

Addiction as a Computational Process Gone Awry

A. David Redish

Addictive drugs have been hypothesized to access the same neurophysiological mechanisms as natural learning systems. These natural learning systems can be modeled through temporal-difference reinforcement learning (TDRL), which requires a reward-error signal that has been hypothesized to be carried by dopamine. TDRL learns to predict reward by driving that reward-error signal to zero. By adding a noncompensable drug-induced dopamine increase to a TDRL model, a computational model of addiction is constructed that over-selects actions leading to drug receipt. The model provides an explanation for important aspects of the addiction literature and provides a theoretic viewpoint with which to address other aspects.

If addiction accesses the same neurophysiological mechanisms used by normal reinforcement-learning systems (1–3), then it should be possible to construct a computational model based on current reinforcement-learning theories (4–7) that inappropriately selects an “addictive” stimulus. In this paper, I present a computational model of the behavioral consequences of one effect of drugs of abuse, which is increasing phasic dopamine levels through neuropharmacological means. Many drugs of abuse increase dopamine levels either directly [e.g., cocaine (8)] or indirectly [e.g., nicotine (9, 10) and heroin (11)]. A neuropharmacologically driven increase in dopamine is not the sole effect of these drugs, nor is it likely to be the sole reason that drugs of abuse are addictive. However, this model provides an immediate explanation for several important aspects of the addiction literature, including the sensitivity of the probability of selection of drug receipt to prior drug experience, to the size of the contrasting nondrug reward, and the sensitivity but inelasticity of drugs of abuse to cost.

The proposed model has its basis in temporal-difference reinforcement models in which actions are selected so as to maximize future reward (6, 7). This is done through the calculation of a value function $V[s(t)]$, dependent on the state of the world $s(t)$. The value function is defined as the expected future reward, discounted by the expected time to reward:

$$V(t) = \int_t^\infty \gamma^{t-\tau} E[R(\tau)] d\tau \quad (1)$$

where $E[R(\tau)]$ is the expected reward at time τ and γ is a discounting factor ($0 < \gamma < 1$) reducing the value of delayed rewards. Equation 1 assumes exponential discounting

in order to accommodate the learning algorithm (6, 7); however, animals (including humans) show hyperbolic discounting of future rewards (12, 13). This will be addressed by including multiple discounting time scales within the model (14).

In temporal-difference reinforcement learning (TDRL), an agent (the subject) traverses a world consisting of a limited number of explicit states. The state of the world can change because of the action of the agent or as a process inherent in the world (i.e., external to the agent). For example, a model of delay conditioning may include an interstimulus-interval state (indicated to the agent by the observation of an ongoing tone); after a set dwell time within that state, the world transitions to a reward state and delivers a reward to the agent. This is an example of changing state because of processes external to the agent. In contrast, in a model of FR1 conditioning, an agent may be in an action-available state (indicated by the observation of a lever available to the agent), and the world will remain in the action-available state until the agent takes the action (of pushing the lever), which will move the world into a reward state. For simplicity later, an available action will be written as $S_k \xrightarrow{a_i} S_l$, which indicates that the agent can achieve state S_l if it is in state S_k and selects action a_i . Although the model in this paper is phrased in terms of the agent taking “action” a_i , addicts have very flexible methods of finding drugs. It is not necessary for the model actions to be simple motor actions. $S_k \xrightarrow{a_i} S_l$ indicates the availability of achieving state S_l from state S_k . The agent selects actions proportional to the expected benefit that would be accrued from taking the action; the expected benefit can be determined from the expected change in value and reward (4, 6, 14, 15).

The goal of TDRL is to correctly learn the value of each state. This can be learned by calculating the difference between ex-

pected and observed changes in value (6). This signal, termed δ , can be used to learn sequences that maximize the amount of reward received over time (6). δ is not equivalent to pleasure; instead, it is an internal signal indicative of the discrepancy between expectations and observations (5, 7, 15). Essentially, if the change in value or the achieved reward was better than expected ($\delta > 0$), then one should increase the value of the state that led to it. If it was no different from expected ($\delta = 0$), then the situation is well learned and nothing needs to be changed. Because δ transfers backward from reward states to anticipatory states with learning, actions can be chained together to learn sequences (6). This is the heart of the TDRL algorithm (4–7).

TDRL learns the value function by calculating two equations as the agent takes each action. If the agent leaves state S_k and enters state S_l at time t , at which time it receives reward $R(S_l)$, then

$$\delta(t) = \gamma^d [R(S_l) + V(S_l)] - V(S_k) \quad (2)$$

where γ^d indicates raising the discounting factor γ by the delay d spent by the animal in state S_k (14). $V(S_k)$ is then updated as

$$V(S_k) \leftarrow V(S_k) + \eta_V \delta \quad (3)$$

where η_V is a learning rate parameter.

Phasic increases in dopamine are seen after unexpected natural rewards (16); however, with learning, these phasic increases shift from the time of reward delivery to cuing stimuli (16). Transient increases in dopamine are now thought to signal changes in the expected future reward (i.e., unexpected changes in value) (4, 16). These increases can occur either with unexpected reward or with unexpected cue stimuli known to signal reward (16) and have been hypothesized to signal δ (4, 7, 16). Models of dopamine signaling as δ have been found to be compatible with many aspects of the data (4, 5, 16, 17).

The results simulated below follow from the incorporation of neuropharmacologically produced dopamine into temporal difference models. The figures below were generated from a simulation by using a TDRL instantiation that allows for action selection within a semi-Markov state space, enabling simulations of delay-related experiments (14). The model also produces hyperbolic discounting under normal conditions, consistent with experimental data (12, 13), by a summation of multiple exponential discounting components (14), a hypothesis supported by recent functional magnetic resonance imaging data (18).

The key to TDRL is that, once the value function correctly predicts the reward, learning stops. The value function can be said to compensate for the reward: The change in

Department of Neuroscience, 6-145 Jackson Hall, 321 Church Street SE, University of Minnesota, Minneapolis, MN 55455, USA. E-mail: redish@ahc.umn.edu

value in taking action $S_k \xrightarrow{a_i} S_l$ counterbalances the reward achieved on entering state S_l . When this happens, $\delta = 0$. Taking transient dopamine as the δ signal (4, 5, 7) correctly predicted rewards produce no dopamine signal (16, 17).

However, cocaine and other addictive drugs produce a transient increase in dopamine through neuropharmacological mechanisms (1, 2, 8). The concept of a neuropharmacologically produced dopamine surge can be modeled by assuming that these drugs induce an increase in δ that cannot be compensated by changes in the value (19). In other words, the effect of addictive drugs is to produce a positive δ independent of the change in value function, making it impossible for the agent to learn a value function that will cancel out the drug-induced increase in δ . Equation 2 is thus replaced with

$$\delta = \max\{\gamma^d[R(S_l) + V(S_l)] - V(S_k) + D(S_l), D(S_l)\} \quad (4)$$

where $D(S_l)$ indicates a dopamine surge occurring on entry into state S_l . Equation 4 reduces to normal TDRL (Eq. 2) when $D(S_l) = 0$ but decreases asymptotically to a minimum δ of $D(S_l)$ when $D(S_l) > 0$. This always produces a positive reward-error signal. Thus, the values of states leading to a dopamine surge, $D > 0$, will approach infinity.

When given a choice between two actions, $S_0 \xrightarrow{a_1} S_1$ and $S_0 \xrightarrow{a_2} S_2$, the agent chooses actions proportional to the values of the subsequent states, S_1 and S_2 . The more valuable the state taking an action leads to, the more likely the agent is to take that action. In TDRL, the values of states leading to natural rewards asymptotically approach a finite value (the discounted, total expected future reward); however, in the modified model, the values of states leading to drug receipt increase without bound. Thus, the more the agent traverses the action sequence leading to drug receipt, the larger the value of the states leading to that sequence and the more likely the agent is to select an action leading to those states.

In this model, drug receipt produces a $\delta > 0$ signal, which produces an increase in the values of states leading to the drug receipt. Thus, the values of states leading to drug receipt increase without bound. In contrast, the values of states leading to natural reward increase asymptotically to a value approximating Eq. 1. This implies that the selection probability between actions leading to natural rewards will reach an asymptotic balance. However, the selection probability of actions leading to drug receipt will depend on the number of experiences. Simulations bear this out (Fig. 1).

In the simulations, drug receipt entails a normal-sized reward $R(s)$ that can be com-

pensated by changes in value and a small dopamine signal $D(s)$ that cannot (14). Early use of drugs occurs because they are highly rewarding (1, 3, 20), but this use transitions to a compulsive use with time (1, 3, 20–22). In the model, the $R(s)$ term provides for the early rewarding component, whereas the gradual effect of the $D(s)$ term provides for the eventual transition to addiction. This model thus shows that a transition to addiction can occur without any explicit sensitization or tolerance to dopamine, at least in principle.

The unbounded increase in value of states leading to drug reward does not mean that with enough experience, drugs of abuse are always selected over nondrug rewards. Instead, it predicts that the likelihood of selecting the drug over a nondrug reward will depend on the size of the contrasting nondrug reward relative to the current value of the states leading to drug receipt (Fig. 1).

When animals are given a choice between food and cocaine, the probability of selecting cocaine depends on the amount of food available as an alternative and the cost of each choice (23, 24). Similarly, humans given a choice between cocaine and money will decrease their cocaine selections with increased value of the alternative (25). This may explain the success of vouchers in treatment (25). This will continue to be true even in well-experienced (highly addicted)

subjects, but the sensitivity to the alternate should decrease with experience (see below). This may explain the incompleteness of the success of vouchers (25).

Natural rewards are sensitive to cost in that animals (including humans) will work harder for more valuable rewards. This level of sensitivity is termed elasticity in economics. Addictive drugs are also sensitive to cost in that increased prices decrease usage (26, 27). However, whereas the use of addictive drugs does show sensitivity to cost, that sensitivity is inelastic relative to similar measures applied to natural rewards (26, 28). The TDRL model proposed here produces just such an effect: Both modeled drugs and natural rewards are sensitive to cost, but drug reward is less elastic than natural rewards (Fig. 2).

In TDRL, the values of states leading to natural rewards decrease asymptotically to a stable value that depends on the time to the reward, the reward level, and the discounting factors. However, in the modified TDRL model, the values of states leading to drug rewards increase without bound, producing a ratio of a constant cost to increasing values. This decreasing ratio predicts that the elasticity of drugs to cost should decrease with experience, whereas it should not for natural rewards (fig. S4).

The hypothesis that values of states leading to drug receipt increase without

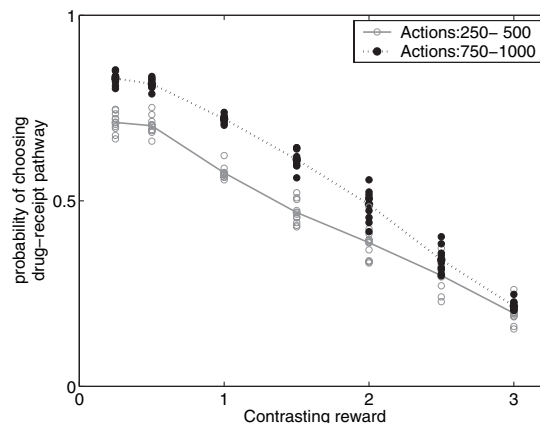


Fig. 1. Probability of selecting a drug-receipt pathway depends on an interaction between drug level, experience, and contrasting reward. Each line shows the average probability of selecting the drug-receipt pathway, $S_0 \xrightarrow{a_2} S_2$, over the contrasting reward pathway, $S_0 \xrightarrow{a_1} S_1$, as a function of the size of the contrasting reward $R(S_2)$. (State space is shown in fig. S1.) Drug receipt on entering state S_4 was $R(S_4) = 1.0$ and $D(S_4) = 0.025$. Individual simulations are shown by dots. Additional details provided in (14).

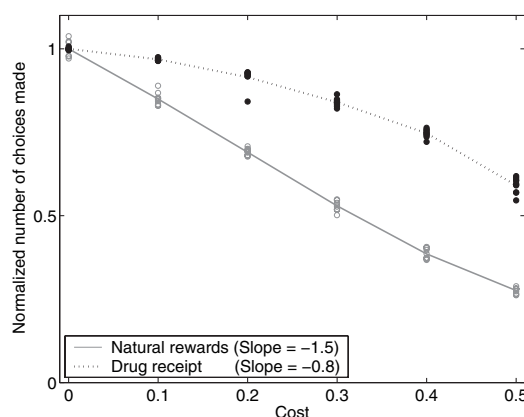


Fig. 2. Elasticity of drug receipt and natural rewards. Both drug receipt and natural rewards are sensitive to costs, but natural rewards are more elastic. Each dot indicates the number of choices made within a session. Sessions were limited by simulated time. The curves have been normalized to the mean number of choices made at zero cost.

bound implies that the elasticity to cost should decrease with use, whereas the elasticity of natural rewards should not. This also suggests that increasing the reward for not choosing the drug [such as vouchers (25)] will be most effective early in the transition from casual drug use to addiction.

The hypothesis that cocaine produces a $\delta > 0$ dopamine signal on drug receipt implies that cocaine should not show blocking. Blocking is an animal-learning phenomenon in which pairing a reinforcer with a conditioning stimulus does not show association if the reinforcer is already predicted by another stimulus (17, 29, 30). For example, if a reinforcer X is paired with cue A , animals will learn to respond to cue A . If X is subsequently paired with simultaneously presented cues A and B , animals will not learn to associate X with B . This is thought to occur because X is completely predicted by A , and there is no error signal ($\delta = 0$) to drive the learning (17, 29, 30). If cocaine is used as the reinforcer instead of natural rewards, the dopamine signal should always be present ($\delta > 0$), even for the AB stimulus. Thus, cocaine (and other drugs of abuse) should not show blocking.

The hypothesis that the release of dopamine by cocaine accesses TDRL systems implies that experienced animals will show a double dopamine signal in cued-response tasks (14). As with natural rewards, a transient dopamine signal should appear to a cuing signal that has been associated with reward (16). However, whereas natural rewards only produce dopamine release if unexpected (16, 17), cocaine produces dopamine release directly (8), thus, after learning both the cue and the cocaine should produce dopamine (Fig. 3). Supporting this hypothesis, Phillips *et al.* (31) found by using fast-scan cyclic voltammetry that, in rats trained to associate an audiovisual signal with cocaine, both the audiovisual stimulus and the cocaine itself produced dramatic increases

in the extracellular concentration of dopamine in the nucleus accumbens.

Substance abuse is a complex disorder. TDRL explains some phenomena that arise in addiction and makes testable predictions about other phenomena. The test of a theory such as this one is not whether it encompasses all phenomena associated with addiction, but whether the predictions that follow from it are confirmed.

This model has been built on assumptions about cocaine, but cocaine is far from the only substance that humans (and other animals) abuse. Many drugs of abuse indirectly produce dopamine signals, including nicotine (10) and heroin and other opiates (11). Although these drugs have other effects as well (1), the effects on dopamine should produce the consequences described above, leading to inelasticity and compulsion.

Historically, an important theoretical explanation of addictive behavior has been that of rational addiction (32), in which the user is assumed to maximize value or “utility” over time, but because long-term rewards for quitting are discounted more than short-term penalties, the maximized function entails remaining addicted. The TDRL theory proposed in this paper differs from that of rational addiction because TDRL proposes that addiction is inherently irrational: It uses the same mechanisms as natural rewards, but the system behaves in a nonoptimal way because of neuropharmacological effects on dopamine. Because the value function cannot compensate for the $D(s)$ component, the $D(s)$ component eventually overwhelms the $R(s)$ reward terms (from both drug and contrasting natural rewards). Eventually, the agent behaves irrationally and rejects the larger rewards in favor of the (less rewarding) addictive stimulus. The TDRL and rational-addiction theories make testably different predictions: Although rational addiction predicts that drugs of abuse will show elasticity

to cost similar to those of natural rewards, the TDRL theory predicts that drugs of abuse will show increasing inelasticity with use.

The rational addiction theory (32) assumes exponential discounting of future rewards, whereas humans and other animals consistently show hyperbolic discounting of future rewards (12, 13). Ainslie (13) has suggested that the “cross-over” effect that occurs with hyperbolic discounting explains many aspects of addiction. The TDRL model used here also shows hyperbolic discounting (14) and so accesses the results noted by Ainslie (13). However, in the theory proposed here, hyperbolic discounting is not the fundamental reason for the agent getting trapped in a nonoptimal state. Rather, the TDRL theory hypothesizes that it is the neuropharmacological effect of certain drugs on dopamine signals that drives the agent into the nonoptimal state.

Robinson and Berridge (22) have suggested that dopamine mediates the desire to achieve a goal (“wanting”), differentiating wanting from the hedonic desire of “liking.” As noted by McClure *et al.* (15), Robinson and Berridge’s concept of incentive salience (22) has a direct correspondence to variables in TDRL: the value of a state reachable by an action. If an agent is in state S_0 and can achieve state S_1 via action $S_0 \xrightarrow{a_i} S_1$ and if state S_1 has a much greater value than state S_0 , then $S_0 \xrightarrow{a_i} S_1$ can be said to be a pathway with great incentive salience. The value function is a means of guiding decisions and thus is more similar to wanting than to liking in the terminology of Robinson and Berridge (15, 22). In TDRL, dopamine does not directly encode wanting, but because learning an appropriate value function depends on an accurate δ signal, dopamine will be necessary for acquisition of wanting.

Many unmodeled phenomena play important roles in the compulsive self-administration of drugs of abuse (1), including titration of internal drug levels (33), sensitization and tolerance (34), withdrawal symptoms and release from them (20), and compensation mechanisms (35, 36). Additionally, individuals show extensive interpersonal variability (37, 38). Although these aspects are not addressed in the model presented here, many of these can be modeled by adding parameters to the model: for example, sensitization can be included by allowing the drug-induced δ parameter $D(s)$ to vary with experience.

TDRL forms a family of computational models with which to model addictive processes. Modifications of the model can be used to incorporate the unmodeled experimental results from the addiction literature. For example, an important question in this model is whether the values of states leading to drug receipt truly increase without bound.

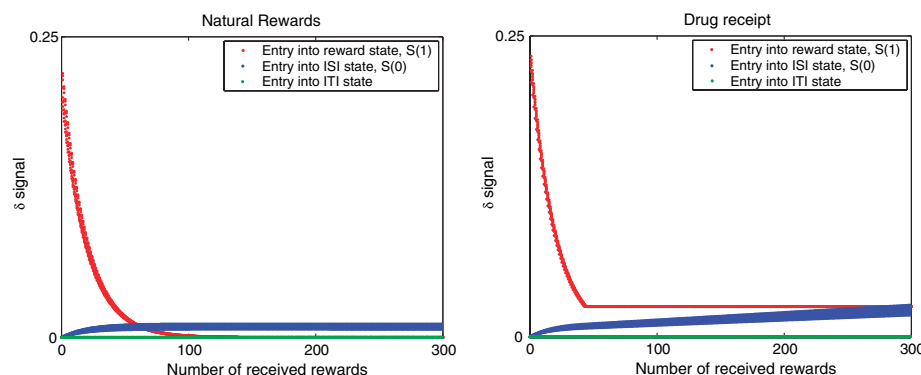


Fig. 3. Dopamine signals. (Left) With natural rewards, dopamine initially occurs primarily at reward receipt (on entry into reward state S_1) and shifts to the conditioned stimulus [on entry into interstimulus-interval (ISI) state S_0] with experience. (State space is shown in fig. S7.) (Right) With drugs that produce a dopamine signal neuropharmacologically, dopamine continues to occur at the drug receipt (on entry into reward state S_1) even after experience, as well as shifting to the conditioned stimulus (on entry into ISI state S_0), thus producing a double dopamine signal.

I find this highly unlikely. Biological compensation mechanisms (35, 36) are likely to limit the maximal effect of cocaine on neural systems, including the value representation. This can be modeled in a number of ways, one of which is to include a global effectiveness-of-dopamine factor, which multiplies all $R(s)$ and $D(s)$ terms. If this factor decreased with each drug receipt, the values of all states would remain finite. Simulations based on an effectiveness-of-dopamine factor that decreases exponentially with each drug receipt (factor = 0.99^n , where n is the number of drug receipts) showed similar properties to those reported here, but the values of all states remained finite.

Another important issue in reinforcement learning is what happens when the reward or drug is removed. In normal TDRL, the value of states leading to reward decay back to zero when that reward is not delivered (6). This follows from the existence of a strongly negative δ signal in the absence of expected reward. Although firing of dopamine neurons is inhibited in the absence of expected reward (16), the inhibition is dramatically less than the corresponding excitation (7). In general, the simple decay of value seen in TDRL (6, 39) does not model extinction very well, particularly in terms of reinstatement after extinction (40). Modeling extinction (even for natural rewards) is likely to require additional components not included in current TDRL models, such as state-space expansion.

A theory of addiction that is compatible with a large literature of extant data and that makes explicitly testable predictions has been deduced from two simple hypotheses: (i) dopamine serves as a reward-error learning signal to produce temporal-difference learning in the normal brain and (ii) cocaine produces a phasic increase in dopamine directly (i.e., neuropharmacologically). A computational model was derived by adding a noncompensable δ signal to a TDRL model. The theory makes predictions about human behavior (developing inelasticity), animal behavior (resistance to blocking), and neurophysiology (dual dopamine signals in experienced users). Addiction is likely to be a complex process arising from transitions between learning algorithms (3, 20, 22). Bringing addiction theory into a computational realm will allow us to make these theories explicit and to directly explore these complex transitions.

References and Notes

1. J. H. Lowinson, P. Ruiz, R. B. Millman, J. G. Langrod, Eds., *Substance Abuse: A Comprehensive Textbook* (Williams and Wilkins, Baltimore, MD, ed. 3, 1997).
2. M. E. Wolf, S. Mangiavacchi, X. Sun, *Ann. N.Y. Acad. Sci.* **1003**, 241 (2003).
3. B. J. Everitt, A. Dickinson, T. W. Robbins, *Brain Res. Rev.* **36**, 129 (2001).
4. P. R. Montague, P. Dayan, T. J. Sejnowski, *J. Neurosci.* **16**, 1936 (1996).
5. W. Schultz, P. Dayan, P. R. Montague, *Science* **275**, 1593 (1997).
6. R. S. Sutton, A. G. Barto, *Reinforcement Learning: An Introduction* (MIT Press, Cambridge, MA, 1998).
7. N. D. Daw, thesis, Carnegie Mellon University, Pittsburgh, PA (2003).
8. M. C. Ritz, R. J. Lamb, S. R. Goldberg, M. J. Kuhar, *Science* **237**, 1219 (1987).
9. M. R. Picciotto, *Drug Alcohol Depend.* **51**, 165 (1998).
10. V. I. Pidoplichko, M. DeBiasi, J. T. Williams, J. A. Dani, *Nature* **390**, 401 (1997).
11. S. R. Laviolette, R. A. Gallegos, S. J. Henriksen, D. van der Kooy, *Nature Neurosci.* **7**, 160 (2004).
12. J. E. Mazur, *Psychol. Rev.* **108**, 96 (2001).
13. G. Ainslie, *Picoeconomics* (Cambridge Univ. Press, New York, 1992).
14. See materials and methods on Science Online.
15. S. M. McClure, N. D. Daw, P. R. Montague, *Trends Neurosci.* **26**, 423 (2003).
16. W. Schultz, *J. Neurophysiol.* **80**, 1 (1998).
17. P. Waelti, A. Dickinson, W. Schultz, *Nature* **412**, 43 (2001).
18. S. C. Tanaka et al., *Nature Neurosci.* **7**, 887 (2004).
19. G. Di Chiara, *Eur. J. Pharmacol.* **375**, 13 (1999).
20. G. F. Koob, M. L. Moal, *Neuropsychopharmacology* **24**, 97 (2001).
21. L. J. M. J. Vanderschuren, B. J. Everitt, *Science* **305**, 1017 (2004).
22. T. E. Robinson, K. C. Berridge, *Annu. Rev. Psychol.* **54**, 25 (2003).
23. M. E. Carroll, S. T. Lac, S. L. Nygaard, *Psychopharmacology (Berlin)* **97**, 23 (1989).
24. M. A. Nader, W. L. Woolverton, *Psychopharmacology (Berlin)* **105**, 169 (1991).
25. S. T. Higgins, S. H. Heil, J. P. Lussier, *Annu. Rev. Psychol.* **55**, 431 (2004).
26. M. E. Carroll, *Drug Alcohol Depend.* **33**, 201 (1993).
27. M. Grossman, F. J. Chaloupka, *J. Health Econ.* **17**, 427 (1998).
28. W. K. Bickel, L. A. Marsch, *Addiction* **96**, 73 (2001).
29. R. A. Rescorla, A. R. Wagner, in *Classical Conditioning II: Current Research and Theory*, A. H. Black, W. F. Prokasy, Eds. (Appleton Century Crofts, New York, 1972), pp. 64–99.
30. A. Dickinson, *Contemporary Animal Learning Theory* (Cambridge Univ. Press, New York, 1980).
31. P. E. Phillips, G. D. Stuber, M. L. A. V. Heien, R. M. Wightman, R. M. Carelli, *Nature* **422**, 614 (2003).
32. G. S. Becker, K. M. Murphy, *J. Polit. Econ.* **96**, 675 (1988).
33. M. Woodward, H. Tunstall-Pedoe, *Addiction* **88**, 821 (1993).
34. C. W. Bradberry, *Neuroscientist* **8**, 315 (2002).
35. S. R. Letchworth, M. A. Nader, H. R. Smith, D. P. Friedman, L. J. Porrino, *J. Neurosci.* **21**, 2799 (2001).
36. F. J. White, P. W. Kalivas, *Drug Alcohol Depend.* **51**, 141 (1998).
37. N. Volkow, J. Fowler, G.-J. Wang, *Behav. Pharmacol.* **13**, 355 (2002).
38. V. Deroche-Gamonet, D. Belin, P. V. Piazza, *Science* **305**, 1014 (2004).
39. R. E. Suri, W. Schultz, *Neuroscience* **91**, 871 (1999).
40. M. E. Bouton, *Biol. Psychiatry* **52**, 976 (2002).
41. This work was supported by an Alfred P. Sloan Fellowship. I thank M. Thomas, E. Larson, M. Carroll, D. Hatsukami, J. Jackson, A. Johnson, N. Schmitzer-Torbert, and Z. Kurth-Nelson for comments on the manuscript and helpful discussions.

Supporting Online Material

www.sciencemag.org/cgi/content/full/306/5703/1944/DC1

Materials and Methods
Figs. S1 to S7

6 July 2004; accepted 19 October 2004
10.1126/science.1102384

The G_s -Linked Receptor GPR3 Maintains Meiotic Arrest in Mammalian Oocytes

Lisa M. Mehlmann,^{1*} Yoshinaga Saeki,^{2†} Shigeru Tanaka,^{2†}
Thomas J. Brennan,^{3‡} Alexei V. Evsikov,⁴ Frank L. Pendola,⁴
Barbara B. Knowles,⁴ John J. Eppig,⁴ Laurinda A. Jaffe^{1*}

Mammalian oocytes are held in prophase arrest by an unknown signal from the surrounding somatic cells. Here we show that the orphan G_s -linked receptor GPR3, which is localized in the oocyte, maintains this arrest. Oocytes from *Gpr3* knockout mice resume meiosis within antral follicles, independently of an increase in luteinizing hormone, and this phenotype can be reversed by injection of *Gpr3* RNA into the oocytes. Thus, the GPR3 receptor is a link in communication between the somatic cells and oocyte of the ovarian follicle and is crucial for the regulation of meiosis.

Meiosis, which reduces the oocyte's chromosome number in preparation for fertilization, begins long before fertilization occurs. In most species, including mammals, DNA replication, entry into meiosis, and chromosomal recombination occur early in oogenesis, but then at late prophase, meiosis arrests. Much later, shortly before ovulation, meiosis resumes: the nuclear envelope breaks down, the chromosomes condense, and a metaphase spindle is formed. In vertebrates, this occurs in response to luteinizing hormone (LH) from the pituitary, which acts on the somatic (granulosa) cells that surround the oocyte in the ovarian follicle (1, 2).

Throughout much of mammalian oogenesis, prophase arrest is maintained by inherent factors in the oocyte and correlates with low levels of activity by cell cycle regulatory proteins, including cyclin B and CDK1 (1). However, once the oocyte reaches its full size and an antral space begins to form between the granulosa cells, prophase arrest in the oocyte becomes dependent on unidentified signals from the granulosa cells. Oocytes that are removed from antral follicles resume meiosis spontaneously (3, 4).

The maintenance of prophase arrest in oocytes within antral follicles requires the activity of signaling molecules within the