Published in final edited form as:

Behav Brain Res. 2010 October 15; 212(2): 204–207. doi:10.1016/j.bbr.2010.03.053.

Learning That a Cocaine Reward is Smaller Than Expected: A Test of Redish's Computational Model of Addiction

Katherine R. Marks, David N. Kearns, Chesley J. Christensen, Alan Silberberg, and Stanley J. Weiss

Psychology Department American University

Abstract

The present experiment tested the prediction of Redish's [7] computational model of addiction that drug reward expectation continues to grow even when the received drug reward is smaller than expected. Initially, rats were trained to press two levers, each associated with a large dose of cocaine. Then, the dose associated with one of the levers was substantially reduced. Thus, when rats first pressed the reduced-dose lever, they expected a large cocaine reward, but received a small one. On subsequent choice tests, preference for the reduced-dose lever was reduced, demonstrating that rats learned to devalue the reduced-dose lever. The finding that rats learned to lower reward expectation when they received a smaller-than-expected cocaine reward is in opposition to the hypothesis that drug reinforcers produce a perpetual and non-correctable positive prediction error that causes the learned value of drug rewards to continually grow. Instead, the present results suggest that standard error-correction learning rules apply even to drug reinforcers.

Keywords

learning; reward prediction errors; dopamine; cocaine; addiction; rats

Redish's [7] computational model of addiction is based in large part on research showing that dopamine serves as the reward-prediction error signal of error-correction learning theories [12], such as the temporal-difference reinforcement learning (TDRL [13]) or Rescorla-Wagner models [9,15]. According to these error-correction models, learning is driven by the discrepancy between expectation and outcome. If a rewarding outcome is larger than expected, a positive prediction error is generated. Learning is the process whereby this error is "corrected" through an increase in future reward expectation. If, on the other hand, a reward is smaller than expected, a negative prediction error occurs. New learning takes place that acts to reduce the expectation-outcome discrepancy by decreasing reward expectation. When expectation matches the outcome received, no prediction error occurs, and therefore no learning takes place.

Much neurophysiological research has shown that the firing of midbrain dopamine neurons displays a high level of correspondence with hypothetical reward-prediction errors (for reviews see [11,12]). For example, Waelti et al. [14] showed in monkeys that when a conditioned stimulus (CS) was followed by an unexpected fruit juice reward, thereby producing a positive

© 2010 Elsevier B.V. All rights reserved.

Katherine R. Marks Psychology Department American University Washington, DC 20016 202-885-1725 km1127a@american.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

prediction error, there was a phasic increase in dopamine activity. However, when monkeys had learned to expect the juice reward during the CS, presentation of the juice no longer produced a change in dopamine activity. Finally, when an expected juice reward was omitted, thereby generating a negative prediction error, there was a phasic decrease in dopamine activity.

Redish [7] noted that in contrast to natural rewards, where phasic dopamine responses stop once the subject has learned to expect the reward, drugs of abuse continue to produce a dopamine increase even after an extensive learning history. This direct, neuropharmacologically produced dopamine increase functions as if it were a positive prediction error, causing the subject to increase its expectation of drug reward even though the just-experienced drug reward was no larger than expected. As drug use continues, the predicted drug reward continues to grow progressively larger than the actually received drug reward. This results in addiction-like behavioral outcomes such as increased choice of drug rewards over natural rewards and decreased elasticity of demand for drug rewards with repeated drug use.

It may seem counterintuitive that an experienced drug user might regularly expect a large drug reward, receive a smaller-than-expected drug reward, and not learn from the difference. However, this failure of error correction is the essential feature that distinguishes addictive drugs from natural rewards for Redish's model. With non-drug reinforcers, the negative prediction error generated when the received outcome is smaller than expected corrects the discrepancy between expectation and outcome. In contrast, for Redish's model, negative prediction errors that should occur with drug rewards are overridden by the dopamine surge produced by the direct effects of the drug. The drug-induced dopamine surge functions as a positive prediction error and prevents the subject from learning that the received reward was smaller than expected (and instead causes reward expectation to increase).

The present experiment tested this prediction of Redish's model by first training rats to expect a large cocaine reward, then presenting them with a small cocaine reward, and finally determining whether they learned from this discrepancy between expectation and outcome. This was accomplished by first training rats to press two levers, each associated with a large dose of self-administered cocaine. Lever preference was measured on a choice test after this equivalent dose training. Then, the dose associated with one of the levers was reduced (while the other lever was not presented). After this devaluation training, lever preference was measured again. A decrease in preference for the reduced-dose lever would be predicted by traditional error-correction learning models (e.g., TDRL, Rescorla-Wagner). In contrast, the central hypothesis of Redish's model is that this error correction fails to occur with drug rewards. According to Redish's model, reward expectation associated with the reduced-dose lever should not have decreased, but instead should have actually continued to grow when rats received the reduced dose. If this occurred, rats should have increased their preference for the reduced-dose lever.

Six experimentally-naïve adult male Long-Evans rats weighing approximately 320 g served as subjects. Rats had unlimited access to water in their home cages and were fed 15 g of standard rat chow per day. The colony room where the rats were housed had a 12-h light: 12-h dark cycle with lights on at 08:00 h. Training sessions were conducted 5-7 days per week during the light phase. Rats were surgically prepared with chronic indwelling jugular vein catheters under ketamine (60 mg/kg) and xylazine (10 mg/kg) anesthesia, using procedures described in detail previously [3].

Experimental sessions were conducted in Coulbourn Modular Test Cages (30.0 cm × 25.5 cm × 29.0 cm). Each chamber was equipped with two Coulbourn retractable levers located on the

chamber's front wall. The left and right levers were positioned approximately 2 cm from the left and right side walls, respectively. Each lever measured 3.3 cm wide and, when inserted, extended approximately 1.5 cm into the chamber approximately 2.25 cm above the grid floor. Three Coulbourn stimulus lights that signaled the start and end of a session were located in the center of the front wall approximately 10 cm above the floor. A speaker connected to a Med-Associates ANL 926 audio generator provided the tone (4000 Hz, approximately 90 dB) and white noise (approximately 90 dB) stimuli. Cocaine (NIDA, Bethesda, MD) in a saline solution (2.56 mg/ml) was delivered through Tygon tubing that was suspended through the ceiling of the chamber and wrapped in a metal spring. This tubing was attached to a 10-ml syringe that was driven by a Med-Associates syringe pump that infused solution at a rate of 3.19 ml/min.

Throughout Phase 1 of the experiment, the dose per infusion was 0.4 mg cocaine (equivalent to 1.25 mg/kg for a 320-g rat). The cocaine was infused over 2.9 s. Rats were initially trained on a discrete-trials procedure where only one lever was inserted into the chamber at a time. Levers were presented in a random order, with the restriction that there were no more than two consecutive presentations of the same lever. Lever pressing was reinforced by a cocaine infusion according to a fixed-ratio (FR) 1 schedule. Each trial lasted until the subject made a response on the inserted lever. Once a lever press was made, the infusion pump was activated (for 2.9 s) and a 10-s auditory stimulus was presented. For the left lever, the auditory stimulus was the tone, which remained on continuously during the 10-s period. For the right lever, the auditory stimulus was the white noise, which pulsed on and off (0.4 s on, 0.2 s off, 0.4 on, each second) throughout the 10-s period. The lever was retracted 5 s after the lever press was made. Once the 10-s auditory stimulus associated with cocaine ended, a variable inter-trial interval (ITI) lasting 120 s on average (range: 90-150 s) began. During this ITI, both levers remained retracted. When a subject began to regularly self-administer cocaine, the response requirement to receive an infusion was increased to FR 3. Sessions lasted 2.5 h.

After regularly self-administering on the FR-3 schedule for at least 2 sessions, a subject advanced to free-choice training. To ensure that subjects continued to sample both levers on the free-choice procedure, each free-choice session began with 6 forced-choice trials that included 3 insertions each of the left and right levers. On subsequent trials, the left and right levers were presented simultaneously. When a subject made a choice by pressing one of the levers, the other lever retracted. Then, the subject was required to complete an FR 3 to receive a cocaine infusion. All other events concomitant with infusion delivery were the same as previously described. Thus, four lever presses (one choice response plus three FR-3 responses) on a lever were required to obtain an infusion. Once the 10-s tone or white noise ended, a 90-150-s ITI began, just as before. After a minimum of five total free-choice sessions and meeting a criterion of making at least two free choices on each lever for at least two consecutive sessions, a preference test was administered.

Lever preference was assessed with a choice test performed under extinction conditions (no cocaine was self-administered during the test). On each trial, both levers were simultaneously inserted into the chamber. When the subject chose one of the levers by pressing it, the other lever retracted. Then, once a rat completed an FR 2 on the chosen lever, the auditory stimulus (the tone or the pulsed white noise) associated with that lever was presented for 4 s. No infusion was delivered. The lever retracted 2 s after the FR-2 requirement was met. After the 4-s auditory stimulus ended, a 10-s ITI began, followed by the insertion of both levers. The test lasted for a maximum of 40 free-choice trials or 30 min, whichever came first.

After the initial lever preference test, 2-3 free-choice self-administration sessions (with the same procedural details as the sessions immediately preceding the initial lever preference choice test) were administered to reestablish the self-administration baseline.

Then, Phase 2 devaluation training began. For each subject, either the left or the right lever, counterbalanced over subjects, was selected as the reduced-dose lever (hereinafter, Lever RD). The dose per infusion available for self-administration on this lever was reduced to 0.08 mg cocaine (0.25 mg/kg for a 320-g rat). The other lever, always associated with the large dose, was not presented during these sessions.

Each session ended after 50 Lever RD presentations or 2.5 h, whichever came first. Completion of an FR 3 on the inserted lever resulted in activation of the syringe pump and a 10-s presentation of the auditory stimulus associated with that lever. The lever retracted 5 s after completion of the FR 3. After the 10-s auditory stimulus ended, a variable 120-s (range: 90-150) ITI began. Rats were trained on this procedure for two sessions. Following this devaluation training, a lever-preference test just like that described above was administered.

In Phase 1, rats received a mean of 11.5 ($\pm$ 1.9 SEM) acquisition sessions plus 7.2 ($\pm$ 1.1 SEM) free-choice sessions, for a mean total of 18.7 ($\pm$ 2.1 SEM) cocaine self-administration sessions with the large dose available on both levers prior to the initial lever preference choice test. During the two free-choice sessions that preceded this initial lever preference test, rats self-administered a mean of 20.8 ($\pm$ 1.1 SEM) total cocaine infusions per session. Averaged over these two sessions, the mean percentage of free-choices made on the to-be-devalued lever (i.e., the lever subsequently designated Lever RD) was 57.0% ($\pm$ 6.4 SEM).

As the black bar in Figure 1 illustrates, the mean percentage of choices made to Lever RD on the initial lever preference test was 57.1% ($\pm$ 6.5 SEM), which very closely approximates the percentage of free choices made to this lever during the final two self-administration sessions. Rats made a mean of 17.0 ($\pm$ 2.7 SEM) choices during this 30-min test performed under extinction conditions.

Averaged over the two baseline recovery sessions that preceded devaluation training, subjects self-administered a mean of 20.0 ($\pm$ 0.8 SEM) infusions per session and made 56.5% ($\pm$ 9.8 SEM) of their free choices to Lever RD. These numbers are comparable to the number of self-administered infusions and choice percentages (20.8 and 57.0%, respectively) on the two baseline sessions that preceded the initial lever preference test.

In Phase 2, the mean number of reduced-dose infusions self-administered per session, averaged over the two devaluation training sessions, was 49.7 ($\pm$ 0.3 SEM) (out of a possible maximum of 50) for five of the six rats. For the sixth subject (S-17), the dose reduction from 0.4 mg/infusion to 0.08 mg/infusion appears to have been too sudden because this subject self-administered only 5 infusions over these two sessions. To ensure that this subject had sufficient exposure to the reduced dose during this phase, two additional sessions were given where the dose for the first several infusions of each session was 0.13 mg before being gradually reduced to 0.08 mg for the majority of the session. This strategy was successful, as S-17 self-administered 43 of the 0.08 mg infusions (out of 50 total infusions) on the final devaluation session.

One subject (S-11), who developed a tendency to hesitate before beginning to respond during self-administration sessions, waited 27 min into this 30-min test to make only one choice (on Lever RD) and did not complete the FR 2 on that trial before the test session ended. To address this issue, the preference-test procedure was modified for this subject so that the 30-min session timer did not begin until the subject made a first response. This subject was first retrained on the Phase 1 free-choice baseline with the large dose available on both levers for 12 sessions and was then re-administered the initial lever-preference choice test (where the outcome was essentially unchanged – 54.5% vs. 50.0% preference for Lever RD on the original test and re-test, respectively), two baseline recovery sessions, and two devaluation sessions. On the modified post-devaluation test, the subject now made eight total choices, with only 25% of

them to Lever RD – a 50% reduction in preference. Only the test results subsequent to re-training on the Phase 1 free-choice baseline are included for this subject in the statistical analysis.

The white bar in Figure 1 presents the mean results from the Phase 2 post-devaluation preference test. Rats made 39.8% (± 11.4 SEM) of their choices on Lever RD, the reduced-dose lever. This was a significant ($t[5] = 2.6$, $p < 0.05$) decrease in preference from the initial lever preference test. Rats made 15.0 (± 3.7 SEM) choices (out of a maximum possible of 40) during this 30-min test.

The present experiment found that rats could learn that a cocaine reward was smaller than expected. As described above, the essential hypothesis of Redish's computational model of addiction is that such learning should not be possible with drugs of abuse. According to the model, reward expectation associated with the reduced-dose lever should not have decreased during Phase 2, but should have continued to grow. The result of the Phase 2 choice test, where rats' preference for the reduced-dose lever decreased, is inconsistent with this prediction of Redish's model, but is consistent with traditional error-correction models (e.g., Rescorla-Wagner).

Redish's model is logically consistent with the broader assumptions that 1) abused drugs produce a phasic dopamine response like that which occurs with reward prediction errors, and 2) phasic dopamine responses associated with prediction errors cause reward learning. The model prediction tested in the present experiment also depends on these two assumptions being true. That the results of the present study (as well as those of a previous study by Panlilio et al. [5]) did not support the model suggests that these assumptions might not provide a complete picture of the potential interactions among dopamine, reward prediction errors, and cocaine.

The assumption that abused drugs produce a dopamine response that is functionally equivalent to the phasic dopamine response associated with reward prediction errors may need to be modified given research showing that the time courses of these two types of dopamine activity differ substantially. For example, significant increases in rats' baseline dopamine levels in the nucleus accumbens following presentation of unexpected food were observed only from 1.0-1.3 s following reinforcer delivery [2]. In contrast, elevations in accumbens dopamine produced by the pharmacological effects of cocaine in rats begin 3-4 s after the onset of an intravenous infusion, reach a plateau after approximately 20 s, and remain elevated for up to 1 h [4]. Given these disparate time courses, it appears unlikely that the dopamine surge produced by cocaine would function comparably to a reward prediction error. The results of the present experiment are consistent with this notion.

The second assumption, that phasic dopamine responses associated with prediction errors drive learning, has also recently been called into question (for reviews, see [1,6]). For example, Berridge [1] reviews evidence suggesting that phasic dopamine responses are neither necessary nor sufficient for reward learning. Instead, Berridge posits that the phasic dopamine responses correlated with prediction errors are only a consequence of reward learning (that takes place in other parts of the brain), rather than a cause of reward learning. Redgrave et al. [6] argue that diverse events other than unexpected rewards also elicit phasic dopamine responses and therefore it is unlikely that these responses cause error-correction learning. They also note that the phasic dopamine responses occur sooner than the time it takes for the sensory processing necessary to identify the eliciting stimulus and to determine whether a reward was smaller or larger than expected. If phasic dopamine responses do not mediate error-correction reward learning, then the dopamine surge produced by cocaine would not be expected to prevent learning that cocaine rewards are smaller-than-expected (as found in the present experiment).

What role, then, might dopamine play in behavior maintained by drugs of abuse? Redish et al. [8] have proposed a more recent theory that views addiction as the result of a decision process with at least ten potential vulnerabilities (which interact with each other) to maladaptive influences. A dopamine-induced overvaluation of actions leading to drug rewards is only one such potential vulnerability. Dopamine is involved in a number of other vulnerabilities and may contribute to heightened motivation for the drug (e.g., [10]), over-sensitivity to drug cues, and enhanced learning of cue-drug associations. As Redish et al. advocate, a more unified framework that considers multiple potential points in the decision process may lead to a better understanding of what causes continued drug-taking behavior despite negative consequences than any single theory that focuses on only one aspect of behavior.

Acknowledgments

This research was supported by NIDA Grant R01-DA-08651 awarded to SJW and by NIDA Grant 5F31-DA-024493 awarded to CJC.

References

1. Berridge KC. The debate over dopamine's role in reward: the case for incentive salience. *Psychopharmacology* 2007;191:391–431. [PubMed: 17072591]
2. Day JJ, Roitman MF, Wightman RM, Carelli RM. Associative learning mediates dynamic shifts in dopamine signaling in the nucleus accumbens. *Nature Neurosci* 2007;10:1020–1028. [PubMed: 17603481]
3. Kearns DN, Weiss SJ, Schindler CW, Panlilio LV. Conditioned inhibition of cocaine seeking in rats. *J Exp Psychol: Anim Behav Proc* 2005;31:247–253.
4. Mateo Y, Budygin EA, Morgan D, Roberts DCS, Jones SR. Fast onset of dopamine uptake inhibition by intravenous cocaine. *Eur J Neurosci* 2004;20:2838–2842. [PubMed: 15548229]
5. Panlilio LV, Thorndike EB, Schindler CW. Blocking of conditioning to a cocaine-paired stimulus: testing the hypothesis that cocaine perpetually produces a signal of larger-than-expected reward. *Pharmacol Biochem Behav* 2007;86:774–777. [PubMed: 17445874]
6. Redgrave P, Gurney K, Reynolds J. What is reinforced by phasic dopamine signals? *Brain Res Rev* 2008;58:322–339. [PubMed: 18055018]
7. Redish DA. Addiction as a computational process gone awry. *Science* 2004;306:1944–1947. [PubMed: 15591205]
8. Redish DA, Jensen S, Johnson A. A unified framework for addiction: vulnerabilities in the decision process. *Behav Brain Sci* 2008;31:415–487. [PubMed: 18662461]
9. Rescorla, RA.; Wagner, AR. A theory of Pavlovian conditioning: variations in the effectiveness of reinforcement and nonreinforcement. In: Black, AH.; Prokasy, WF., editors. *Classical conditioning II: current research and theory*. Appleton-Century-Crofts; New York: 1972. p. 64-99.
10. Robinson TE, Berridge KC. The neural basis of drug craving: an incentive-sensitization theory of addiction. *Brain Res Rev* 1993;18:247–291. [PubMed: 8401595]
11. Schultz W. Predictive reward signal of dopamine neurons. *J Neurophysiol* 1998;80:1–27. [PubMed: 9658025]
12. Schultz W, Dickinson A. Neuronal coding of prediction errors. *Annu Rev Neurosci* 2000;23:473–500. [PubMed: 10845072]
13. Sutton, RS.; Barto, AG. *Reinforcement learning*. The MIT Press; Cambridge, MA: 1998.
14. Waelti P, Dickinson A, Schultz W. Dopamine responses comply with basic assumptions of formal learning theory. *Nature* 2001;412:43–48. [PubMed: 11452299]
15. Wagner, AR.; Rescorla, RA. Inhibition in Pavlovian conditioning: application of a theory. In: Boakes, RA.; Halliday, MS., editors. *Inhibition and learning*. Academic Press; San Diego: 1972. p. 301-336.

Percentage of Choices Made on Lever RD (reduced-dose lever)

Figure 1.

Mean ($\pm$ SEM) percentage of choices made on Lever RD (the reduced-dose lever) during the initial lever-preference test administered at the end of Phase 1 (black bar) and during the post-devaluation preference test administered at the end of Phase 2 (white bar). Preference for Lever RD on the Phase 2 test was significantly ($p < 0.05$) lower than that on the Phase 1 test.