Neth, H., Sims, C. R., & Gray, W. D. (2005). Melioration despite more information: The role of feedback frequency in stable suboptimal performance. Proceedings of the 49th annual meeting of the Human Factors and Ergonomics Society, pp. 357–361.

Melioration Despite More Information: The Role of Feedback Frequency in Stable Suboptimal Performance

Hansjörg Neth, Chris R. Sims & Wayne D. Gray

Cognitive Science Department
Rensselaer Polytechnic Institute
[nethh; simsc; grayw]@rpi.edu

Situations that present individuals with a conflict between local and global gains often result in a behavioral pattern known as melioration — a preference for immediate rewards over higher long-term gains. Using a variant of a paradigm by Tunney & Shanks (2002), we explored the potential role of feedback as a means to reduce this bias. We hypothesized that frequent and informative feedback about optimal performance might be the key to enable people to overcome the documented tendency to meliorate when choices are rewarded probabilistically. Much to our surprise, this intuition turned out to be mistaken. Instead of maximizing, 19 out of 22 participants demonstrated a clear bias towards melioration, regardless of feedback condition. From a human factors perspective, our results suggest that even frequent normative feedback may be insufficient to overcome inefficient choice allocation. We discuss implications for the theoretical notion of rationality and provide suggestions for future research that might promote melioration as an explanatory mechanism in applied contexts.

It is an intuitively appealing assumption that rational organisms maximize their expected reward when making choices between options. The idea of optimal resource allocation (or maximization of subjective utility) is frequently equated with the very concept of rationality and one of the main guiding principles of experimental psychology, decision sciences, and economic theory. Despite its intuitive appeal, this notion of utility maximization might be mistaken. In an extensive series of experiments, Richard Herrnstein and colleagues (see Herrnstein, 1997) have documented many instances of motivated and systematic deviations from the rational ideal. When faced with a dilemma between short-term rewards and long-term gain, both animals and humans appear to systematically and reliably favor high immediate reinforcements over a higher overall gain — a phenomenon known as *melioration* (Herrnstein & Vaughn, 1980).

Imagine a simple gambling scenario involving a repeated forced-choice between two risky alternatives A and B (see Figure 1). The solid lines indicate the probability of receiving a reward by choosing options A or B as a function of recent choices to B. While choosing A at any moment yields a higher expected payoff, every choice of A reduces the subsequent probability of reward for both options. As the dashed line represents the expected reward for any mix of choice allocations, the optimal long-term strategy (indicated by the maximum point of the dashed line) is to allocate *all* choices to option B.

Even though the environmental contingencies just described concern repeated forced-choice decisions in a fairly abstract laboratory-type task, we believe that the underlying dynamics apply to human choice behavior in a wide range of practical contexts. For instance, novice typists face a dilemma between completing their immediate task of producing a document and investing additional time and effort into perfecting their typing skills. Yechiam and colleagues (2003) have empirically demonstrated that even extensive training in efficient strategies (such as touch-typing) does not automatically lead to their adoption if alternative methods with a

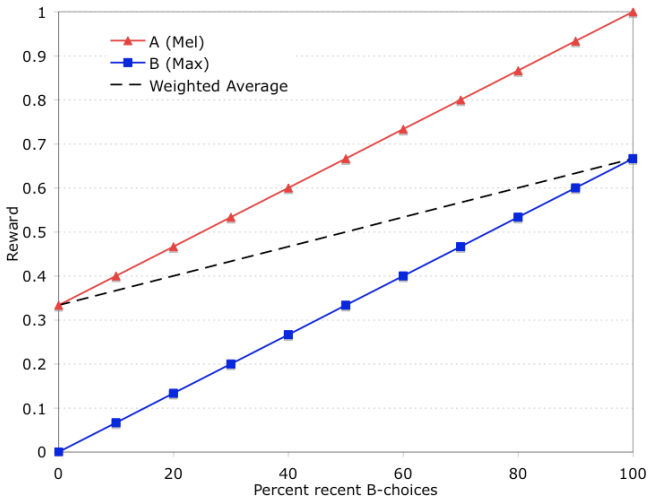

Figure 1: Environmental contingencies that have been known to induce melioration behavior. Option A is always more preferable than B. However, as the abscissa is the percentage of B choices over last N trials the weighted average (dashed line) is maximal when option B is chosen 100% of the time.

higher instant gratification exist. Typists tended to relapse into a less efficient visually-guided strategy after training had ceased and document production became their primary objective.

Numerous other examples of stable suboptimal behavior nurture further doubts about utility maximization as the driving force underlying human choice behavior. When interactive software packages offer more than one method to achieve a goal, even highly experienced users routinely select inefficient strategies (Bhavani & John, 2000). Fu and Gray (2004) have recently explained this 'paradox of the active user' (Carroll & Rosson, 1987) in terms of cost-benefit tradeoffs that favor small incremental gains of an interactive nature over less interactive but globally more efficient strategies.

Whereas previous research has often cast melioration in clinical terms of self control, addiction, and impulsiveness (see Herrnstein, 1997, Ch. 5–9) we approach the phenomenon as a problem of incomplete knowledge and a challenge to human information-processing limits.

An Information-Processing Perspective

In a series of experiments using the repeated forced-choice dilemma described above, Tunney and Shanks (2002) demonstrated that small changes in the type of payoffs can have large effects on behavior. Whereas participants maximized when payoffs systematically varied in magnitude (Exp-1), they tended to meliorate when payoffs were probabilistic (Exp-2). This bias to focus on immediate gains was alleviated when payoffs were negative (Exp-3) or when the test phase was preceded by an exploration phase (Exp-4). In the absence of a principled account, these results appear like an assortment of unrelated phenomena, suggesting that people's choice allocation is heavily context-dependent and subject to relatively random situational constraints.

The conflict between melioration and maximization is a consequence of the competition between two different timescales: attention to short-term rewards (on a local timescale) would favor option A, whereas attention to long-term gains (or adopting a global perspective) would favor option B.

One possible way to systematically influence the adoption of a short-term or long-term perspective is by manipulating feedback. Trial-by-trial feedback in terms of outcome is controlled by the current payoff contingencies, which in turn depend on the choice distribution to A and B over the last N trials. The span of recent trials over which this ratio of A/B is calculated can be defined as the *reward window*. As the reward window is shifted from a larger (e.g., 40) to a smaller (e.g., 10) span, melioration decreases and maximization increases (Herrnstein, 1991).

In addition to the reward window, an explicit *feedback window* can be defined as the number of trials that are considered whenever feedback is provided. This window can vary not only in size, but can also independently vary in its frequency. Whereas the feedback window size determines over how many recent trials the feedback information spans, feedback frequency divides a sequence of trials into discrete learning episodes. In our experiment, both feedback window size and frequency were set to the same value.

We felt that the combination of a reward window size of 10 with a feedback window of size 100 employed by Tunney and Shanks (2002) was somewhat arbitrary. Moreover, a learning episode affording 2^{100} possible different sequences seems likely to exceed the limits of human information processing. In contrast, a feedback presentation every 10 trials offers ten times as many feedback instances and simultaneously affords 2^{90} fewer candidate strategies to be explored.

Providing Optimal Feedback

Although it is clear that feedback plays a central role in determining performance in melioration experiments, it is less clear what exactly would constitute appropriate or helpful feedback to a participant. A rigorous theoretical account of feedback is offered by the literature of control theory. In the classical example, an idealized controller is given the task of regulating a system in order to optimize some measure of performance. For example, a thermostat can be viewed as an operator given the task of minimizing temperature deviations from a reference value. The ability to adjust to the environmental changes depends on two pieces of information being available to the operator: a reference signal (such as the temperature set on the dial of a thermostat), and an output signal (the current room temperature). Only by comparison of both can the operator steer performance in the appropriate direction.

Applied to our melioration scenario, the most obvious choice of output signal is the actual reward earned by the participant during recent trials. However, the choice of reference signal, while seemingly straightforward, can introduce subtle biases into behavior. Consider the most straightforward choice of a reference signal. If it is possible to earn \$10 over the course of the entire experiment and there are five feedback presentations it is reasonable to assume that each feedback presentation delimits an independent block of trials and that the possible reward for each block is a maximum of \$2.

Neth, H., Sims, C. R., & Gray, W. D. (2005). Melioration despite more information: The role of feedback frequency in stable suboptimal performance. Proceedings of the 49th annual meeting of the Human Factors and Ergonomics Society, pp. 357–361.

However, the reward window can be based on a span of trials that is longer than the feedback window. Suppose that during the first block of the experiment the participant adopts a pure melioration strategy. Upon receiving feedback the participant may realize the inefficiency of his or her strategy and switch to a pure maximization strategy for the second block. However, as the rewards in the second block are partially dependent upon the suboptimal choices made during the first block, it will be impossible to achieve the full $2 as indicated by the reference signal. A perfectly reasonable response then would be to abandon the maximizing strategy because it failed to achieve the provided reference signal.

A related problem is that for very small feedback windows it would actually seem that it was to the participant's benefit to meliorate. Consider the extreme case of a feedback window of size 1. In this case, the expected reward for a meliorating response exceeds that for a maximizing response. Upon reflection, this exemplifies the inherent dilemma of a melioration scenario: locally (for small feedback windows) it is always better to meliorate. Only on a global scale are the benefits of maximization observable. Thus, any feedback mechanism that attempts to reduce the tendency to meliorate has to strike a balance between short and long feedback windows and extrapolate from short-term behavior to its global consequences.

Taking into account these considerations, we believe that an intermediate feedback window span provides the best compromise and opted for a feedback window size that matches the size of the reward window.

Experiment

In this experiment participants faced a simple choice between two alternatives for 500 trials. On each trial, the chosen option was probabilistically rewarded by a small fixed amount of $.03. The size of the feedback window was varied as a between-subjects manipulation. In the f-10 condition, participants received feedback every 10 trials, while in the f-100 condition participants received feedback every 100 trials.

Method

Participants

Twenty-two RPI undergraduate students volunteered to participate in this study. Participants were informed that they could win between $5 and $10.

Apparatus

Participants were tested individually in a quiet room. The experimental software was written in LispWorks Common Lisp and run on a Macintosh-G4 computer. Two buttons marked 'Left' and 'Right' were displayed at the bottom of the task window (see Figure 2). The top of the window contained information on the participant's cumulative winnings, as well as the previous trial number, their choice on that trial, and the reward received for that choice.

Figure 2: A screenshot of the experimental task window.

Procedure

At the beginning of the study participants were informed that the amount of money earned during the study was based directly on their decisions and that an optimal strategy could earn them as much as $10. Each individual choice was indicated by pressing either the left or the right button. After each choice, both buttons were disabled for .5 sec and the feedback from the previous trial was erased. All visible information was subsequently updated and the buttons were re-enabled. The participant was then free to make the next choice.

Payoff in this experiment was a fixed $.03 reward that was received probabilistically, and the Max button was randomly assigned to either the left or right button.

The current probability of receiving a reward upon selecting an alternative was based on the participant's distribution of choices over the last 10 trials, using the reward function illustrated by Figure 1. For the Max button, the probability of receiving a reward equaled 0 + 2/3 x (% of previous 10 trials spent on the Max button). For the Mel button, the probability of receiving a reward equaled 1/3 + 2/3 x (% of previous 10 trials spent on the Max button). Over the course of 500 trials, consistently choosing the Max button would provide an expected reward of $10, while consistently choosing the Mel button would provide an expected reward of $5.

Neth, H., Sims, C. R., & Gray, W. D. (2005). Melioration despite more information: The role of feedback frequency in stable suboptimal performance. Proceedings of the 49th annual meeting of the Human Factors and Ergonomics Society, pp. 357–361.

After every 10 or 100 trials (depending on condition), a feedback screen was displayed. This window contrasted the actual amount earned by the participant with the amount that could be expected by following the optimal strategy ($.20 for f-10, $2 for f-100).

Results

Participants in the f-10 condition earned an average payoff of $6.77 (SD = 0.48), while in the f-100 condition participants earned on average $6.57 (SD = 0.67). Figure 3 shows the percentage of maximizing choices for each of the two feedback conditions, divided into blocks of 100 trials. Contrary to our hypotheses, there is no hint that the experimental manipulation of feedback had any impact on performance. Consistent with this interpretation, a 2x5 mixed-design ANOVA (feedback condition x block) yielded no main effect of feedback condition [$F(1, 20)=.40$, $MSE=447.8$, $p=.53$]. Thus, varying feedback window size had no impact on performance in the task. Additionally, there was no significant interaction [$F(2.9, 57.7)=.44$, $MSE=176.7$, $p=.72$, Huynh-Feldt correction for sphericity violation], but a significant main effect of block [$F(2.9,57.7)=3.0$, $MSE=176.7$, $p=.04$]. Subsequent pairwise comparisons revealed that the only significant difference was a reduction of maximization responses from block 1 to 5 [$p = .017$, Sidak adjustment for multiple comparisons].

In addition to the ANOVA, non-parametric binomial tests were conducted for each participant on the number of choices to the Max button across the entire experiment. Out of 22 participants, 19 exhibited significant deviations from chance performance ($p < 0.001$) in the direction of a melioration bias. Two of the remaining three participants who did not differ significantly from chance were in the f-100 group, again indicating that more frequent feedback in the f-10 condition did not facilitate maximization.

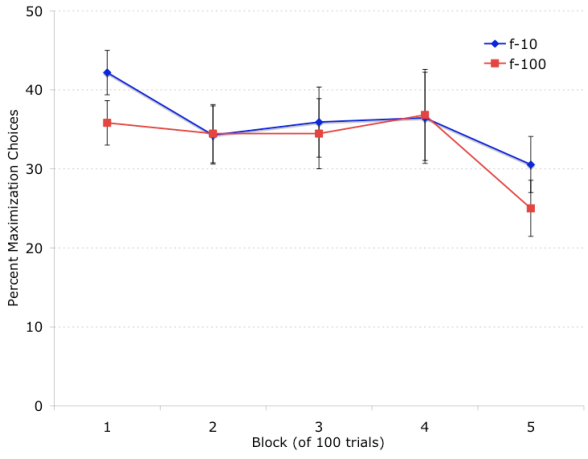

Figure 3: Mean percentage of maximization choices by feedback condition. (Error bars represent standard errors.)

Curiously, an analysis of the number of button switches (in either direction) yielded a significant main effect of feedback condition [$F(1, 20)=6.7$, $MSE=417.9$, $p=.02$]. Whereas participants in the f-100 condition on average switched on 26.6% of trials, participants in the f-10 condition switched on 36.7% of trials.

Discussion

In designing this study we were confident that reducing the feedback window from 100 to 10 trials would significantly improve the participants' performance, if not completely eliminate all evidence of melioration. Much to our surprise, there is no evidence that more frequent feedback had any impact at all on performance. Despite their strongly professed desires, none of our participants earned the maximal reward, or even came close to achieving it. Although this result proved our intuitions wrong, we believe this dramatic failure to be more interesting than a positive result.

In hindsight there are several possible explanations for the results obtained. The most prosaic account would be that our participants simply devoted too little attention to the feedback. However, as participants also switched buttons significantly more due to a smaller feedback window we believe this to be unlikely. Paradoxically, the fact that we succeeded in manipulating their choice allocation might have steered people away from an optimal solution that demands a complete absence of switches.

A more intriguing possibility is hinted at by the fact that all of our participants were highly curious to find out the optimal strategy upon completion of the experiment. This might suggest that their inability to discover the maximization strategy did not stem from lack of attention but rather a profound bias to explore and utilize all aspects of the environment. One peculiar aspect of the task environment is that global maximization requires the participant to completely abandon one of two available options. This necessity might conflict with an inherent *variability bias*, or drive to distribute choices across all available options in the environment.

While controversial at first glance, this hypothesis draws support from a number of seemingly disparate findings. Herrnstein (1990, 1997) has strongly argued that the atomic unit of individual behavior is not a single choice, but rather a distribution of choices over all available alternatives that would bias against any extremely uniform response strategy. Additionally, data from children's strategy discovery in mental arithmetic

suggests that older and inefficient strategies are still occasionally employed despite familiarity with better methods (Siegler & Stern, 1998). In isolation, this result might appear maladaptive, but in a dynamic environment strategic variability might be a valuable survival mechanism.

Conclusion

The history of the study of human decision-making has long held the assumption that the underlying motivation for rational behavior is to maximize goal achievement. However, forty years of results in the study of melioration have remained a thorn in the side of this assumption, and our current study only furthers our understanding of the extent of the problem. In short, whenever local and global rewards are in direct competition, people may be strongly drawn towards immediate reinforcement even though this incurs an overall loss. Far more than just a problem for controlled psychology paradigms, melioration affects performance in all aspects of routine, everyday behavior.

If melioration is an ubiquitous phenomenon, it also presents challenges to designers, e.g., of software applications offering user assistance. When facing difficult tasks, users might become overly reliant on software assistance rather than acquiring the necessary skills to solve the task themselves. Examples include spell checkers that might diminish users' ability to spell, users becoming addicted to the assistance provided by adaptive user interfaces, or over-reliance on external memory aids (e.g., storing contact information in mobile devices instead of making an effort to rehearse and recall the information).

Ideally, software designers and the human factors community would benefit most from an easy-to-implement, generalizeable fix to the problem. Unfortunately, our inability to induce more successful strategies raises skepticism that suboptimal performance can be avoided simply through the incorporation of more informative feedback. This is of particular importance in the context of supervisory control tasks, in which complex and often delayed system parameters are conveyed through interactive information displays.

A pessimistic interpretation of this study would point out that there might be no simple fix to the problem of melioration. On the positive side, the absence of an improvement despite frequent informative feedback not only poses an intriguing puzzle for further investigation but may help to promote melioration as an important explanatory mechanism in Human Factors research.

Future research will explore the variability bias by studying multi-alternative decisions and establish how the mechanisms investigated in abstract task scenarios manifest themselves in more applied contexts.

Acknowledgments

The work reported was supported by grants from the Air Force Office of Scientific Research (AFOSR #F49620-03-1-0143), as well as the Office of Naval Research (ONR #N000140310046).

References

Bhavnani, S. K., & John, B. E. (2000). The strategic use of complex computer systems. *Human-Computer Interaction, 15*(2-3), 107–137.

Carroll, J. M., & Rosson, M. B. (1987). Paradox of the active user. In J. M. Carroll (Ed.), *Interfacing thought: Cognitive aspects of human–computer interaction*. Cambridge, MA: MIT Press.

Fu, W.-T. & Gray, W. D. (2004). Resolving the paradox of the active user: Stable suboptimal performance in interactive tasks. *Cognitive Science, 28(6)*, 901–937.

Herrnstein, R. J. (1990). Behavior, Reinforcement and Utility. *Psychological Science, 1(4)*, 217–224.

Herrnstein, R. J. (1991). Experiments on stable suboptimality in individual behavior. *The American Economic Review*, *81(2)*, 360–364.

Herrnstein, R. J. (1997). *The matching law*. H. Rachlin & D. I. Laibson (Eds.). Cambridge, MA: The Harvard University Press.

Herrnstein, R. J. & Vaughn, W. (1980). Melioration and behavioral allocation. In Staddon, J. E. R. (Ed.), *Limits to action: The allocation of behavior* (pp. 143–176), New York: Academic Press.

Siegler, R. S. & Stern, E. (1998). Conscious and unconscious strategy discoveries: A microgenetic analysis. *Journal of Experimental Psychology: General, 127*, 377–397.

Tunney, R. J., & Shanks, D. R. (2002). A re-examination of melioration and rational choice. *Journal of Behavioral Decision Making, 15(4)*, 291–311.

Yechiam, E., Erev, I., Yehene, V., & Gopher, D. (2003). Melioration and the transition from touch-typing training to everyday use. *Human Factors, 45(4)*, 671–684.