Chapter 7
Computational Models of Incentive-Sensitization in Addiction: Dynamic Limbic Transformation of Learning into Motivation

Jun Zhang, Kent C. Berridge, and J. Wayne Aldridge

Abstract Incentive salience is a motivational magnet property attributed to reward-predicting conditioned stimuli (cues). This property makes the cue and its associated unconditioned reward 'wanted' at that moment, and pulls an individual's behavior towards those stimuli. The incentive-sensitization theory of addiction posits that permanent changes in brain mesolimbic systems in drug addicts can amplify the incentive salience of Pavlovian drug cues to produce excessive 'wanting' to take drugs. Similarly, drug intoxication and natural appetite states can temporarily and dynamically amplify cue-triggered 'wanting', promoting binge consumption. Finally, sensitization and drug intoxication can add synergistically to produce especially strong moments of urge for reward. Here we describe a computational model of incentive salience that captures all these properties, and contrast it to traditional cache-based models of reinforcement and reward learning. Our motivation-based model incorporates dynamically modulated physiological brain states that change the ability of cues to elicit 'wanting' on the fly. These brain states include the presence of a drug of abuse and longer-term mesolimbic sensitization, both of which boost mesocorticolimbic cue-triggered signals. We have tested our model by recording neuronal activity from mesolimbic output signals for reward and Pavlovian cues in the ventral pallidum (VP), and a novel technique for analyzing neuronal firing "profile", presents evidence in support of our dynamic motivational account of incentive salience.

Definition Box:
Incentive salience: Also called 'wanting', incentive salience represents motivation for reward (UCS), and is typically triggered in anticipation by a reward-related cue (Pavlovian CS) when the cue is encountered by an individual whose brain mesocorticolimbic circuits are in a highly reactive state (determined by a modulation parameter kappa in our model). Attribution of incentive salience to the cue or reward representations make them more attractive, sought after, and likely to be consumed. Brain mesolimbic systems, especially those involving dopamine, are espe-

J. Zhang · K.C. Berridge (✉) · J.W. Aldridge
Department of Psychology, University of Michigan, Ann Arbor, MI, USA
e-mail: berridge@umich.edu

B. Gutkin, S.H. Ahmed (eds.), *Computational Neuroscience of Drug Addiction*,
Springer Series in Computational Neuroscience 10,
DOI 10.1007/978-1-4614-0751-5_7,

cially important to "wanting." Ordinarily "wanting" occurs together with other reward components of "liking" and learning, but can be dissociated both from other components and subjective desire under some conditions. Incentive salience may occur in the absence of conscious, declarative goal in the ordinary sense of the word wanting. This cognitive form of wanting involves additional cortical brain mechanisms beyond the mesolimbic systems that mediate "wanting" as incentive salience. The difference between conscious desire (want) and incentive salience ('want') can sometimes confer an irrational feature on the excessive urges of sensitized addicts who are not in withdrawal yet still 'want' to take a drug that they know will not give much pleasure.

7.1 Introduction

Incentive salience is a psychological process and neural mechanism to explain the acquisition and expression of motivational values of conditioned stimuli (Berridge and Robinson 1998; Berridge 2007). Incentive salience arises typically as a consequence of reward learning (Bindra 1978; Toates 1986; Berridge 2004), and involves a fundamental dissociation in brain mechanisms of learning, "liking" (hedonic impact or pleasure associated with the receipt of a primary reward) and "wanting" (incentive salience itself; features that makes a stimulus a desirable and attractive goal), see Berridge and Robinson (2003); Robinson and Berridge (2003). Incentive salience is attributed to a sensory stimulus after prior learning of cue-reward associations (between a Pavlovian cue for reward [CS], or the reward itself [UCS]), and transforms it from a sensory representation into a salient and attractive goal representation capable of grabbing the animal's attention and motivating the animal's approach and consumption behaviors.

Beyond learning, physiological brain states relevant to drugs and addiction for the relevant reward, such as activation of mesocorticolimbic dopamine circuits or their regulatory inputs, also modulate attributions of incentive salience on a moment-to-moment basis, in part via alteration in mesolimbic dopamine activation. The incentive salience hypothesis specifically suggests Pavlovian-guided attribution of incentive salience to be dynamically modulated by physiological states that impact NAcc-related circuitry, including dopamine neurotransmission. Regarding addiction, the incentive-sensitization hypothesis suggests that drugs of abuse induce compulsion to take drugs by hijacking neural circuits of incentive salience that evolved to motivate behavior for natural rewards (Robinson and Berridge 1993, 2003, 2008). The hypothesis is not exclusive: it does not deny an important role for drug pleasure or drug withdrawal in motivating drug taking behavior (Koob and Le Moal 2006; Gutkin et al. 2006; Redish et al. 2008). But it suggests that the compulsive and persistent nature of addiction, may be best explained by the concept of a sensitized 'wanting' systems in susceptible individuals, mediated by long-term neuroadaptations that may involve alterations in gene expression, neurotransmitter release and receptor levels, and dendritic sprouting patterns in mesocorticolimbic

structures. Incentive-sensitization can create addiction even to drugs that are not particularly pleasant, and still produce cue-triggered relapse back into drug taking even long after recovery from withdrawal.

The relation to mesocorticolimbic modulation makes incentive salience particularly influenced by natural appetite states, by psychostimulant drugs that promote dopamine and by enduring neural sensitization of mesolimbic NAc-VP systems. Previous computational models have suggested that incentive salience can be construed purely by reinforcement learning mechanisms such as the temporal difference or prediction error model (McClure et al. 2003; Redish 2004). Such models account for incentive salience in terms of dopamine-based learning mechanisms (Schultz et al. 1997; Schultz 2002), without invoking a role for physiological modulation of motivation after learning. For example, McClure et al. identified incentive salience with the "(state) value function" or expected total discounted reward. Redish (2004) identified drug sensitization with a mechanism that amplifies the temporal difference prediction error signal itself. This is different from our view, which posts incentive salience to be dynamically modulated from moment to moment, based on inputs from current physiological/brain states as well as from learned associations to a reward cue (Zhang et al. 2009). In this chapter, we first review those standard learning models, and then contrast them to data that indicate incentive salience involves more than merely learning. Namely, cue-triggered 'wanting' also involves an additional physiological factor that dynamically transforms transforms static learned values into a flexible level of motivation appropriate to the moment (Tindell et al. 2009). Rather than simply reflecting a previously-learned value, our model proposes a specific gain-control mechanism for modulating on the fly the expected values of reward-predicting stimuli to dynamically compute incentive salience.

7.1.1 Dopamine and Reinforcement Learning: The "Standard" Model and Critique

Contemporary reinforcement learning theory posits an actor-critic architecture for reward prediction and action control. The critic computes the error in reward-prediction—the discrepancy between the reward expected from a stimulus (technically, a state) and the reward actually received. The temporal-difference (TD) method provides an explicit formula for calculating such expected reward through incorporating the subsequent prediction made by the same reward-predicting system as a part of predicted reward of the current state, thereby allowing a refined estimate of the value of a state in the sequential context. The actor, on the other hand, evaluates the merits of policies and selects for each state an action associated with highest long-term reward values. Critic and actor are often discussed as potential functions of ventral striatum and its mesolimbic dopamine inputs from ventral tegmental area (VTA) and substantia nigra pars compacta (SNc), respectively (O'Doherty et al. 2004).

There is growing consensus (though not without controversy, see Redgrave et al. 1999; Redgrave and Gurney 2006), that the predictive error signal, which lies

at the core of temporal difference learning, is carried by the firing of dopaminergic neurons projecting to nucleus accumbens and neostriatum (Schultz 1998; Schultz et al. 1997). Phasic firing in midbrain dopaminergic neurons (Schultz et al. 1997; Schultz 1998) has been suggested to express the TD error in reward prediction. Such signal has also been posited to update working memory and goal stack representations in prefrontal cortex (Montague et al. 2004; O'Reilly 2006), consistent with neuropsychological proposals that the prefrontal cortex (PFC) controls goals and goal-directed action (Miller and Cohen 2001).

The actor-critic architecture and TD-based learning rule derives its computational power from its consistent and effective scheme for optimizing sequential decision-making in a stationary Markov environment (e.g., Puterman 1994), without the need of an elaborate model of the world. However, an important missing piece for this framework is how learned predictions are used to generate motivation on a moment-to-moment basis in a way that incorporates the current physiological state such as hunger, satiety, psychostimulant sensitization, or the immediate impact of drugs. This was pointed out by Dayan (2009) and Dayan and Balleine (2002). Attempts to grapple with this issue have been made via tree-search devaluations of goals (Daw et al. 2005a, 2005b) or via satiation decrements that reduce motor arousal or limit generalization (Niv et al. 2006). A problem that has remained unaddressed is how the motivation value of specific reward stimuli may be dynamically *increased* in targeted fashion by physiological states involving mesolimbic activation, including neural sensitization states relevant to addiction (Robinson and Berridge 2003; Berridge 2007).

Going beyond the act-outcome ("A-O") devaluation that can be successfully modeled by the cognitively-based tree-search mechanism (Daw et al. 2005a, 2005b) the concept of incentive salience posits an additional Pavlovian-based cue-triggered ("S-S") motivation process (i.e., incentive salience attribution), which can be either increased or decreased phasically by physiological hunger-satiety states that are relevant to the brain's calculation of hedonic value for particular sensory goals (for example, sweet versus salty tastes during caloric hunger versus salt appetite) (Berridge and Valenstein 1991; Robinson and Berridge 1993; Berridge and Robinson 1998; Tindell et al. 2006). Incentive salience takes Pavlovian associations as its primary learned input, but also takes a separate input in the form of current mesolimbic states that amplify or dampen 'wanting' for specific cues and their rewards. Further, the incentive-sensitization theory of addiction posits that mesolimbic activation, drugs of abuse and persisting sensitization can tap into those motivation-amplifying brain circuits to incrementally raise the incentive salience carried by particular reward stimuli, so that cues may trigger compulsive motivation for their rewards by the same S-S incentive modulation mechanism (Robinson and Berridge 1993, 2003) The incentive salience attribution mechanism is thus essentially Pavlovian-guided, but calculates value anew on a moment-by-moment basis that is influenced by mesolimbic dopamine states (see Berridge 2001 chapter). This view assigns a very different role to dopamine function from traditional learning models, but one that we suggest below can be made compatible with the existing computational theories of reinforcement learning.

7.2 Previous Computational Approaches to Learning and Incentive Salience

Several models of dopamine function with respect to computation of incentive salience and incentive sensitization have been proposed, including our own. They are all anchored on the now-standard reinforcement learning framework along with the temporal difference (TD) learning method.

Reinforcement learning theory provides a mathematical framework to model action control and reward prediction by an agent in a stable Markov environment. In temporal difference models (Sutton and Barto 1981), the expected total future discounted reward V associated with an environmental state s (i.e., the conditioned stimulus [CS] associated with reward) is

$$V(s_t) = \left\langle \sum_{i=0} \gamma^i r_{t+i} \right\rangle = \langle r_t \rangle + \gamma \langle r_{t+1} \rangle + \gamma^2 \langle r_{t+2} \rangle + \cdots, \tag{7.1}$$

where $\gamma \in [0, 1)$ is the discount factor, $r_t, r_{t+1}, r_{t+2}, \ldots$, representing the sequence of primary rewards (UCS) starting from the current state (subscripted t, predictive CS), and the expectation $\langle \cdot \rangle$ is taken over generally stochastic state transition and reward delivery. The estimated value of reward prediction $\hat{V}$ (denoted with a hat) is a cached value that becomes gradually established through temporal difference learning over past instances in which r and s are paired. On each trial, specifically, a prediction error δ concerning deviation from consistent successive predictions is calculated, based on instantaneous reward r_t (which might be stochastic)

$$\delta(s_t) = r_t + \gamma \hat{V}(s_{t+1}) - \hat{V}(s_t), \tag{7.2}$$

and is used to update $\hat{V}$ via $\delta \hat{V}(s_t) \propto \delta(s_t)$. After learning has completed, $\delta(s_t) = 0$, so

$$\hat{V}(s_t) = \langle r_t \rangle + \gamma \hat{V}(s_{t+1}). \tag{7.3}$$

In the early application to 'wanting' mentioned above, McClure et al. (2003) proposed that the notion of incentive salience be mapped directly to the computational concept of total expected future discounted reward, namely V (see Eq. (7.1)). In TD learning theory, V is a cached, incrementally-learnt value function. However, a difficulty arises from identifying incentive salience with the value function V, as V is usually defined in TD models. That difficulty is that V can change if a reward is revalued only after further relearning about the new prediction error introduced by re-encounters with the revalued reward. Thus, to change incentive salience of a CS requires further pairing with its revalued UCS, according to such a pure learning model based on re-training via new prediction errors. That contradicts our idea described above that CS incentive salience is also modulated on the fly by relevant physiological states that alter mesolimbic function, which produces a synergistic interaction between prior learning and current mesolimbic reactivity in determining the current level of CS-triggered motivation (Robinson and Berridge 1993; Berridge and Robinson 1998).

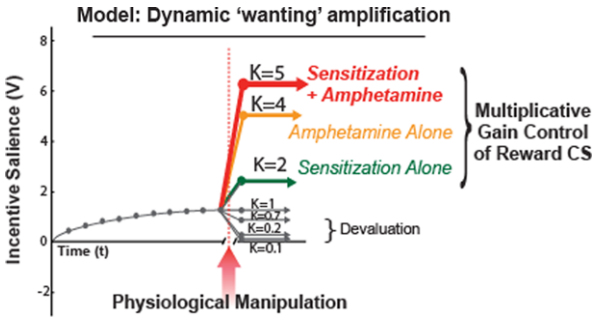

Fig. 7.1 Simulations of dynamic shifts in incentive salience. Initial learning is assumed to proceed by a TD type of rule initially. At time step $t = 11$, a new mesolimbic activation introduced via amphetamine administration, sensitization, or both. The change in incentive salience occurs as indicated by the *arrows*, multiplicatively ($V \cdot \kappa$). Modified from Zhang et al. (2009)

A variant on prediction-error approach applied to addiction has been to posit that sensitization elevates learning itself (e.g., Redish 2004). For example, by magnifying the drug-elicited prediction error signal carried by dopamine neurons, this has been suggested to result in a surge in the value of δ itself (Redish 2004). Such a change would then induce an increase in the learned value function V via the standard, δ-driven TD learning mechanism on the next trial that paired the initial CS with the UCS. An exaggerated TD error signal, repeated again and again in drug addicts who continue to take drugs, has been posited to increase V without upper bound, hence explaining addiction as over-learning of V prediction (Redish 2004). Such ideas are elegant applications of TD learning theory to addiction, but nevertheless rely exclusively on the assumption that dopamine causes a predictive error that functions as a teaching signal in the TD framework. There are reasons to question that assumption (e.g., Berridge 2007).

7.3 Our Dynamic Model of Incentive Salience: Integrating Learning with Current State

To incorporate motivational modulation of previously learned values (Fig. 7.1), we propose that the *incentive salience* or motivational value $\tilde{V}(s_t)$ of a reward-predicting CS be described as

$$\tilde{V}(s_t) = \tilde{r}(r_t, \kappa) + \gamma V(s_{t+1}), \tag{7.4}$$

where the value of the UCS (i.e., the primary reward value r_t) is modulated by a factor κ reflecting current physiological state, such as hunger, thirst, salt appetite, amphetamine administration, drug sensitization, etc. Two specific forms of r were postulated in Zhang et al. (2009): an additive mechanism and a multiplicative mechanism. Only the multiplicative mechanism is required regarding addiction in terms

of drug activation of dopamine systems and long-term sensitization of those systems, and so in this chapter, we concentrate on the multiplicative form

$$\tilde{r}(r_t, \kappa) = \kappa \langle r_t \rangle. \tag{7.5}$$

In our model, the multiplicative constant κ can be specific to a particular appetitive system. Incentive salience can be either increased or decreased, with $\kappa < 1$ representing decrease (such as satiation or devaluation) and $\kappa > 1$ representing increase (such as hunger or sensitization). Equation (7.4) along with (7.5) suggests that moment-to-moment evaluation of the incentive value of the goal associated with the current state s_t, is contributed to by two parts: a gain-controlled evaluation of the immediately available reward (first term), and a γ-discounted evaluation of future rewards based on stored prediction value (second term). Physiological state κ factors may couple with geometric discounting under γ, in that satiation ($\kappa < 1$) may increase the temporal horizon γ, whereas sensitization or an increased physiological appetite (κ becomes greater than 1) may decrease γ, and disproportionately raise the motivational value of temporal proximity to reward UCS (see Giordano et al. 2002). The incentive value of a state s_t is the motivationally-modulated value of the immediate reward r_t plus the discounted value of the expected reward in the next state s_{t+1}; both are loaded into the goal representation as s_t is presented.

The multiplicative modulation (i.e., the calculation of $\tilde{V}$) is the key distinction from learning-based models of incentive salience—it corresponds to model-based action control via goal representation versus model-free action control using stored predictions. Subtracting both sides of Eq. (7.5) from those of Eq. (7.3), and substituting in the multiplicative relation (7.5), we obtain

$$\tilde{V}(s_t) - V(s_t) = (\kappa - 1)\langle r_t \rangle, \tag{7.6}$$

that is, the incentive salience $\tilde{V}(s_t)$ reduces to $V(s_t)$ in the absence of devaluation/sensitization manipulation ($\kappa = 1$). We believe, however, that the κ parameter does not act as an indiscriminate tide that floats all boats to raise the incentive salience of all CSs equally. Rather, specific physiological appetite states, such as drug addiction, caloric hunger or salt appetite, each amplify the hedonic value of their own reward (drugs, sweets, salty foods), and hence specifically amplifies the incentive salience of particular CSs related to that UCS (Bindra 1978; Toates 1986; Berridge 2001, 2004; Dickinson and Balleine 2002), presumably each modulated by its own κ parameters.

It follows from Eqs. (7.1), (7.4) and (7.5) that

$$\tilde{V}(s_t) = \kappa \langle r_t \rangle + \gamma \left(\left\langle \sum_{i=0} \gamma^i r_{t+1+i} \right\rangle \right). \tag{7.7}$$

In essence, Eq. (7.7) is an expression of what is known as the "quasi-hyperbolic" discounting model (Laibson 1997; Frederick et al. 2002). Hence, our model provides an incentive salience explanation of why mesolimbic NAc-VP systems may sometimes be activated by an immediately available reward more than by temporally distant reward (McClure et al. 2004), and suggests that the degree of discounting in such situations will be modulated by current mesolimbic states.

Note that our postulated gain-control mechanism effectively modulates the tradeoff of immediate primary reward and future expected reward. Recent neuro-imaging studies implicating ventral striatum and parietal cortex in mediating the relative weighting of immediate versus future rewards in humans (McClure et al. 2004; Glimcher and Kable 2005) and in monkeys (Louie and Glimcher 2005), with mesolimbic neural activation specifically potentiating the motivational value of short-term rewards at the expense of long-term rewards, are consistent with our proposed gain-control function for dopamine.

In short, our model proposes that incentive salience requires (1) an online gain control (gating) mechanism κ that dynamically modulates CS reward value according to changes in reward-relevant physiological or neurobiological states, including mesolimbic dopamine activation or sensitization, along with (2) potential adjustment of the temporal horizon γ for evaluating stored prediction values; both motivational consequences are adjustable on-the-fly without the need for (re)learning.

7.4 Testing Model Predictions Using a Serial Conditioning Task

The above-mentioned models of incentive salience and mesolimbic dopamine function were teased apart previously by our colleagues and us in an electrophysiological recording study that employed post-learning mesolimbic modulation in a Pavlovian conditioning task (Tindell et al. 2004). Let us consider the serial conditioning paradigm involving two conditioned stimuli in sequence, CS1 and CS2, that predict a terminal reward UCS: the full series is CS1 $\rightarrow$ CS2 $\rightarrow$ UCS (Fig. 7.2). After the rat is trained on this sequential paradigm, it learns the values associated with V_1 (of CS1) and V_2 (of CS2), as well as with the value r of the terminal reward. In later tests, the rat's mesolimbic systems may be activated by amphetamine administration; or between the training and the test, the rat may be sensitized by a hefty regimen of exposure to psychostimulant drugs. In all test cases, the first CS still predicts all following stimuli, and because of temporal discounting, their magnitude will be in descending order: $V_1 < V_2 < r$. So a pure TD value-coding model would predict that neuronal coding of incentive salience should follow the same ordering, with activation to UCS being the largest. Under sensitization manipulation, the primary UCS reward values will be magnified, $r \rightarrow \kappa r$.

A TD error model by comparison would predict that, after learning is complete in the CS1/CS2/UCS paradigm, $\delta_2 = \delta_3 = 0$, whereas $\delta_1 > 0$. Allowing the possibility for incomplete learning, one still has $\delta_1 > \delta_2 > \delta_3$, where the ordering reflects the propagation of learning gradient from reward-distal to reward-proximal direction. Assuming the effect of sensitization or acute amphetamine challenge to be either additive or multiplicative on the existing δ signal, it follows from the above line of reasoning that the response of TD error-coding neurons to CS1 would be the strongest, though it would not appear until after a new learning trial once the drug elevated δ. In short, a prediction error coding model (e.g., Redish 2004) specifies increments in the neural code for δ, most prominently for CS1. That is in stark contrast to the specification (by our incentive salience model below) of CS2 as the stimulus

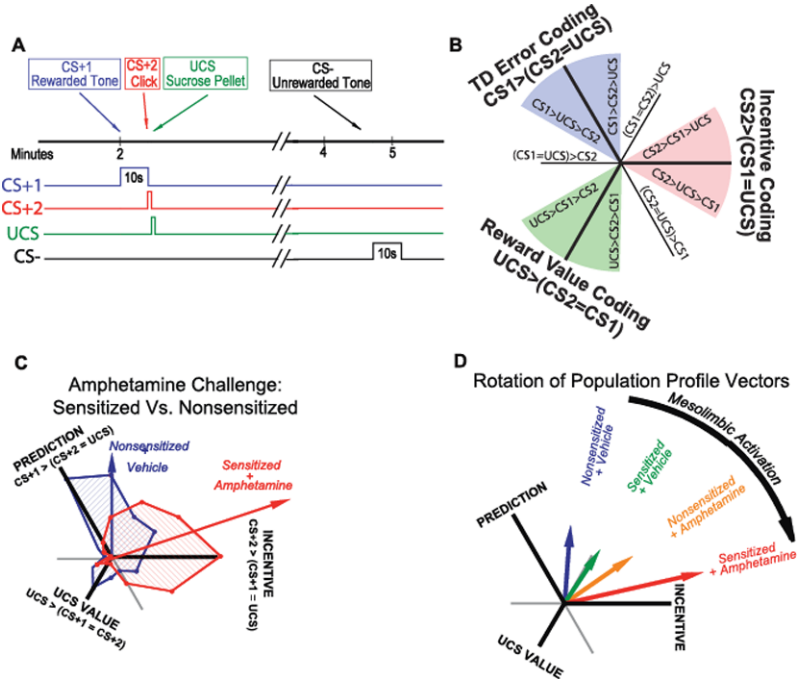

Fig. 7.2 Selective amplification of CS incentive salience (not CS prediction or UCS hedonic impact) by transient amphetamine intoxication and more permanent drug sensitization. Experimental design of the serial CS1/CS2/UCS procedure, and effects of sensitization and amphetamine on neuronal firing profiles in ventral pallidum (**A**). The relative rank-ordering of neuronal responses to CS1/CS2/UCS is defined as the "profile" of a neuron; it can be represented mathematically as the angle of a vector in a two-dimensional space, where the two axes represent two orthogonal contrasts formed from the three responses (**B**). The computation is such that this angular value indexing a response profile exists in a continuum which (1) exhausts all possible firing patterns (i.e., relative orders in firing rates to these three types of stimuli); and (2) guarantees that nearby values represents similar firing patterns. Temporal difference error-coding implies maximal response to CS1 which has the greatest prediction, whereas value-coding implies maximal firing to UCS which has the highest hedonic value. By contrast, incentive-coding implies maximal firing to CS2 that has the greatest motivational impact as it immediately precedes the actual reward. The data panel shows firing in control condition contrasted to the combination of amphetamine plus sensitization (**C**). The summary *arrow panel* shows the averaged neuronal response for each group of rats, illustrating the additive increments produced by sensitization, amphetamine and combination of both (**D**). From Zhang et al. (2009) and modified from Tindell et al. (2005)

most enhanced in incentive salience by a neurobiological activation of mesolimbic systems.

By contrast, our gain-control κ model of incentive salience predicts that mesolimbic activation, even if occurring after learning (e.g., by psychostimulant sensitization or by acute amphetamine administration), may immediately modulate the neuronal computation of CS incentive salience (especially for CS2). In the serial conditioning paradigm, $r_1 = 0$ and $r_2 = r$, so the motivationally controlled incen-

tive salience value is $\tilde{V}_1 = \gamma r$, $\tilde{V}_2 = \kappa r$. Pan et al. (2005) showed that on certain trials when the CS2 is omitted after CS1 there is a "dip" in the firing of dopamine neurons in substantial nigra and VTA. This suggests that following CS1, the animal is indeed expecting CS2 rather than UCS. (Because the temporal discount factor $\gamma < 1$ and sensitization manipulation $\kappa > 1$, $\tilde{V}_1 < \tilde{V}_2$.) In short, with respect to ordering of magnitudes, our model of incentive salience anticipates that CS2 should receive greater motivational impact than CS1, because it is closer in time to UCS, facilitating transfer of incentive properties directly from UCS to CS2 (see Tindell et al. 2005 for supporting evidence regarding relative roles of CS1 and CS2). The reward-proximal CS2 should be most potently enhanced by either persisting neural sensitization induced after learning, or acute amphetamine administration on the day of test, because both manipulations activate mesolimbic dopamine systems.

7.4.1 VP Neuronal Coding for Incentive Salience

These models were tested in a study that focused on mesolimbic output signals that were intercepted in the ventral pallidum (Tindell et al. 2005). The ventral pallidum (VP) lies at the base of the brain behind the nucleus accumbens and in front of ventral tegmental area (VTA) and lateral hypothalamus, and serves as a final common path for mesocorticolimbic circuits. The VP processes compressed representations of mesocorticolimbic reward signals before relaying them back up into corticolimbic and mesocorticolimbic loops, and downward to motor structures (Zahm 2000; Zahm 2006). Regarding mesolimbic dopamine inputs, VP also receives direct dopamine projections from VTA that has been implicated in drug reward (Gong et al. 1996; McFarland et al. 2004), as well as most efferent projections from the nucleus accumbens. Its outputs project upward to mediodorsal thalamus and thence back to corticolimbic loops involving prefrontal, cingulate, and insular cortical areas, and downwards to subcortical sites. VP neurons fire to Pavlovian conditioned stimuli (CS+) that predict rewards as well as to reward UCSs themselves (Tindell et al. 2004). Hence, VP is a prime candidate to study mesolimbic modulations of CS incentive coding.

The Tindell et al. (2005) study tested whether mesolimbic activation by sensitization or pharmacological dopamine release enhances VP firing that codes incentive salience as a motivational transform of CS+ in a manner specified by our model of incentive salience computation, as opposed to pure learning-based models. Rats were trained on a serial Pavlovian conditioning paradigm, CS1/CS2/UCS, where a Pavlovian CS1 (a tone) followed after a 10 second delay by a CS2 (a click) predicted immediate (i.e., within 1 second) reward in the form of a sucrose pellet (UCS); they were also trained on a negative CS− (another tone) that did not lead to any reward. After two weeks of Pavlovian training, physiological activation of dopamine-related mesolimbic brain systems was induced in either or both of the following two ways: by neural sensitization caused by repeated, pulsed psychostimulant drug administration followed by a prolonged drug-free incubation period ("sensitization condition"), or later during testing sessions by acute amphetamine administration that

immediately causes dopamine to be released into synapses ("amphetamine challenge condition"). Recordings were done during testing sessions. An offline review of the monitoring video ruled out that VP neuron activity in response to CS2 and UCS presentation were due simply to pure movements.

Each neuronal unit's responses to the three stimuli (CS1, CS2, UCS) were analyzed by a novel data analysis and presentation method, called Profile Analysis (Tindell et al. 2005; Zhang et al. 2009), to derive a profile direction (an angle within 360°). A total of 524 recorded units show that VP neuronal response profiles were broadly dispersed, covering the entire spectrum of value-coding, predictive error-coding, and incentive-coding regions of the profile space. Histograms (polar plot) were constructed by plotting the number of units at each directional angle versus the angular values themselves. Population averages of such profile vectors (i.e., vector sum called a Population Profile Vector), were also plotted for VP neurons under various conditions including normal control, sensitization and/or acute amphetamine conditions. Normally, VP neurons signaled prediction-error preferentially, responding maximally to CS1, secondarily to CS2, and least to UCS. But mesolimbic dopamine activations enhanced incentive salience computations on the fly. Sensitization (Fig. 7.2) and acute amphetamine (Fig. 7.2) both shifted the distributions of response profiles away from predictive error coding (CS1-maximal response) and toward incentive coding (CS2-maximal response). The greatest shift occurred when rats were both pre-sensitized and exposed to acute amphetamine challenge on the same test day (Fig. 7.2).

The effects of mesolimbic dopaminergic activation can be visualized as the rotation of the Population Profile Vectors away from a CS1-maximal/prediction-coding axis (CS1) and towards the CS2-maximal/incentive-coding axis (CS2) (Fig. 7.2). Thus, it can be concluded that while VP neurons in control animals (after training) tend to follow a TD error coding profile, mesolimbic dopaminergic activation causes the neuronal response profiles to shift towards encoding incentive salience. Mesolimbic activation by sensitization, amphetamine administration, or both, specifically and progressively caused VP neurons to increase their firing rates predominantly to CS2, compared with CS1 or UCS. Such results are anticipated by our motivational-based model of incentive salience (with $\kappa > 1$).

7.5 Discussion

Reward cues trigger motivation "wanting", as well as hedonic affects, cognitive expectations and procedural habits. Incentive salience theory posits the motivational value triggered by a CS+ to be based on two separate but integrated inputs: (a) current physiological/neurobiological state; and (b) previously learned associative values. This integration of physiological signals allows drug states, sensitization states, or natural hunger, thirst and other appetitive states to immediately enhance the incentive salience attributed to a previously learned CS+ for relevant reward, without necessarily requiring additional re-learning trials.

To summarize, our analysis supports a computational account of incentive salience as a motivational gain control mechanism that dynamically responds to post-learning shifts in physiological states when computing 'wanting' triggered by a relevant CS for reward. This gain control mechanism modulates motivation on a moment-by-moment basis as brain states vary, gauging the relative importance (tradeoff in values) between primary reward versus expected future reward. Finally, VP circuits, as a crucial node in mesocorticolimbic circuits, may be an important stage in computing the motivational transforms of CS and UCS values alike.

7.5.1 Multiple Motivation-Learning Systems

We stress that other types of learning and motivation exist aside from Pavlovian incentive salience: in particular, cognitive incentives and reward-oriented habits. For example, evidence described elsewhere indicates that 'wanting' (with quotation marks: incentive salience) exists alongside ordinary wanting (without quotation marks: cognitive predictions), which may plausibly be based on full look-ahead cognitive representations of expected goal values and their related act-outcome strategies to obtain those goals (Dayan and Balleine 2002; Dickinson and Balleine 2002; Berridge and Robinson 2003). Ordinarily, wanting and 'wanting' act together to guide behavior toward the same goals, with incentive salience serving to add motivation 'oomph' to cognitive representations. But under some important conditions cognitive and Pavlovian motivation mechanisms may diverge. For example, divergence can lead to 'irrational wanting' in addiction for a target that the individual does not cognitively want, nor predicatively expect to be of high value. Our current model may help to computationally capture the Pavlovian incentive salience limb of that divergence (Berridge and Aldridge 2008).

7.5.2 Contrasting Dynamic Incentive Salience to Cognitive Tree Goals

Loosely speaking, our model could be considered similar to one-step look-ahead in a model-based (tree-search) approach. However, there are important differences between our model and most tree-search models. A full tree-model is usually thought to have an advantage of providing a stable cognitive map of declarative goals and available actions within the tree representation of the world. Our model nevertheless posits a dynamic synergy between current mesolimbic reactivity and the presence of a cue (with its previously acquired association to reward). For example, cue-triggered 'wanting' shoots up upon presentation of a CS, but importantly, also goes down again nearly as soon as the CS is taken away—even when the brain remains in a mesolimbic-activated state (e.g., after amphetamine administration; after sensitization; or after combination of both). Coupling of incentive salience to CS is

evident in behavioral cue-triggered 'wanting' experiments (Pavlovian instrumental transfer), where lever-pressing peaks fade away as soon as the CS is removed—even though the dopamine drug or sensitization state that enhanced the cue's motivation-eliciting power persist.

This type of transience is quite typical of motivational states. In particular, the incentive salience mechanism is especially compatible with transient peaks in 'wanting' being tied to CS presence because the rules that underlie Pavlovian controls of incentive salience specify that a synergy exists between CS presence and current mesolimbic state (Robinson and Berridge 1993; Berridge 2007). The physical presence of a Pavlovian CS is a crucial factor in generating incentive salience, and a sporadic CS can lead to up-and-down changes in 'wanting'. This synergy feature is precisely why a drug CS triggers relapse in an addict as a phasic peak of temptation—at least if that CS is encountered in a mesolimbic-activated state.

Our model for the computation of incentive salience implies the motivational magnet property of a drug reward cue is dynamically recomputed based on current physiological states of sensitization and drug intoxication. This dynamic amplification of motivation in addicts may maladaptively pull the addict like a magnet towards compulsively 'wanted' drugs, and so make it harder to escape from the addiction.

Acknowledgements Collection of experimental data that gave rise to this computational model was supported by NIH grants DA017752, DA015188 and MH63649. The writing of this book chapter was also supported by AFOSR grant FA9550-06-1-0298. We thank Dr. Michael F.R. Robinson for helpful comments on an earlier version of the manuscript.

References

Berridge KC (2001) Reward learning: Reinforcement, incentives, and expectations. In: Medin DL (ed) The psychology of learning and motivation, vol 40. Academic Press, New York, pp 223–278

Berridge KC (2004) Motivational concepts in behavioral neuroscience. Physiol Behav 81:179–209

Berridge KC (2007) The debate over dopamine in reward: the case for incentive salience. Psychopharmacology 191:391–431 (2007)

Berridge KC, Aldridge JW (2008) Decision utility, the brain, and pursuit of hedonic goals. Social Cogn 26:621–646

Berridge KC, Robinson TE (1998) What is the role of dopamine in reward: hedonic impact, reward learning, and incentive salience? Brains Res Rev 28:309–369

Berridge KC, Robinson TE (2003) Parsing reward. Trends Neurosci 26:507–513

Berridge KC, Valenstein ES (1991) What psychological process mediates feeding evoked by electrical stimulation of the lateral hypothalamus? Behav Neurosci 105:3–14

Bindra D (1978) How adaptive behavior is produced: a perceptual-motivation alternative to response reinforcement. Behav Brain Sci 1:41–91

Daw ND, Niv Y, Dayan P (2005a) Uncertainty-based competition between prefrontal and dorsal striatal systems of behavioral control. Nat Neurosci 8:1704–1711

Daw ND, Niv Y, Dayan P (2005b) Actions, policies, values, and the basal ganglia. In: Bezard (ed) Recent breakthroughs in basal ganglia research. Nova Publ, New York, pp 91–106

Dayan P (2009) Dopamine, reinforcement learning, and addiction. Pharmacopsychiatry 42(S 01):S56–S65

Dayan P, Balleine BW (2002) Reward, motivation and reinforcement learning. Neuron 36:285–298

Dickinson A, Balleine B (2002) The role of learning in the operation of motivational systems. In: Gallistel CR (ed) Stevens' handbook of experimental psychology: learning, motivation, and emotion, vol 3, 3rd edn. Wiley, New York, pp 497–534

Frederick S, Loewenstein G, O'Donoghue T (2002) Time discounting and time preference: A critical review. J Econ Lit 40:351–401

Giordano LA, Bickel WK, Loewenstein G, Jacobs EA, Marsch L, Badger GJ (2002) Mild opioid deprivation increases the degree that opioid-dependent outpatients discount delayed heroin and money. Psychopharmacology 163:174–182

Glimcher PW, Kable O (2005) Neural mechanisms of temporal discounting in humans. Abstract for 2005 annual meeting of the society for neuroeconomics

Gong W, Neill D, Justice JB Jr (1996) Conditioned place preference and locomotor activation produced by injection of psychostimulants into ventral pallidum. Brain Res 707(1):64–74

Gutkin BS, Dehaene S, Changeux JP (2006) A neurocomputational hypothesis for nicotine addiction. Proc Natl Acad Sci USA 103(4):1106–1111

Koob GF, Le Moal M (2006) Neurobiology of addiction. Academic Press, New York

Laibson D (1997) Golden eggs and hyperbolic discounting. Q J Econ 112(2):443–477

Louie K, Glimcher PW (2005) Intertemporal choice behavior in monkeys: interaction between delay to reward, subjective value, and area LP. Abstract for 2005 annual meeting of the society for neuroeconomics

McClure SM, Daw ND, Montague PR (2003) A computational substrate for incentive salience. Trends Neurosci 26:423–428

McClure SM, Laibson DI, Loewenstein G, Cohen JD (2004) Separate neural systems value immediate and delayed monetary rewards. Science 306:503–507

McFarland K, Davidge SB, Lapish CC, Kalivas PW (2004) Limbic and motor circuitry underlying footshock-induced reinstatement of cocaine-seeking behavior. J Neurosci 24(7):1551–1560

Miller EK, Cohen JD (2001) An integrative theory of prefrontal cortex function. Annu Rev Neurosci 24:167–202

Montague PR, Hyman SE, Cohen JD (2004) Computational roles for dopamine in behavioral control. Nature 760–767

Niv Y, Joel D, Dayan P (2006) A normative perspective on motivation. Trends Cogn Sci 10:375–381

O'Doherty J, Dayan P, Schultz J, Deichmann R, Friston K, Dolan RJ (2004) Dissociable roles of ventral and dorsal striatum in instrumental conditioning. Science 304:452–454

O'Reilly RC (2006) Biologically based computational models of high-level cognition. Science 314:91–94

Pan W-X, Schmidt R, Wickens JR, Hyland BI (2005) Dopamine cells respond to predicted events during classical conditioning: evidence for eligibility traces in the reward-learning network. J Neurosci 25:6235–6242

Puterman ML (1994) Markov decision processes. Wiley, New York

Redgrave P, Gurney K (2006) The short-latency dopamine signal: a role in discovering novel actions? Nat Rev, Neurosci 7:967–975

Redgrave P, Prescott TJ, Gurney K (1999) Is the short-latency dopamine response too short to signal reward error? Trends Neurosci 22:146–151

Redish AD (2004) Addiction as a computational process gone awry. Nature 306:1944–1947

Redish AD, Jensen S, Johnson A (2008) A unified framework for addiction: Vulnerabilities in the decision process. Behav Brain Sci 31(4):415–437; discussion 437–487

Robinson TE, Berridge KC (1993) The neural basis of drug craving: an incentive-sensitization theory of addiction. Brains Res Rev 18:247–291

Robinson TE, Berridge KC (2003) Addiction. Annu Rev Psychol 54:25–53

Robinson TE, Berridge KC (2008) The incentive sensitization theory of addiction: some current issues. Philos Trans R Soc Lond B Biol Sci 363(1507):3137–3146

Schultz W (1998) Predictive reward signal of dopamine neurons. J Neurophysiol 80:1–27

Schultz W (2002) Getting formal with dopamine and reward. Neuron 36:241–263
Schultz W, Dayan P, Montague PR (1997) A neural substrate of prediction and reward. Science 275:1593–1599
Sutton RS, Barto AG (1981) Toward a modern theory of adaptive networks: expectation and prediction. Psychol Rev 88(2):135–170
Tindell AJ, Berridge KC, Aldridge JW (2004) Ventral pallidal representation of pavlovian cues and reward: population and rate codes. J Neurosci 24:1058–1069
Tindell AJ, Berridge KC, Zhang J, Peciña S, Aldridge JW (2005) Ventral pallidal neurons code incentive motivation: amplification by mesolimbic sensitization and amphetamine. Eur J Neurosci 22:2617–2634
Tindell AJ, Smith KS, Pecina S, Berridge KC, Aldridge JW (2006) Ventral pallidum firing codes hedonic reward: when a bad taste turns good. J Neurophysiol 96(5):2399–2409
Tindell AJ, Smith KS, Berridge KC, Aldridge JW (2009) Dynamic computation of incentive salience: "wanting" what was never "liked". J Neurosci 29(39):12220–12228
Toates F (1986) Motivational systems. Cambridge University Press, Cambridge
Zahm DS (2000) An integrative neuroanatomical perspective on some subcortical substrates of adaptive responding with emphasis on the nucleus accumbens. Neurosci Biobehav Rev 24:85–105
Zahm DS (2006) The evolving theory of basal forebrain functional-anatomical 'macrosystems'. Neurosci Biobehav Rev 30:148–172
Zhang J, Berridge KC, Tindell AJ, Smith KS, Aldridge JW (2009) Modeling the neural computation of incentive salience. PLoS Comput Biol 5:1–14