms to the data quality fed and to the generated outputs. Furthermore, the explorations and incorporation of positive emotion dynamics (even when modelling negative emotional responses) would be another direction to advance modelling approaches and techniques for experiential phenomena, acknowledging, in that way, the complexity and contextuality of human emotions.

References

1. Azevedo, R., Hadwin, A. F. Scaffolding self-regulated learning and metacognition–Implications for the design of computer-based scaffolds. Instructional Science, 33, 367–379 (2005).
2. Baker-Sennett, J., Matusov, E., Rogoff, B. Planning as developmental process. Advances in child development and behavior, 24, 253-281. (1993).
3. Baldi, P., Atiya, A.F. How delays affect neural dynamics and learning. IEEE Transactions on Neural Networks, 5(4), 612-621. (1994).
4. Bandura, A., Walters, R. H. Social learning theory (Vol. 1). Prentice Hall: Englewood Cliffs. (1977).
5. Bitterman, M. E. Classical conditioning since Pavlov. Review of General Psychology, 10(4), 365-376 (2006).

6. Blum, A., Mansour, Y. Learning, Regret Minimization, and Equilibria. Algorithmic Game Theory. doi:10.1017/CBO9780511800481.006, (2007).
7. Brown, N., Lerer, A. and Gross, S. and Sandholm, T. Deep Counterfactual Regret Minimization. In Proceedings of the 36th International Conference on Machine Learning, Proceedings of Machine Learning Research, 97:793—802 (2019). https://proceedings.mlr.press/v97/brown19b.html
8. Butz, M., Achimova, A., Bilkey, D., Knott, A. Event-Predictive Cognition: A Root for Conceptual Human Thought. Special issue on Event-Predictive Cognition: From Sensorimotor via Conceptual to Language-Based Struc-tures and Processes . Topics in Cognitive Science. (first available online 04 December 2020) doi:10.1111/tops.12522
9. Byrne, R. Cognitive processes in counterfactual thinking about what might have been. In: Medin, DL., editor. "The psychology of learning and motivation: Advances in research and theory. 37. Academic Press; San Diego, CA. p. 105-154 (1997).
10. Byrne R. Mental models and counterfactual thinking. Trends in Cognitive Sciences, 6, 405–445 (2002).
11. Byrne, R. The rational imagination: How people create alternatives to reality. MIT Press; Cambridge, MA. (2005).
12. Clark, R. E. The classical origins of Pavlov's conditioning. Integrative Physiological and Behavioral Science, 39(4), 279-294 (2004).
13. Connolly, and Zeelenberg, M. Regret in decision making. Current directions in psychological science, 11(6), 212-216 (2002).
14. Dix, A. The adaptive significance of regret. unpublished essay. (August 2005.) https://alandix.com/academic/essays/regret.pdf
15. Egan, K. Memory, imagination, and learning: Connected by the story. Phi Delta Kappan, 70(6), 455-459 (1989).
16. Epstude, K., Roese, N. J. The functional theory of counterfactual thinking. Personality and social psychology review, 12(2), 168-192 (2008).
17. EPSRC. Human-like Computing:: Report of a Workshop held on 17 & 18 February 2016, Bristol, UK. (2016).
18. Fehrer, E. Effects of amount of reinforcement and of pre-and post-reinforcement delays on learning and extinction. Journal of Experimental Psychology, 52(3), 167 (1956).
19. Friedman, S. L., Scholnick, E. K. (Eds.). The developmental psychology of planning: Why, how, and when do we plan?. Psychology Press. (2014).
20. Lattal K. A. Delayed reinforcement of operant behavior. Journal of the experimental analysis of behavior, 93(1), 129–139 (2010). doi:10.1901/jeab.2010.93-129
21. Loomes, G., Sugden, R. Regret Theory: An Alternative Theory of Rational Choice Under Uncertainty. The Economic Journal, 92(368), 805–824 (1982). doi:10.2307/2232669
22. Kahneman, D., Tversky, A. Prospect Theory: An Analysis of Decision under Risk. Econometrica, 47(2), 263–291 (1979). doi:10.2307/1914185
23. Kahneman, D., Miller, D. T. Norm theory: Comparing reality to its alternatives. Psychological Review, 93(2), 136–153 (1986). doi:10.1037/0033-295X.93.2.136
24. MacIntyre, P., Gregersen, T. Emotions that facilitate language learning: The positive-broadening power of the imagination. Studies in Second Language Learning and Teaching Department of English Studies, Faculty of Pedagogy and Fine Arts, Adam Mickiewicz University, Kalisz SSLLT 2 (2). 193-213 (2012). http://www.ssllt.amu.edu.pl

25. Maes, P., Mataric, M., Meyer, J-A., Pollack, J., Wilson, S. Skinnerbots, in From Animals to Animats 4: Proceedings of the Fourth International Conference on Simulation of Adaptive Behavior , MIT Press, pp.285-294 (1996).
26. Markman KD, McMullen MN. A reflection and evaluation model of comparative thinking. Personality and Social Psychology Review, 7, 244–267 (2003).
27. Markman, K., McMullen, M., Elizaga, R. Counterfactual thinking, persistence, and performance: A test of the reflection and evaluation model. Journal of Experimental Social Psychology, 44:421–428 (2008).
28. McCormack, T., Feeney, A., Beck, S. R. Regret and Decision-Making: A Developmental Perspective. Current Directions in Psychological Science, 29(4), 346-350 (2020). doi:10.1177/0963721420917688
29. McSweeney, F.K., Bierley, C. Recent developments in classical conditioning. Journal of Consumer Research, 11(2), 619-631 (1984).
30. Muggleton, S., Chater, N. (eds). Human-Like Machine Intelligence. Oxford University Press (2021).
31. O'Connor, E., McCormack, T., Feeney, A. Do children who expe-rience regret make better decisions? A developmental study of the behavioral conse-quences of regret. Child development, 85(5), 1995–2010 (2014). doi:10.1111/cdev.12253
32. Pearce, J., Bouton, M. Theories of associative learning in animals. Annu Rev Psychol., 52, 111–39 (2001). doi:10.1146/annurev.psych.52.1.111
33. Sanna, L., Stocker, S., Clarke, J. Rumination, imagination, and personality: Specters of the past and future in the present. (2003).
34. Shalev-Shwartz, S. Online Learning and Online Convex Optimization. Found. Trends Mach. Learn. 4(2), 107–194 (2012). doi:10.1561/2200000018
35. Skinner, B. Operant conditioning. The encyclopedia of education, 7, 29-33 (1971).
36. Smith, T., Kimball, D. Learning from feedback: Spacing and the delay–retention effect. Journal of Experimental Psychology: Learning, Memory, and Cognition, 36(1), 80 (2010).
37. Staddon, J., Cerutti, D.. (2003). Operant conditioning. Annual review of psychology, 54(1), 115-144.
38. Touretzky, D., Saksida, L. Operant conditioning in Skin-nerbots. Adaptive Behavior, 5(3–4), 219–247 (1997).
39. Van de Ven, N., Zeelenberg, M. Regret aversion and the reluctance to exchange lottery tickets. Journal of Economic Psychology, 32(1), 194–200 (2011).
40. Weick, K. The role of imagination in the organizing of knowledge. European Journal of Information Systems, 15(5), 446–452 (2006).
41. Zeelenberg, M., Van Dijk, W., Manstead, A., der Pligt, J. The experience of regret and disappointment. Cognition and Emotion, 12(2), 221–230 (1998).
42. Zeelenberg, M., Pieters, R. Consequences of regret aversion in real life: The case of the Dutch postcode lottery. Organizational Behavior and Human Decision Processes, 93(2), 155–168 (2004).
43. Zhang, Z Ji, X. Regret Minimization for Reinforcement Learning by Evaluating the Optimal Bias Function. In Advances in Neural Information Processing Systems, 32 (2019). https://proceedings.neurips.cc/paper/2019/file/9e984c108157cea74c894b5cf34efc44-Paper.pdf
44. Zinkevich, M., Johanson, M., Bowling, M,. Piccione, C. Regret minimization in games with incomplete information. In Proceedings of the 20th Inter-national Conference on Neural Information Processing Systems (NIPS'07). Curran Associates Inc., Red Hook, NY, USA, 1729–1736 (2007).