Note: I got a reviewer report on this piece pointing out significant overlap with Richard Holton & Kent Berridge's "Addiction between compulsion and choice" in N. Levy (ed.) *Addiction and Self-Control: Perspectives from Philosophy, Psychology, and Neuroscience*, Oxford, 2013, 239-265.

I may use some of the material in another piece in due course. But I won't revise this one. If you want something to cite, cite Holton & Berridge!

Liking versus Wanting

Abstract: This paper is an attempt to make sense of the empirical dissociation between *liking* and *wanting* (Berridge & Kringelbach, 2008, 2013; Pool et al., 2016). It will argue that the "wantings" in question are really *goals* or *intentions*, and that it is possible (indeed, common) for new intrinsic (non-instrumental) goals to be created in ways that bypass any anticipation of value from the result.

1. Introduction

Berridge & Kringelbach (2008, 2013) demonstrate that there are distinct subcortical networks in both humans and animals that underlie valence-production, on the one hand (which they call "liking"), and motivation, on the other (which they call "wanting"). These networks normally interact, but are potentially dissociable.[1] I shall begin by reviewing what is known about the nature and functional role of valence, before turning to its dissociation from "wanting".[2]

It seems increasingly likely that valence constitutes a single natural-psychological kind, the same in nature across all the different varieties of affective state. (These states constitute a broad class, including emotions like fear, anger, and disgust, but also moods like depression, felt desires like a longing to drink when thirsty or eat when hungry, enjoyment in an activity, and bodily states like pain and orgasm.) Valence-processing appears to be underlain by a single (albeit multicomponent) neurobiological network, involving not just subcortical evaluative regions in the basal ganglia (notably

[1] In addition, each is partly dissociable from the networks involved in reward learning. See Smith et al. (2011).

[2] I use quotation-marks throughout for Berridge & Kringelbach's technical notion of "wanting", since this only partly overlaps with the common-sense use of the term. The latter can equally well designate felt states of desiring or longing, which are better-characterized in terms of "liking". When one longs for something, positive valence is directed at the longed-for thing, which is thereby seen as good or desirable. The state of "wanting", in contrast, is a motivational state that can exist in the absence of any perceived desirability.

the ventral striatum), but also the anterior insula and anterior cingulate, together especially with orbitofrontal and ventromedial prefrontal cortex (Leknes & Tracey, 2008; FitzGerald et al., 2009; Plassmann et al., 2010; Bartra et al., 2013). The latter regions are the primary projection areas for valence signals in the cortex. These signals are thought to provide an evaluative "common currency" for use in affectively-based decision making (Levy & Glimcher, 2012; Ruff & Fehr, 2014). Positive and negative valence produced by many different properties of a thing or event can be summed and subtracted to produce an overall evaluative response, and such responses can be compared to enable us to choose among options that might otherwise appear incommensurable.

Valence also holds a central position in accounts of human decision-making that emphasize the role of *prospection* (Gilbert & Wilson, 2007; Benoit et al., 2014; Seligman et al., 2014). In prospection one envisages some of the options open to one, responding affectively to each. Those options seem to some degree good or to some degree bad. One likes or dislikes the thought of them, experiencing positive or negative valence directed at them, and takes a decision (thereby forming a new goal) accordingly. This is the mode of decision-making that seems to have become especially important in the human lineage, although it also appears to be present, to some degree, in other animals (Author, date 1). In prospection, valence signals are produced when the subcortical evaluative networks in question receive as input representations of future options or events. But it is these same networks that produce liking for a currently-experienced event.

Author (date 2) provides an extended argument for the view that positive and negative valence are best understood, not in terms of sensory pleasure or displeasure (thought of as intrinsic properties of experience), but rather as kinds of nonconceptual representation of value. A current activity or future option that is *liked* is one that is positively valenced, which means (I argue) that it is *seen* or *experienced* as to some degree *good*. But the goodness in question is represented in a fine-grained nonconceptual manner, without necessarily embedding the concept GOOD. In contrast, an activity or option that is *disliked* is one that is seen as to some degree bad.

Turning now to the dissociation between liking (positive valence) and "wanting": Berridge & Kringelbach (2008, 2013) show that there are distinct regions of the ventral striatum devoted to liking, disliking (negative valence), and to "wanting" (understood as the motivation or impulse to perform an action). Stimulation of the liking area will cause an animal to enjoy its food more intensely (as measured by a "lip-smacking" hedonic reaction, common to both rats and human infants), but without causing it to eat in greater quantities; whereas suppression of this area will reduce liking while having no impact on the amount eaten. (There are similar findings for the involvement of the disliking area in reactions to

bitter tastes.) In contrast, stimulation of the "wanting" region will cause an animal to eat more vigorously and to work harder for food, but without impacting its pleasure-reactions. It is thought that there are homologous networks in humans.[3]

Not only can liking and "wanting" be modulated independently of one another, but it is possible to "want" something in the absence of any positive valence directed toward it, or toward the thought of it. This can happen, for example, in the late stages of certain forms of addiction (for drugs, food, or gambling), as Berridge & Kringelbach (2008) point out. An addict can experience a powerful urge to take a drug (thereby "wanting" it), while seeing nothing good about taking it (neither anticipating that he will like it, nor actually liking it when consumed). How is this possible? How are we to understand the nature of this motivational state?

It is a familiar fact that anticipated and actual liking can come apart, of course. One can want something because one anticipates one will like it, but then when one gets it one is disappointed. This is no more problematic than any other form of mistaken prediction. What is much more puzzling is that people (and other animals) can be motivated to pursue things (and pursue them intrinsically, not for any further end) that they don't anticipate liking and don't perceive as in any respect good. (Likewise, they can like things that happen to them that they are neither motivated to continue nor to repeat; but this won't be our focus here.)

Although addiction will form our central example of "wanting" without anticipated-liking throughout, it might be helpful to have a more-familiar and mundane example on the table as well. Here is one. I often find it useful to break up intensive-writing days with a few games of solitaire. When I do so, I frequently reach a point where I think I have played enough, but nevertheless don't stop. I might say to myself, "OK, so this is the last round, then I'll get back to work." But if the game finishes without a win, I often find myself embarked on the next game, generally in full awareness of my previous decision. By this point, it is not that I am enjoying playing the game any longer. (Indeed, I may be bored with it.)

---

[3] Indeed, there is some older direct-stimulation work done with humans (Heath, 1972; Portenoy et al., 1986). When participants were provided with a button that provided brief electrical impulses to subcortical sites near the striatum, they self-stimulated almost continually, sometimes to the exclusion of all else. Although originally heralded as the discovery of the "pleasure center" in the brain, later reviews of the data provide a different interpretation (Berridge & Kringelbach, 2008). Participants displayed little sign of enjoyment, and didn't report feelings of pleasure. While they reported a variety of physical sensations, their most dominant experience was a brute urge to press the button once again, in the absence of any anticipation of pleasure. It seems that self-stimulation produced "wanting", not liking.

Nor do I anticipate any enjoyment from playing another round, nor from winning it. But neither does the thought of returning to work seem at all aversive; quite the contrary. It seems that, at some level and in some sense, I must want to win, and must want this more than I want to return to work. What kind of state is this? How is it to be understood?

I should stress that it is no part of my purpose here to explain what *differentiates* the two examples given above. Why is it that the "wanting" of an addict is felt so intensely, to the point where it can come to dominate life entirely, whereas the sort of "wanting" that leads me to play another round of solitaire is a minor nuisance? This is an important question. And a great deal of scientific work has been done on the underlying causes of addiction (Everitt & Robbins, 2016). But that is not my topic. My focus, rather, is on what the two cases have in common: the state of "wanting" without anticipated-liking. My goal is to understand the nature of this state.

Notice that if *all* decision-making is supposed to be prospection-based, then cases of "wanting" without anticipated-liking pose a problem. For the addict certainly seems to take *decisions* when pursuing drugs, or food, or the next gambling opportunity. Yet in some cases, at least, these goals are pursued in the absence of any anticipated value (except release from the urge to pursue them). And although the urges in question can feel irresistible, the actions taken are nevertheless fully intentional and controlled, it appears. After all, addicts can sometimes plot and execute elaborate strategies to achieve their goals. What we should conclude from this, I shall ultimately argue, is that there is a mode of goal-formation that humans share with other animals that is causally-direct and unconscious, and that doesn't involve valence-based prospection. But first we need to consider some other possibilities.

2. Motivation by habit?

One proposal for understanding "wanting" without anticipated-liking would be to assimilate it to the kind of motivation for action provided by *habits*. For habitual actions are nevertheless intentional ones, despite not being initiated or guided by prospection. (They certainly aren't mere reflexes.) Indeed, habits provide a direct sensorimotor route to action, whereby perception of habit-evoking cues leads directly to action, by-passing evaluative processing altogether. This is made possible through a network centered on a region of the dorsal striatum, which receives direct projections from perceptual cortices and has strong direct projections to premotor and motor cortex (Burton et al., 2015; Wood & Rünger, 2016). Moreover, people do often talk about someone's "drug habit", after all (as do many of the scientists who study addiction; see Everitt & Robbins, 2016).

This suggestion is implausible, however, at least as a complete account of the phenomenon of

"wanting" without anticipated-liking. This is because habits are independent of affective, evaluational, processing altogether, whereas states of "wanting", in contrast, seem capable of causing negative affect. (Consider how *upset* an addict can become when the object of addiction is unattainable.) Habits, although formed on the basis of repeated actions that may have an affective component, thereafter comprise a direct sensorimotor route to action. In fact, as noted above, perception of habit-evoking cues directly activates the motor routines that constitute the habitual action (Wood & Rünger, 2016).

Now, it is true that habits need to be suppressed through executive control-signals if they aren't to be carried out. And sometimes suppression can be difficult. But the difficulty is largely attentional, not affective. It can be difficult to "catch" a habitual action prior to its execution in time to prevent it. But resisting a habit when one is fully aware and prepared for the habit-evoking cues is generally not difficult, and rarely causes emotional distress. Why should it, since habits have their life independent of the affective system? It is also true that well-rehearsed habitual action-sequences can be hard to modify effectively. (Try repeating the alphabet without sounding—either aloud or in inner speech—the F and the G.) But in general they aren't hard to interrupt. (It is easy to *stop* oneself in the middle of saying the alphabet.)

A drug addict, in contrast, experiences a powerful urge to take the drug. He may work hard to get into a position to do so, growing increasingly anxious and frustrated if he fails, all the while wishing (whenever he reflects) that he could do something else, and while not anticipating any pleasure should he succeed. Moreover, it is the drug that is wanted, not (or not just) the familiar actions involved in drug-taking. For if the only available means of taking the drug involves a novel set of actions (snorting rather than injecting, say), then that will do just fine, and can fully satisfy the urge.

3. Emotional motivation?

A second suggestion would be to assimilate "wanting" without anticipated-liking to the sort of direct motivation for action provided by emotional states. A hiker who experiences fear when a bear looms out of the nearby bushes will likely have a strong urge to run away. This urge can be hard to resist, even when one knows full well that it is a black bear (not a grizzly), and that the best strategy is confrontation and noise, not retreat. If one can't resist the urge to run, one would be doing something that one doesn't anticipate ending well. Perhaps the addict is like that: someone who experiences a powerful, unreasoned, affectively-motivated urge to act, despite not envisaging any good from acting (except relief from the urge). If so, then again there is no threat to the prospection-based model of decision-making. For the impulse to run from the bear doesn't result from any decision. Decision-making is

required to *inhibit* the directly-caused impulse to run, but not to produce it.

Moreover, note that the kinds of action that can be directly motivated by emotional states aren't restricted to simple evolutionarily-ancient behaviors (running, laughing, and so on). When angry with a colleague at a meeting one might need to repress an urge to make an insulting remark, for example. Anger motivates actions intended to damage the object of one's anger. But what counts as "damage", and the sorts of actions that might secure it, can be quite varied. So it isn't ruled out that the addict's actions in preparing and consuming the drug might be directly motivated by an affective state in something like the way that emotion-expressive actions are.

Although more promising, this suggestion, too, seems inadequate. This is because what is motivated by addictive urges are actions (any actions) intended to achieve a specific end-state or goal-state (consuming the drug, consuming food, gambling). Achieving these goals may require elaborate and deliberate planning (all the while foreseeing nothing good about the outcome). This seems quite unlike cases of directly-caused emotion-expressing action, like an urge to run, an urge to laugh, or an urge to make a cutting remark. Indeed, the challenge of reconciling the dissociation between liking and "wanting" with the supposed ubiquity of valence-based prospection in decision-making can be located just here: for the addict who is planning a way to achieve the addictive goal is surely engaged in decision-making. Yet addicts can sometimes, apparently, take these decisions while envisaging no good from achieving the goal—neither anticipating, nor actually experiencing, positive valence in its achievement.

Some emotions can be chronic or persistent, however; and these come closer to our target. Consider someone who is motivated by revenge. The initial impulse is anger-like. There is an urge to damage those who have damaged us. In revenge this motivation can be long-standing, however (albeit intermittent). Whenever reminded of the object of one's revenge, feelings of anger might well up, and one may again resume plotting to find a way to exact a penalty. Those motivated by revenge seem to have an anger-caused long-standing goal: to cause damage, by whatever means, to the person in question. Indeed, in someone consumed by revenge, this goal can come to dominate one's attention and thoughts, somewhat like the goal of obtaining the object of one's addiction does for an addict. It seems that emotions can produce, not just direct action-tendencies, but also powerfully-motivating longer-term *goals*. This is much closer to the sort of "wanting"-state we are interested in (except that those motivated by revenge generally *do* see damage caused to the target as in some way good, or positively valenced). The suggestion that "wanting" without anticipated-liking is really a form of motivation by goals is what we explore next.

4. Motivation by goals?

A better suggestion—and the one I propose to develop and defend—is that the addict has an active goal (to take the drug) caused directly and unreflectively by a sub-component of the affective system. The affect the addict experiences then results from an appraisal that failure to act will *frustrate* that goal. The goal itself is not an affective state of any sort. Rather, it is more like an *intention* to consume, which directly motivates actions that one believes might achieve the intended outcome (the goal). Yet it is familiar to us all that things that frustrate our goals and intentions can give rise to negative affect. The negative affect that the addict experiences, I suggest, is a feeling that results from *frustration of the goal* of consuming the drug. Moreover, the goal in question, like all goals, can motivate reasoning and decision-making about the best way to achieve the goal. The *instrumental* decision-making that takes place in pursuit of the goal can involve valence-based prospection, but adoption of the goal itself isn't reflectively decided upon. It is, rather, directly activated by the reward-systems operating independently of any valence production.

What I will suggest, then, is that the dissociation between liking and "wanting" shows us something interesting about the nature of decision-making and goal adoption. What it shows, in fact, is that there are two distinct routes to the adoption of a new goal. One is valence-based prospection, in which one envisages the various alternatives open to one and responds affectively to each, forming one's goals accordingly. The other is a direct causal route to goal adoption, produced by the activity of (another component of) the affective system. In fact, both reflective valence-based (liking-based) prospection and intuitive causally-direct goal adoption ("wanting") converge on the acquisition or activation of goals. The latter then directly motivate non-habitual forms of action, leading one to engage in reflective kinds of planning where necessary, guided by one's beliefs and the affordances of the environment.

In order to make the case for this account, I first need to say something about what goals are, and how they are related to intentions (Section 5). I will then discuss some nonhuman-animal literature that establishes the existence of a form of unreflective goal-formation that nevertheless results in fully-controlled intentional action (Section 6), before reviewing some of what is now a robust literature on unconscious (unreflective) goal-activation and goal-pursuit in humans (Section 7). What these literatures will show is that even if goal adoption is always dependent on previous reward-based learning, goals can thereafter be created in ways that bypass valence-production.

5. Goals and intentions

In the scientific literature (as well as in ordinary talk) the terms "goal" and "intention" are often used interchangeably. One might equally well describe a participant in a psychology experiment as forming the *goal* of complying with the experimental instructions, or as forming the *intention* to comply, for example. But it can sound more natural to talk about intentions where the intended thing is an action or activity of some sort. (It sounds odd, but not obviously false, to say that one has the goal of going hiking on Saturday; one would normally say that one *intends* to hike on Saturday.) In contrast, it is more natural to talk of goals where the end-state is a state of affairs. (One has the goal of being a home-owner; it can sound odd to say that one intends to be one, unless one has in mind some specific action that might achieve it.) Likewise, goals and intentions seem to exist on a spectrum of abstractness. When one reaches toward a glass of water for a drink it would be more natural to say that one *intends* to pick it up to drink, than that one has the goal of doing so.[4]

Whatever the differences between goals and intentions (and I think they are slight and perhaps merely verbal), neither kind of state is affective in nature. Indeed, both are thought to be realized outside of the main affective networks in the brain. In the neuroscientific literature, intentions are generally thought to be realized in premotor and supplementary motor cortex, anterior to primary motor cortex, whereas current goals are often described as being realized somewhat more anterior still, in the control-structures of dorsolateral prefrontal cortex (Morsella et al., 2009). (Note that this is consistent the idea of a gradation of increasing abstractness between intentions and goals.) Moreover, both can guide and constrain further planning and decision-making without needing any involvement from affective states, as described in some detail by Bratman (1987).

Consider a simple intention for the here-and-now, for example. Suppose the room is uncomfortably hot, and one has just decided to open the window for cooler air (thereby forming an intention to do so). This will initiate an increasingly fine-grained cascade of motor plans, guided by one's initial situation (sitting, standing, or whatever) as well as the affordances of the environment (the shape, size, and orientation of the window-handles, for example). This may involve unconscious calculations of relative *ease* of action, which at some level might implicate valence signals. But it certainly needn't

[4] Notice that one could also say quite naturally that one *wants* to go hiking or to become a home owner, or has a *desire* to drink water from the glass. The ordinary notion of desire includes both affective valence-involving states as well as more-cognitive states of intending and taking-as-a-goal. The common-sense notion therefore fails to pick out a single natural-psychological kind.

involve valence-involving conscious prospection. If execution of the intention is blocked, however (for example, the window turns out to be stuck, and won't budge), then this will be felt as *frustrating*. One will experience negative affect (annoyance, or disappointment) directed at the situation. It seems that intentions, although not intrinsically affective in nature, have the power to activate affective networks when frustrated.

Intentions and goals are often more distantly future-oriented, of course. One can form the intention to go running tomorrow, or have the goal of vacationing in Martinique next summer. When something reminds one of such intentions, they motivate consideration of the means of implementing them. Yet the trouble with such open-ended goals is that they make procrastination too easy. If something tomorrow morning reminds one of one's intention to go running, one can defer implementation until the afternoon without giving up on the intention; one can do the same thing in the afternoon, until the next time one is reminded, time has run out. This is why it is advisable to formulate what are called in the psychological literature "implementation intentions", such as: "when I return from walking the dog around the block in the morning, I shall go for a run" (Gollwitzer & Sheeran, 2006). This is because, when triggered into activity by fulfillment of the antecedent clause, they either have to be acted on or changed. People are much more likely to do what they intended to in these circumstances.

Now notice that when a goal remains unfulfilled, it will be apt to direct attention and planning while active, and then again whenever it gets triggered into activity thereafter. Suppose that, unable to open the window, one takes off one's jacket instead and returns to work. But if one happens to look up and notice the closed window later, or if one finds oneself hot once again, one's intention to get the window open will be apt to become active once again. One may become distracted from one's work, and get up to make a second attempt; or one might break off work to call the maintenance department. Even for a trivial matter like this, an intention will tend to motivate action towards its own fulfillment whenever it becomes active.

The relevance of this to "wanting" without anticipated-liking in addiction should be obvious, on the assumption that the "want" here is actually a goal or intention to take the drug. Supposing such a goal can be created by the sight of drug paraphernalia or by physical cues from one's own body, it will thereafter be apt to distract from one's current activity, dominating attention and planning in the service of its own fulfillment. If the state in question is goal-like, then this can happen in the absence of any anticipated liking for the result. At this point it remains unclear, however, how it is possible for goals to be created in the absence of prospective decision-making. For after all, prospective reasoning may already have caused the addict to *abandon* that goal (just as it causes me to decide that I have been

playing solitaire for long enough). So the addict isn't *reminded* of a standing intention to take drugs. He may no longer have any such intention. Rather, such a goal must somehow be (re)created afresh. This is where we go next.

6. Incentive salience

How are goals formed? In particular, how are intrinsic, or ultimate, goals formed? (Of course, sub-goals or instrumental goals can be formed purely cognitively by reasoning, although instrumental reasoning can also implicate valence-based prospection.) One way in which such goals are formed is through prospection. One envisages the options and end-states open to one, responding affectively to each as one moves one's attention back and forth between them; issuing eventually in a decision that one of them is best. The result is a new intrinsic goal. But there is another mode of goal formation that is more basic and ancient, which has been much studied in the animal literature. It happens when an animal perceives a cue that signals a way of achieving something previously experienced as rewarding. This creates the goal of getting that thing. This phenomenon is referred to in the literature as "incentive salience." Moreover, that this is the right place to look for an understanding of the states of "wanting" that interest us is suggested by the fact that Berridge & Kringelbach (2013) themselves describe "wanting" as a form of incentive salience.[5]

Careful experimental work done more than 30 years ago established that animal behavior comprises both habits and goal-directed actions (Dickinson, 1985). Depending on the sort of training schedule employed, or on how long training lasts, an animal might develop a habit that is independent of its own motivational state; and independent, too, of its beliefs about the contingencies of the situation. Once a rat has developed the habit of pulling on a chain (to get food), it will continue to pull even when food is no longer available, and even when pre-fed to satiety so that it doesn't actually eat any of the food when it *is* available. In contrast, given other forms of training, the rat's actions can be genuinely goal-directed. Believing that pulling the chain will obtain the food, and having the goal of obtaining food, the rat will pull—but only if it is hungry, and not if it is shown that pulling the chain is no longer effective. Importantly for our purposes, is seems that renewed sight of the chain is sufficient to create the goal of obtaining food, even in an animal that had until that moment been engaged with other things or with other forms of search.

[5] See also Smith et al. (2011). And see Robinson & Berridge (2008) for an account of addiction as incentive-salience sensitization.

Can we be confident, however, that goal-formation, here, isn't prospection-based? How do we know that the rat doesn't decide to pursue the food (by pulling the chain) by first imagining the outcome and responding with positive valence to the thought of the result? In fact, we know that the rat's decision-making is unlikely to be prospection-based, because of an interesting and well-replicated finding (Dickinson & Balleine, 1994; Balleine, 2005). This is that the animal has to *learn*, through actual consumption, the reward-value of the goal when in an altered physiological state. For example, an animal that has learned to pull a chain to get food but has now been given an emetic that will provoke disgust on contact with the food will still work hard initially at chain-pulling (in conditions of extinction, when no reward is actually available). It is only when the animal is first allowed renewed contact with the food in its altered physiological state that it will no longer pull the chain.

In contrast, we all know from our own experience that the mere thought of eating something that has recently made one ill (or is appraised by the disgust system as having made one ill) is enough to provoke disgust. We don't have to re-experience eating that food in order not to want it any longer. This is the way the prospection system works, after all: one imagines the outcome or action in question, and one's affective system responds with positive or negative valence much as if the outcome or action were currently being experienced.

What these findings show is that there is a way for goals to be created that is independent of valence-based prospection. Perceiving that a situation S affords an action A, and believing that A will bring about result R, where R is something that has previously been experienced as consistently rewarding, will cause one to adopt of the goal of achieving R. The stored value of R interacts with one's beliefs about the available means to create a new active goal: get R. This will then guide attention, subsequent planning, and subsequent action; normally until R is achieved. For this to work, one doesn't need to *imagine* achieving R or performing A, nor need one enter into a positively valenced state directed at R. The causal process in question is unconscious and direct.

For this account to be viable, it needs to be the case that the value attaching to the end-state R is somehow stored in the reward system in a way that doesn't contribute to valence-production. Otherwise, when one imagines that end-state in prospection, the stored value should create positive valence directed at it—that is to say, one should anticipate liking it. But we know that in cases of "wanting" without anticipated-liking, it doesn't. As I will now briefly explain, however, this condition is met. There are distinct and dissociable subcortical networks that contribute to "wanting" (goal-formation) and to liking respectively.

We noted in Section 5 that goals and intentions are thought to be realized in regions of

dorsolateral prefrontal cortex, premotor cortex, and the supplementary motor area—all regions on the upper outside surface of frontal cortex. Valence signals, in contrast (pleasure and displeasure, and representations of subjective value more generally) are primarily realized in orbitofrontal cortex and ventromedial prefrontal cortex—regions located on the underneath and lower medial surface of prefrontal cortex (Berridge & Kringelbach, 2008, 2013; Leknes & Tracey, 2008; Bartra et al., 2013). Both sets of regions interact richly with subcortical value-processing structures, especially in the striatum. Importantly for our purposes, the striatum itself has significant internal organization. Ventromedial regions (especially the Nucleus Accumbens, or NAcc) are especially implicated in hedonic and valuational processes, and project strongly to orbitofrontal and ventromedial prefrontal cortex. In contrast, the dorsolateral region of the striatum forms part of a sensorimotor network that underlies habitual responding, receiving input from sensory cortices, and with strong projections to premotor and motor cortex (Burton et al., 2015; Wood & Rünger, 2016).

Where does previously-rewarded goal-formation fit into this picture? It turns out that when animals are trained in rewarding tasks until they acquire a habit (which is then no longer sensitive to the value of the outcome), control of behavior can be seen to shift gradually from the NAcc up to dorsolateral striatum, with the intermediate region of dorsomedial striatum seemingly being especially implicated in incentive-salience, or goal-directed behavior (Burton et al., 2015). This is fully consistent with the idea being put forward here, that "wanting" without anticipated-liking should be thought of as goal-directed motivation in the absence of anticipated value—a phenomenon that can be seen to be half-way between valence-based decision-making and value-independent habitual responding. (It has the behavioral flexibility of the former while sharing the insensitivity to anticipated reward of the latter.)

Obviously this only scratches the surface of what is now an immensely rich body of neuroscientific work. The important point for our purposes is just that the neuroscience seems to bear out the reality of three distinct modes of selection and control of behavior. One is habitual responding, which is independent of the affective system (except historically, in habit *formation*), and is dependent on stimulus cues. The second is valence-based prospection of alternatives leading to the formation of a novel goal or intention. And the third is *direct* activation of a goal from stimulus cues (incentive salience), on the basis of previous affective learning.

I suggest, then, that cases of "wanting" without anticipated-liking can be understood on the model of incentive salience outlined here. When the addict sees or is reminded of drug-taking, perhaps by sight of drug paraphernalia of some sort, this creates the goal of consuming the drug. This happens as a result of previous reward-based learning and independently of any capacity that the drug might now

have to give pleasure (whether real or in prospection), much as the sight of the chain creates the goal of obtaining food in an animal that wouldn't actually enjoy it because of its altered physiological state. The goal then motivates actions that are believed to achieve it, producing an urge to perform them and feelings of frustration when they aren't successful.

7. Unconscious goal-activation

The account outlined here is consistent with, and receives additional indirect support from, a robust literature on unconscious goal-activation and unconscious goal-pursuit (Dijksterhuis & Aarts, 2010; Huang & Bargh, 2014). In particular, it seems that goals can be created by *priming*. This can either be subliminal, or by solving word-puzzles in which associated concepts are embedded. Once active, these goals are apt to consume executive resources, presumably by initiating searches of memory and attentional search of the environment, on the lookout for a means to achieve the goal. For example, participants in a working-memory task can be primed with the goal of socializing, or with a goal known from a previous questionnaire to be individually important to them. Having been so primed, they are significantly slower at the experimental task they have undertaken than are control participants (Marien et al., 2012). The newly-active goal, acting alongside the existing one, divides their attention to some degree.

Consider another example. Lau & Passingham (2007) were able to demonstrate that subliminal priming can activate goals that would otherwise not be present. Participants engaged in a task where they either had to make a semantic judgment (indicating whether a word was concrete or abstract) or a phonological one (indicating whether the word was bi-syllabic). The judgment that they were to make on each trial was signaled by a prior visible cue (a square or a diamond respectively). Shortly before presentation of the cue, participants were presented with a smaller square or diamond shape, in conditions where it was rendered either visible or invisible through the timing of a backward mask. They were told to ignore these initial shapes, which could be either congruent or incongruent with the target cues.

What the experiment found was that when the prime was both subliminal and incongruent with the target cue, people made more errors and were slower to respond, suggesting that the prime had activated the opposite goal, which then conflicted with the cued goal. (There was no effect when the prime was visible.) Moreover, brain imaging revealed increased activity in cortical areas consistent with the subliminal prime (associated with semantic and phonological judgments respectively), as well as increased activity in dorsolateral prefrontal cortex in cases where the subliminal prime and target cue

were in conflict. (Recall that this is the brain region in which goals are thought to be realized). These findings strongly suggest that the goal of making a semantic judgment, for example, was unconsciously activated by the subliminal square and that it interfered with the operation of the consciously-caused goal of making a phonological judgment.

While such findings are obviously quite remote from the phenomenon of addiction, they do seem to demonstrate that goals can be created in ways that bypass valence-based prospection. Indeed, participants in the experiment just described can be said to be in a state of "wanting" without anticipated-liking. For they have adopted a new goal (to make a semantic judgment, say) without experiencing any anticipatory positive valence from achieving that goal.

8. Conclusion

This paper has been an attempt to understand the phenomenon of "wanting" something (and wanting it intrinsically, not instrumentally) without liking it, and without even anticipating liking it. This dissociation is displayed most dramatically in certain forms of addiction, although it can also manifest itself in ordinary life. (Recall the example of compulsive solitaire-playing.) What I have suggested is that the sort of wanting in question is really a non-affective state of *intending to achieve something* or *having something as a goal*. And we should recognize that there are two distinct routes to goal formation. One involves the liking-system, as when one engages in prospective valence-imbued reasoning over a set of alternatives. The other is a causally-direct route to goal formation, triggered by cues that signal the presence of a way of achieving something whose value has previously been stored through reward-based learning.

I have not attempted to provide a complete theory of addiction, of course. I have said nothing about why the addict's goal seems so *pressing* in comparison to the other examples we have discussed. Why does that goal dominate all others? Nor have I attempted to explain what causes the "wanting"-network and the liking-network to dissociate so strongly in addiction, although they normally operate in close concert with one another. My aim has only been to make sense of what sort of state "wanting" is in the absence of anticipated liking. I have suggested that it is best thought of as a goal-state that has been caused independently of valence-involving prospection.

Acknowledgments

I am grateful to RS for comments on a previous draft of this paper.

References

Balleine, B. (2005). Neural bases of food-seeking: Affect, arousal and reward in corticostriatolimbic circuits. *Physiology & Behavior*, 86, 717-730.

Bartra, O., McGuire, J., & Kable, J. (2013). The valuation system: A coordinate-based meta-analysis of BOLD fMRI experiments examining neural correlates of subjective value. *NeuroImage*, 76, 412-427.

Benoit, R., Szpunar, K., & Schacter, D. (2014). Ventromedial prefrontal cortex supports affective future simulation by integrating distributed knowledge. *Proceedings of the National Academy of Sciences*, 111, 16550-16555.

Berridge, K. & Kringelbach, M. (2008). Affective neuroscience of pleasure: Reward in humans and animals. *Psychopharmacology*, 199, 457-480.

Berridge, K. & Kringelbach, M. (2013). Neuroscience of affect: Brain mechanisms of pleasure and displeasure. *Current Opinion in Neurobiology*, 23, 294-303.

Bratman, M. (1987). *Intentions, Plans, and Practical Reason*. Cambridge MA: Harvard University Press.

Burton, A., Nakamura, K., & Roesch, M. (2015). From ventral-medial to dorsal-lateral striatum: Neural correlates of reward-guided decision-making. *Neurobiology of Learning and Memory*, 117, 51-59.

Dickinson, A. (1985). Actions and habits: The development of behavioral autonomy. *Philosophical Transactions of the Royal Society of London, Series B, Biological Sciences*, 308, 67-78.

Dickinson, A. & Balleine, B. (1994). Motivational control of goal-directed action. *Animal Learning & Behavior*, 22, 1-18.

Dijksterhuis, A. & Aarts, H. (2010). Goals, attention, and (un)consciousness. *Annual Review of Psychology*, 61, 467-490.

Everitt, B. & Robbins, T. (2016). Drug addiction: Updating actions to habits to compulsions ten years on. *Annual Review of Psychology*, 67, 23-50.

FitzGerald, T., Seymour, B., & Dolan, R. (2009). The role of human orbitofrontal cortex in value comparison for incommensurable objects. *Journal of Neuroscience*, 29, 8388-8395.

Gilbert, D. & Wilson, T. (2007). Prospection: Experiencing the future. *Science*, 317, 1351-1354.

Gollwitzer, P. & Sheeran, P. (2006). Implementation intentions and goal achievement: A meta-analysis of effects and processes. *Advances in Experimental Social Psychology*, 38, 69-119.

Heath, R. (1972). Pleasure and brain activity in man. *Journal of Nervous Mental Disorders*, 154, 3-18

Huang, J. & Bargh, J. (2014). The selfish goal: Autonomously operating motivational structures as the

proximate cause of human judgment and behavior. *Behavioral and Brain Sciences*, 37, 121-135.

Lau, H. & Passingham, R. (2007). Unconscious activation of the cognitive control system in the human prefrontal cortex. *Journal of Neuroscience*, 27, 5805-5811.

Leknes, S. & Tracey, I. (2008). A common neurobiology for pain and pleasure. *Nature Reviews Neuroscience*, 9, 314-320.

Levy, D. & Glimcher, P. (2012). The root of all value: A neural common currency for choice. *Current Opinion in Neurobiology*, 22, 1027-1038.

Marien, H., Custers, R., Hassin, R., & Aarts, H. (2012). Unconscious goal activation and the hijacking of the executive function. *Journal of Personality and Social Psychology*, 103, 399-415.

Morsella, E., Bargh, J., & Gollwitzer, P. (eds.) (2009). *Oxford Handbook of Human Action*. NY: Oxford University Press.

Plassmann, H., O'Doherty, J., & Rangel, A. (2010). Appetitive and aversive goal values are encoded in medial orbitofrontal cortex at the time of decision making. *Journal of Neuroscience*, 30, 10799-10808.

Pool, E., Sennwald, V., Delplanque, S., Brosch, T., & Sander, D. (2016). Measuring wanting and liking from animals to humans: A systematic review. *Neuroscience and Biobehavioral Reviews*, 63, 124-142.

Portenoy, R., Jarden, J., Sidtis, J., Lipton, R., Foley, K., & Rottenberg, D. (1986). Compulsive thalamic self-stimulation: A case with metabolic, electrophysiologic and behavioral correlates. *Pain*, 27, 277-290.

Robinson, T. & Berridge, K. (2008). The incentive sensitization theory of addiction: Some current issues. *Philosophical Transactions of the Royal Society of London: B*, 363, 3137-3146.

Ruff, C. & Fehr, E. (2014). The neurobiology of rewards and values in social decision-making. *Nature Reviews Neuroscience*, 15, 549-562.

Seligman, M., Railton, P., Baumeister, R., & Sripada, C. (2013). Navigating into the future or driven by the past. *Perspectives on Psychological Science*, 8, 119-141.

Smith, K.S., Berridge, K., & Aldridge, J. (2011). Disentangling pleasure from incentive salience and learning signals in brain reward circuitry. *Proceedings of the National Academy of Sciences*, 108, E255-E264.

Wood, W. & Rünger, D. (2015). Psychology of habit. *Annual Review of Psychology*, 67, 289-314.