On The Nature and Rationality of Desire

Phil7018: Research Report

Submitted in partial fulfilment of the requirements for the degree of Master of Arts (by Coursework and Research Report).

Supervisor: Prof. Mark Leon

Khondlo Mtshali

0509489h

University of Witwatersrand

27 July 2018

ACKNOWLEDGEMENTS

It has taken me quite a long time to complete this project, throughout the journey I have had diligent support and help from friends, family, my supervisor and the philosophy department. I would like to thank Prof Mark Leon for his assistance, without Prof Leon's encouragement and admonishments when necessary the project would have not been completed. I am grateful for his support and the space he gave me to do what I wanted.

I would also like to thank my family. The "when are you submitting question" as frustrating as it was provided me with the fuel I needed to finish this project. I would like to thank my mother for her constant and unwavering support. When most people questioned my doing philosophy her advice was "do what you love and do it well". To my siblings "Y'all are gangsta" thank you for the support. I want to single out my sister, Zamantshali, dude thank you for reading massive amounts of "weird stuff" and giving me invaluable advice.

To my friends thank you for the support. I cannot mention all of you by name but thank you to each and every one of you for the input and support.

Finally, the Philosophy Department, thank you for the unwavering support. The Wits Philosophy Department is, in my opinion, the best philosophy department.

Ngiyabonga! Thank you!

Declaration

I Khondlo Quett Sithembiso Mtshali (Student number: 0509489h) declare that this thesis, which I submit to University of Witwatersrand for examination in consideration of the award of a higher degree, Master of Arts in Philosophy by Coursework and Research Report, is my own personal effort. Where any of the content presented is the result of input or data from a related collaborative research programme this is duly acknowledged in the text such that it is possible to ascertain how much of the work is my own. I have not already obtained a degree at University of Witwatersrand or elsewhere on the basis of this work. Furthermore, I took reasonable care to ensure that the work is original, and, to the best of my knowledge, does not breach copyright law, and has not been taken from other sources except where such work has been cited and acknowledged within the text.

Signed _____

Student Number _____**0509489H**_____

Date _____**28/07/2018**_____

Introduction

What is desire? What do we mean when we say that an organism has a desire? Can desires be rational? In this research report, I attempt to answer these questions. I take a different approach to that often taken by philosophers where desire is concerned. I do not set out defend any of the traditional approaches towards desire; instead I will be favouring a novel approach, an approach inspired by work done in neuroscience. In the first chapter, I discuss desire three traditional methods of defining desire: as motivation, as pleasure, and as goodness. My aim is to illustrate that these *a priori* definitions of desire are inadequate, because whatever position we take results in exclusions (which, upon analysis, are viewed as unjustifiable). If, for example, we define desire as motivation (and by this, we mean that organisms only desire things which they can bring about as a result of acting), we *exclude* things that action is ill-suited to bring about from the set of things that can be desired (and therefore thought of as being under the purview of desire). I and many other philosophers feel that this is a very restrictive definition. Some philosophers mean to solve this issue by replacing 'motivation' with some other equally restrictive criterion; I, however, propose an *a posteriori* definition of desire which I give a detailed account of in the second chapter. In the second chapter, I propose a reward-based account of desire. A reward-based account states that an organism is viewed as desiring something if the thing that is desired results in reward, the release of dopamine, and subsequently results in the repetition of the rewarding action. I think that this theory of desire is preferable because it accounts for all the phenomena that we associate with desire. I illustrate the commensurability of this theory with all the phenomena that we associate with desire by illustrating that there are neuro-anatomical connections between the reward system and the motivational and pleasure centres of the brain. Accepting the reward-based theory of desire means that we no longer have to give an *a priori* definition of desire — desire is whatever activates the reward system, dopamine — and results in repeated action.

Having put forward this theory, I then go on, in the third chapter, to argue for the existence of two types of desires: simple and complex desires. I define simple desires as desires that are created by departures from homeostasis, i.e., responses to physiological need, while complex desires are desires that are informed by the beliefs that an organism has. My argument for simple and complex desires is twofold. I first argue for the acceptance of desire as action setting states: they are like beliefs, in that they set goals without dictating specific actions that ought to be taken to achieve those goals. I then argue that physiological responses, i.e., simple desires, ought to be considered as desires given that they also set goals. I then move on to the final chapter, chapter four, where I look at whether or not desires ought to be viewed as rational. In this chapter, I argue that rationality ought to be understood as reason-responsiveness and given this, the rationality of simple and complex desires ought to be determined in virtue of whether or not they are responsive to their proper determinants. By reason-responsive, I mean that desires are not formed arbitrarily, but rather in response to the reward experienced by the organism when it performs certain actions under certain circumstances. I then conclude that if the dopamine-releasing system is responsible for the creation of desire, and its functionality is reason-responsive, and reason-responsiveness implies rationality, desires must also be rational

in both how they are formed and maintained. I conclude that both simple and complex desires are reason-responsive and thus both are rational.

In this thesis, I propose the re-imagining of desire. By re-imagination, I propose the naturalisation of desire: the empirical study of desire and the other parts of the brain that it is connected to. I prose that desire, which I define as *reward*, is produced by the function of the dopamine system. In suggesting that desire be studied using neuroscientific methods, I am suggesting that what is desirable (and therefore, rewarding) can only be determined *a posteriori*, on the basis of how it activates the dopaminergic system, and finally that this process is rational.

Chapter one

An exploration and critique of the most common theories of desire

Action or motivation-based theories of desire assume that something is desirable in so far as it motivates action. However, merely specifying that an organism has desires based on the fact that it engages in action can result in confusion. Not all actions are associated with desire. Some actions are indicators of aversion. Given this, a more precise definition of action based theories of desire is needed. To better define these theories we could say an organism has a desire(s) when it acts in a way that is intended to bring about a certain state of affairs. Alternatively, an organism has a desire if it acts in a way that is intended to result in the acquisition of a certain object(s). Motivational accounts of desires are intuitively plausible: something about desire seems intrinsically motivational. These accounts, however, have a major defect. Saying that an organism has a desire if it is motivated to act in a particular way does not explain what about the motivating object or state of affairs is motivational. Desire-as-motivation does not, for example, refer to conceptions of goodness or pleasure as the source of motivation. In as much as we assume desire to be motivational, we assume organisms are motivated by something and that something results in action. We assume action, motivation, to be a means to an end and not the end itself. If I walk to the shop it is because there is something at the shop which I desire. Walking is a means to the object of my desire but not the desire itself *ceteris paribus*.

If we do not differentiate between what is motivational and the culmination of motivation in action we find ourselves in a situation where every action could be an indicator of desire. There are actions that we would want to classify as unplanned and therefore not endorsed by desire. Consider reflex actions. Reflex actions are involuntary movements prompted by an instantaneous reaction to a stimulus. If we do not distinguish between motivation as the initiation of action and motivation as made manifest in action, all actions including reflex actions will be viewed as indicative of desire, in virtue of their being actions. In as much as we do not want reflex actions to be viewed as actions indicative of desire, we also do not want the unintended consequences of our actions to be viewed as signalling our desires. To use Donald Davidson's illustration, my flipping a switch with the intention of illuminating a room may also alert a criminal about my presence, but only one of those outcomes was desired by me "A reason rationalizes an action only if it leads us to see something the agent saw, or thought he saw, in his action-some feature, consequence, or aspect of the action the agent wanted, desired, prized, held dear, thought dutiful, beneficial, obligatory, or agreeable." (1960:685). Distinctions of this sort can only be drawn if we distinguish between performing an action and performing an action intentionally, i.e., performing an action for a reason. Desire ought to be associated with the intentional performance of an action, not merely action.

The abovementioned issues do not exhaust the problems faced by motivational accounts of desire. In addition, these accounts do not allow for cases where an organism has a desire but that desire does not ultimately issue in action. Consider a person who wants to go to a concert but lacks the funds to do so. Can this person be thought of as having a desire if they do not act on what they desire? Michael Smith (1987 and others) answer 'Yes' to this question. Some philosophers claim that motivational theories of desire ought to be understood as *dispositions*

to action as opposed to the successful execution of actions. In the concert example, the individual has a desire because he has a disposition – an inclination – to act in a particular way. The importance of associating desire with dispositions to act (as opposed to successfully executed actions) is predicated on the fact that desire need not (and often does not) issue in action, yet the absence of action does not imply the absence of desire. My inability to go to a concert does not mean that I do not desire it; it only means that I cannot affect that desire (perhaps because I did not have money to buy the tickets, or the tickets were sold out).

Some philosophers, such as Daniel Friedrich (2014) and Stephen Darwall (2003), believe that to desire, one requires *more* than a disposition to act. A desire requires a disposition to action plus a belief about the action one is disposed to, namely that the action taken will result in the satisfaction of the desire: "It is the function of belief to link up the desired state of affairs with actions in the agent's repertoire." (Friedrich, 2014: 257).This theory emphasizes the idea that desires are states that an organism is aware of and which factor in action selection

> To sum up, to the extent that desires are based on reasons, the fact of desire is not an additional reason. It is something like the agent's correctly registering reasons, where these are facts about the desire's object. These reasons warrant the desire (and, if conclusive, the action), but, so far as the desire is concerned, they are all the reasons there are. (Darwall, 2003: 438).

However, accounts of this sort are also problematic. This type of account specifies that an action is chosen on the basis of a belief, namely a belief that if the action is selected it will lead to the satisfaction of a desire. This theory is predicated on the assumption that there is a reliable link between beliefs and dispositions, namely that beliefs are infallible and that they pick out the right dispositions.

Beliefs do not (or rather, the mechanism responsible for creating beliefs does not) always function correctly. Beliefs can malfunction. Because our beliefs can malfunction, it is important to cash out a theory of desire that places a premium on motivation but also accounts for possible malfunctions. An account of desire that factors in malfunction conceives of desires as states (mental/psychological) whose function is to produce or dispose one toward action in virtue of a belief about the action's likelihood to result in satisfaction. Theories that account for malfunctions are advantageous because they specify that to be a desire is to retain the function of disposing one to action irrespective of whether the disposition to action is in fact produced or whether the action toward which one is disposed is appropriate given some belief(s).

Some philosophers, including Timothy Schroeder and Galen Strawson, believe that motivational theories of desire are false, or, if not false, they are overly restrictive. People who hold this view believe that there are things that people desire for which action is ill-suited to bring about. Consider a person who desires the rain to stop so that he can go on a picnic. There is no action available to the individual in question which when undertaken would result in the realisation of his desire. The individual in question cannot bring about what they desire, but they do have an emotive relationship to their desire (namely, pleasure at its realisation). Given this, some philosophers have gone on to propose pleasure-based theories of desire. For an

organism to desire under the pleasure thesis, the organism must be disposed toward enjoyment/pleasure, at it seeming that ***P***, and disposed toward suffering/displeasure at it seeming that ***–P***. Cashing out desire in terms of pleasure and displeasure ameliorates the problem presented by desires that action is ill-suited to bring about. Take the person who desires the rain to stop. Such a person has no action they can enact to ensure that the rain stops; they would, however, feel pleasure as a consequence of the rain stopping and therefore have a desire in that respect.

If desires dispose us to action it would be incorrect to speak of the person wanting the rain to stop as having a desire. If we take this position, some of the things that we speak of in terms of desire would have to be expressed differently. In the case of the person desiring the rain to stop we could make the claim that he *wishes* for the rain to stop, as opposed to desiring it. We could make the claim that people who speak of having desires for things that action is ill-suited to bring about equivocate between *desiring* and *wishing*. While this claim can be made, it seems overly restrictive and unnecessary. We could simply make the claim that when an organism desires something, they are disposed toward feeling pleasure at its realisation or feeling displeasure at its non-realisation.

Speaking about desire in terms of pleasure and displeasure allows us to maintain desire talk in situations where there are no mechanisms for realising action. It may seem like needless repetition to speak about the lack of mechanisms intended to realise a desire (i.e. to produce an action) having spoken about situations where no action is available to realise a desire. However, the distinction between these cases is subtle and requires nuanced treatment. When a person desires that the rain stop, the person lacks access to a set of actions that, if undertaken, would result in the rain stopping. The person is, however, able to act *ceteris paribus*; the person's limitation is only with respect to stopping the rain. A theory of desire that places a premium on pleasure states that an organism can desire without being able to bring about what is desired, so long as the organism has proper dispositions. A person who desires the rain to stop would be pleased if it stopped raining and displeased if it continued, and therefore possesses the requisite dispositions to count as desiring. What about a quadriplegic? Can a quadriplegic desire? Quadriplegia is interesting because proponents of the desire-as-pleasure thesis can claim quadriplegia as a counterexample to desire-as-motivation. Proponents of desire-as-pleasure can make this claim because nothing about quadriplegia seems to contradict desire. Given this, quadriplegia seems to contradict the idea that desire ought to be cashed out in motivational terms. Quadriplegia seems to be affirming desire understood as pleasure as opposed to desire understood as action. Consequently, proponents of desire-as-pleasure would argue that their theory of desire is right, given its broader applicability, and its intuitive value. However, proponents of motivational accounts of desire would deny that quadriplegia offers a counterexample to their theory.

Quadriplegia does not render motivational accounts of desire invalid. To make the claim that it does is to misunderstand motivational theories of desire. Motivational theories of desire state that an organism cannot have a desire for something that action is ill-suited to bring about. If something cannot be brought about by an action, it is not a desire. If, on the other hand, an organism has a desire for something that action is best-suited to bring about, but the organism

lacks the capacity to initiate and execute actions that would bring about the thing that is desired, the organism can still be thought of as having desire. Proponents of motivational theories of desire would say that quadriplegic individuals who have desires for things that action is best-suited to bring about *do* have desires: desire understood in motivational terms. This is due to the fact that a quadriplegia is a form of malfunction. The failure to dispose one toward action – malfunction – does not imply that the function, disposing action, ceases to exist, or more importantly, that bringing it about is impossible. If a quadriplegic desires something that an able-bodied person can bring about, the quadriplegic has desire because had it not been for the quadriplegia, i.e., malfunction, he could have acted to bring about his desire. Malfunction in this instance prevents action, but action is still possible (possible in principle).

The quadriplegic and picnic cases are vastly different. The picnic case presents a desire that is ill-suited for action and therefore it is not a desire according to the motivational perspective. No action can be taken to prevent it from raining or stopping the rain once it starts. The classification of "rain stoppage" as a desire does not turn on the properties of the organism that is desiring, but rather on the ontological status of the thing being desired. It is important to note that the property of being ill-suited for action does not refer to the impossibility of something occurring (the rain can and does stop), but rather refers to the impossibility of something being brought about intentionally / being caused intentionally via the actions of an organism. This point can be cashed out in Frankfurtian terms. Frankfurt makes a distinction between effective desires and weak desires / non-effective desires. An effective desire is a desire that moves an individual to act: "Rather, it is the notion of an *effective* desire – one that moves (or will and would move) a person all the way to action." (Frankfurt, 1971:5). Some quadriplegic desires could be effective and therefore are desires (understood in motivational terms), while all rain-stoppage desires can never be effective and are therefore not desires, according to the motivational perspective. While quadriplegia and similar causes are explicable in motivational terms, pleasure theorists do raise a valid point. Organisms, especially human beings, do at times desire things that action is ill-suited to bring about. Merely saying that these things are not desires seems unsatisfactory.

Galen Strawson (1994) tries to defend the pleasure-based perspective by introducing us to the Weather Watchers. Weather Watchers are individuals who watch the weather. They are incapable of movement, like quadriplegics. But unlike quadriplegics, who do not move because of malfunction, Weather Watchers do not move because they do not have the capability to produce movement. Weather Watchers, however, have an emotive relationship with the weather: they prefer sunshine to rain, and are therefore happy when the sun shines and sad when it does not. Do Weather Watchers have desires, namely, desires for more sunshine and less rain? Strawson says we ought to view Weather Watchers as possessing desires even though they do not have dispositions to action. Weather Watchers have desires because they feel pleasure and displeasure relative to certain states of affairs.

Strawson has in mind imaginary entities when he speaks about Weather Watchers. He uses this example to illustrate that it is not logically impossible to conceive of organisms that cannot move yet still have desire. The example used by Strawson is important in that while it refers to purely imaginary cases, there are empirical cases which have properties in common with the

imaginary case. Individuals with certain neurological deficits can present with symptoms like lacking dispositions to action. Consider a person with akinesia. People with akinesia suffer from a loss or impairment of voluntary movement. Can such a person have desires? Unlike a quadriplegic — a person who is paralyzed but has the will to move — a person with akinesia lacks the *will* to move. According to motivational theories of desire, a person of this sort does not have desire. They lack desire not because they lack dispositions to action, but rather because they lack the function of producing dispositions to action like the Weather Watchers. Strawson's Weather Watchers example is interesting because it cannot be accounted for in motivational terms and as a consequence brings to the fore an interesting conflict. Being able to explain the quadriplegia case in motivational terms maintains the intuition that desire is inherently about action. The Weather Watchers example, on the other hand, maintains the intuition that at least some of our desires are not motivational.

Our intuitions seem to be in conflict. It seems both intuitions cannot be right. There is, however, a way in which these two intuitions can be aligned. Earlier in the chapter, I made the claim that pleasure theorists could argue that their theory is correct because of its broader applicability. I did not, however, fully interrogate what "broader applicability" means. We still do not know why certain things have the motivations they do; we do not know why certain things dispose us toward action. One way of explaining this phenomenon is by making the claim that organisms are disposed towards actions that result in pleasure. Pleasure theorists could claim that action results in pleasure, and as a consequence, action is a means to pleasure. Organisms are inherently disposed to pleasure, and *not* action, and as such can feel pleasure even in cases where pleasure is not derived from action. In essence, organisms are hedonists. If this is true, the pleasure thesis can subsume the motivational thesis in that it explains motivation in pleasure terms, thus preserving both our intuitions. How would we prove this? One way of doing this would be by turning to neuroscience.

Why neuroscience? It is well established that the mind is what the brain does. Given this understanding, how the brain works will give us insights into how things actually work, as oppose to how we perceive them to work. Neuroscience can be used to establish whether or not our intuitions align with empirical evidence. If the claim that organisms are disposed towards pleasure (and are therefore motivated to act in ways that result in pleasure) is true, we ought to see connectivity between the parts of the brain that are responsible for motor function and pleasure respectively, assuming these functions are executed by different systems. Alternatively, these two functions could be facilitated by the same part of the brain. If the parts of the brain that facilitate motor function and pleasure are separate and have no connectivity, it leads us back to the initial position, a point where we have two intuitions and no way to decide. What follows will be brief and schematic neuroscientific exegesis. The following exegesis will be light on detail because the purpose is to establish whether or not the neural correlates of motor function and pleasure are different, connected, or the same. This can be done without exploring the intricacies of how each of the parts that collectively produce motor function and pleasure function.

Evidence from neurology and neuroscience suggests that the motor and pleasure systems have different neural correlates. Voluntary movement, movement generally, is controlled by the

motor cortex: “The remarkable findings of Penfield and Boldrey which have been supported by many subsequent studies emphasize a key role for motor cortex in mammalian movement control.” (Matyas *et al*, 2010:1240).The motor cortex forms part of the cerebral cortex. The motor cortex is involved in the planning, control, and execution of voluntary movement. The motor cortex can be divided into three major parts: the primary motor cortex, the premotor cortex and the supplementary motor area: “…the motor cortex could be divided into a primary motor area (area 4) and a premotor area (area 6).” (Chouinard *et al*, 2006:143). There are other parts of the cerebral cortex, parts that do not form part of the motor cortex but which also play a pivotal role in the planning and execution of movement. Chief amongst these areas are the posterior parietal cortex and the primary somatosensory cortex. Control of motor function is, however, not limited to the cerebral cortex. There are other brain regions that play an important part in motor function: notably, the cerebellum and basal ganglia.

Pleasure-activated brain networks are widespread. The following brain regions and brain parts are associated with the production of pleasure: the nucleus accumbens, the ventral pallidum and the brainstem: “The medial shell of nucleus accumbens (NAc) contributes to the hedonic impact of sensory pleasures (e.g., “liking” reactions to sweetness)…” (Catro *et al*, 2014: 4239). There are other candidates for pleasure which are found in the cortex: areas like the orbitofrontal, cingulate, medial prefrontal and insular cortices. The limbic system is also implicated in the production of pleasure, given that pleasure is mostly associated with positive emotion and emotion is controlled by the limbic system. The abovementioned brain regions are in no way exhaustive. The main idea is to show, without giving an exhaustive list, that there is a separation between the parts of the brain that produce movement, (the motor cortex), and those that determine the hedonic tone of stimuli (the pleasure centre[s]).

The fact that we have functional separation does not imply that we have functional independence. There are times when there is functional specialisation in the brain and as a consequence different parts of the brain handle different functions. Functional separation does not however always indicate functional independence. There are times when functionally specialised units of the brain interact with each other. This interaction leads to either connectivity or reciprocal connectivity amongst functionally separate systems. Where motor function and pleasure are concerned we find both functional separation and lack of connectivity or reciprocal connectivity. The difference in neural correlates and lack of connectivity between these neurological structures suggests that we cannot make the claim that pleasure results in motivation. We, therefore, cannot accept the broad interpretation of the desire-as-pleasure thesis, thus explaining motivation in terms of pleasure. This brings us back to our original position. We again find ourselves with two contending intuitions: is desire motivational or pleasure-based? The functional separation and functional independence of the motor and pleasure systems do not resolve this question.

Functional separation and functional independence lend credibility to the claim that an organism could be immobile – specifically, lacking the will to move – but still desire. An organism desires in this situation if the parts of its brain that determine pleasure are functional, and the organism can, therefore, feel pleasure at it seeming like ***P***. The claim that there is functional separation and functional independence between the motor and pleasure system does

not imply that desire as dispositions to action is implausible. Functional separation and functional independence also lends credibility to the claim that desire can be understood as dispositions to action. Just as it is possible to have an immobile organism desiring if it feels pleasure, it is also possible for an organism that does not feel pleasure to have desire if it has dispositions to action.

The following example illustrates the fact that motivation and pleasure can be dissociated such that an organism could be motivated whilst not feeling or deriving pleasure from what is desired (motivated): consider the case of the willing, the unwilling, and wanton addicts as discussed by Frankfurt. All three types of addicts may have first-order desire conflicts: they all may desire to take and also not to take a drug. The difference between the three is that the willing addict endorses the-take-the-drug desire, while the unwilling addict endorses the-do-not-take-the-drug-desire. The wanton addict endorses neither desire: he is indifferent. The willing addict's endorsement of the drug can be explained in pleasure terms. The willing addict endorses the-take-the-drug-desire because the drug gives him pleasure. The unwilling addict does not endorse the-take-the-drug-desire because he does not derive pleasure from the drug or, if he does, he does not want drug-induced pleasure. Both cases lead to desires that can be accounted for using the pleasure thesis. The wanton addict is different and therefore interesting: because the wanton addict fails to endorse any of the urges, one can argue that he is not moved by considerations pertaining to pleasure or displeasure. This means that the wanton addict's desire cannot be accounted for using the pleasure thesis and is, therefore, an example of how one can desire without pleasure.

The indifference of the wanton addict may cause some to question addiction as described by Frankfurt and furthermore the conclusions drawn from it. Addiction implies compulsion; compulsion does not, however, imply indifference, as the wanton addict case suggests. Compulsion also complicates issues pertaining to autonomy. We often think that when a person is compelled, they are not free / they do not have autonomy. If addiction is synonymous to compulsion, the idea of a willing addict does not make sense. Mark Leon's (2001) critique of willing addiction presents a better account of addiction and consequently a better account of pleasure-neutral desiring. Leon does not set out to give an account of pleasure neutral desiring but it can be inferred from his work. Leon sets out to illustrate that an addict can never act willing because addiction, properly understood, implies irresistibility or compulsion. When an addiction is irresistible, it operates whether or not it is endorsed by an addict. The fact that addiction operates without endorsement, and that the addict is passive with respect to his desire, implies that the desire could very well not result in pleasure and yet still be motivational, i.e., dispose one toward action, simply because it is irresistible. Neuroscientific evidence supports this conclusion. Neuroscientists say that liking and wanting (i.e., desiring) are different, and as a consequence one can want (one can desire) without liking: "At extreme, the addict may come to "want" what is neither "liked" nor expected to be liked, a dissociation possible because "wanting" mechanisms are largely subcortical and separable from cortically mediated declarative expectation and conscious planning." (Kringelbach *et al*, 2010:668).

As things stand, we have no way of determining which approach to desire is correct. We lack evidence to show that pleasure and motivation can be unified. We do, however, know that at

times we desire things that actions are ill-suited to bring about, while at other times we desire things that action is best-suited to bring about. Given that we need to explain both circumstances, things that action is ill-suited to bring about and things that action is best-suited to bring about, we have to have a broad theory. The need to account for things action is ill-suited to bring about means we cannot endorse desire-as-motivation because this would lead to exclusions. We have to endorse desire-as-pleasure because it has broader applicability. Endorsing the pleasure-based approach to desire requires making an *ad hoc* manoeuvre. We have to explain motivation in pleasure terms, yet the evidence maintains that we cannot. To counter this we can merely stipulate that given the fact that the explanation of motivation in terms of pleasure is both logically possible and intuitively appealing, it ought to be accepted, hence the claim to broader applicability. This move is *ad hoc* but it at least gets us to a theory. Pleasure-based theories of desire, however, face another problem. If it is true that desire ought to be understood in terms of pleasure, it ought to be the case that an organism cannot desire something that is un-pleasurable. Frankfurt's unwilling addict illustrates that this is not the case. It may be worthwhile to make the point with another example.

A proponent of the pleasure thesis could make the argument that the lack of pleasure experienced by an unwilling addict refers to psyche dis-pleasure, displeasure as a result of not being able to constitute his will as he pleases. This form of displeasure is nothing more than frustration. Frustration is explicable in terms of displeasure but it is not dis-pleasure in the truest sense. Dis-pleasure in its truest sense refers to physical discomfort, as opposed to psyche discomfort. As a consequence, while the unwilling addict can be spoken of as desiring something that causes him displeasure, the dis-pleasure experienced by the unwilling addict is not genuine. This seems *ad hoc*, but we can accept it. An addict experiencing withdrawal experiences physical discomfort. That discomfort is ameliorated by the consumption of drugs; yet, when the drug is consumed and the physical pain of withdrawal is over, the addict experiences dis-pleasure associated with having a weak will. In light of this, the distinction between psyche and physical dis-pleasure makes sense.

If we accept this distinction we have to accept that it is possible to desire things that cause psyche dis-pleasure, but that it is impossible to desire things that cause physical displeasure. If we accept this, the argument that one cannot desire something that causes dis-pleasure stands but only if dis-pleasure refers to physical displeasure. It is a known fact that this is not the case. We know as a matter of fact that organisms can desire things that cause physical displeasure. Consider masochism, the derivation of pleasure from physical pain. If it were true that an organism would never desire physical pain, masochists and masochism would not exist, but they do: "There are certainly cases where pain increases sensual pleasure, and there have been experiments where animals have been fooled into doing more of a behaviour the more it brought on pain" (Ainslie, 2001:58). A proponent of the pleasure thesis could explain this by arguing that masochistic organisms experience pleasure where non-masochistic organisms experience pain, but I do not think that this is right. I do not think that masochists misinterpret painful stimuli for pleasurable stimuli. I think it is pain *qua* pain that results in pleasure for masochists. Where a non-masochist feels pain, and as a result displeasure (prompting them to discontinue the activity that is the cause of the pain), a masochist feels pain and *then* pleasure,

and as a result, may continue the activity. Masochism illustrates that organisms can desire things that cause them displeasure or pain. I do not intend to pursue masochist argument any further. My intention was to illustrate that even an expansive, liberal interpretation of the pleasure thesis still has problems.

We could perhaps explain desire in terms of goodness. Defining desire in terms of goodness entails assuming that a desire for ***P*** is based on a belief that ***P*** is good. There are many iterations of the guise of the good theory. One could use Aristotle's formulation or more contemporary articulations from philosophers like Joseph Raz. Irrespective of what articulation of the guise of the good theory one uses, one ends up with the assumption that desiring something means that the thing being desire is thought to be good. A good based theory of desire presents a hybrid account of desire in comparison to the abovementioned theories. Motivational theories of desire state that we desire when we have a disposition to act. Pleasure theories of desire state that we desire when have a disposition to pleasure. Good based theories of desire have a motivational component. A good based theory is however different from a motivational theory because, under the good based theory, a person is motivated by the good, whereas under motivational theories a person merely has motivation. This theory solves one of the major problems we face when defending motivational accounts: the problem of what motivates. If we define desires as aiming at the good whenever we are motivated or disposed toward act, we are motivated because we assume that the action will result in the good. Desire as goodness can also help explain pleasure. We can claim that organisms desire what is good and identify what is good in virtue of whether or not it solicits pleasure. As a consequence, the pursuit of pleasure is actually the pursuit of the good. This definition incorporates the pleasure thesis in both its egocentric and altruistic forms. It seems as if the guise of the good theory solves our problems. It gives us a means of accounting for both our intuitions: the intuition that desires is both motivational and directed at pleasure.

The guise of the good theory is however not without its problems. What does it mean to say something is good? How do we define the good? I will sidestep these issues and focus on issues that I think are more telling. The guise of the good theory is based on having beliefs about goodness: this is to say that an organism must believe that something is good for it to be desirable. A belief based conception of the good is preferred because it allows for the possibility of desires which would follow as a matter of necessity once one knows the good.

The above claim is predicated on two assumptions. The first is that desires have beliefs as their proper determinants. The second assumption is that desires always track belief. However, having a particular belief does not entail having a corresponding desire. The manner in which desires are generated can malfunction. As a consequence of malfunction, being in possession of beliefs about the good does not entail having desires that track the good. A proponent of the guise of the good thesis could defend themselves in a similar manner to the proponents of motivational accounts. Proponents of motivational accounts state that irrespective of whether or not dispositions to action are created, or whether the right dispositions are created, is irrelevant, just as long as the function of disposing action is maintained. I do not, however, think that this defence applies to the guise of the good theory.

Function, as it applies to disposing action, refers to the motor cortex and the function in this regard is understood in teleosemantic or biological (and therefore evolutionary) terms. A motor cortex is a motor cortex in virtue of having the appropriate evolutionary lineage and as such even if it fails at disposing action still possesses its function (Millikan 1984). Goodness, on the other hand, is not explicable or rather should not be explicated in teleological terms like biological function. Goodness is understood as an evaluative concept, not a biological entity and therefore not something that is heritable. Evaluative concepts are acquired but not in a manner in which biological functions are acquired, and therefore ought not to be understood in teleological terms.

If goodness is an evaluative concept and desiring requires thinking or believing that something is good, do organisms that cannot think about the good or have not yet acquired goodness as a conceptual schema have desires? How can children, for example, have desires? Velleman writes that "the capacity to desire requires the possession of evaluative concepts. Yet a young child can want things long before it has acquired the concept of their being worth wanting, or desirable. Surely, the concept of desirability-of something's being a correct or fitting object of desire-is a concept that children need to be taught." (Velleman, 1992:7). Assessments of goodness seem to be rationalisations intended to ameliorate issues emanating from the fact that cashing out desire using only dispositions to action seems unsatisfactory. Beliefs, as I understand, would lead to desire even if the beliefs were not viewed as good. Making the claim that beliefs are good and therefore desire is good is not necessary for desiring as such; it is, however, necessary for rationalising action. Nothing prevents children from acquiring beliefs and, from those beliefs, desires. It does, however, seem that children do not have a conception of the good, given that part of child-rearing is inculcating concepts of the good and bad and the right and wrong.

Children present an interesting difficulty for the guise of the good. We assume children have desires because they exhibit behaviours that we associate with desiring. If a door suddenly opens and a child makes a dash for the door, we can assume, uncontroversially, that the child makes a dash for the door because they desire to go outside. Under the guise of the good thesis, we would have to assume that the child in question desires going outside because there is something good about going outside, but this does not seem plausible. Velleman responds by giving an account of the good which does not focus on evaluative or normative concepts: "…to say that the desire involves regarding *p* as good is not to say that it consists in a judgment with an evaluative proposition as its object." (Velleman, 1992:8). If regarding *p* as good is not to say that it consists in a judgment with an evaluative proposition as its object, what does regarding *p* as good mean? Velleman claims that regarding *p* as good (and therefore desirable) entails assuming that *p* is attainable: "One cannot desire something if it seems impossible or if it seems already to have come about; one can desire that p only if p seems attainable, in the sense of being a possible future outcome." (Velleman, 1992:17).

Velleman provides a solution for the problem posed by child cases. However, this is not the end of the problems faced by guise of the good theories. If people desire the good it ought to be the case that people cannot desire something that is bad, or at the very least, something that they know is not good. We do, however, know that people can and do desire things that are bad

either knowingly or unknowingly. There are cases where organisms might desire things that they do not value or think of as good: " First, it is possible that what one desires is not to *any degree* valued, held to be worthwhile, or thought good; one assigns no value whatsoever to the object of one's desire." (Watson, 1975:210). An organism might also find itself in a situation where it values something and therefore desires it but it also desires something else, something considered not as valuable, yet desired more than the valued object or state of affairs. The thing which is most desired, if goodness or value determines desire, ought to be more highly valued, but this does not, however, follow as a matter of necessity. A person may value one thing more than he values something else, but the thing that is most valued may not be the thing most desired and the thing that therefore motivates the person: "Secondly, although one may indeed value what is desired, the strength of one's desire may not properly reflect the degree to which one values its object; that is, although the object of desire is valuable, it may not be deemed the most valuable in the situation yet one's desire for it may be stronger than the want for what is most valued." (Watson, 1975:210). As an illustration, let us return to Frankfurt's unwilling addict. Frankfurt's unwilling addict values – desires — sobriety. However, he also values/desires satiating his craving. In virtue of being an unwilling addict and thus having a second-order volition (a preference about which first-order desire ought to be effective), the unwilling addict can be assumed to value and thus desire sobriety *more*, since he is unwilling in his addiction. This valuation does not, however, result in him being motivated by the thing that he values the most. The unwilling addict is instead motivated by the thing that he values the least.

The above discussion focused on the good understood normatively; this, however, is not the only way in which the good can be defined. The notion of goodness can be defined in egocentric terms, where good means anything that benefits the organism in question. If goodness is cashed out in egocentric terms it ought to be the case that organisms do not engage in behaviours that go against their own interest. Perhaps, more stringently, it ought to be the case that organisms do not (or rather, cannot) desire things that cause them harm, because harm is not in their self-interest. Evolutionary theory suggests that organisms can and do often desire outcomes that are not good for them, i.e., are not in their interest, and at times, desire outcomes that harm them if those outcomes increase evolutionary fitness. Evolutionary theory says that when organisms are making fitness-enhancing decisions, they look to the propagation of genes as opposed to the survival of specific instantiations of those genes. This is due to the fact that the unit of selection is the gene as opposed to a specific organism: "The evolution of species occurs through the survival of genes, not of individuals." (Ainslie, 2001:45). If the unit of selection, and therefore the thing that ought to survive, is the *gene*, it is possible for an organism to desire states of affairs (or engage in behaviours) that result in *harm* to the organism, yet *increase* the probability of genes being propagated: "There are many familiar cases where an individual organism must sacrifice itself to maximize survival of its genes." (Ainslie, 2001:45). The notion of self-sacrifice consists in dis-pleasure (physical dis-pleasure): the very thing that a good-/egocentric-based account of desire says an organism cannot or ought not to desire. It seems that good based definitions of desire also have problems: problems which render them untenable.

To summarise so far: while theories that account for desire in terms of dispositions to pleasure incorporate things that we normally think of as desires (but which we would have to exclude under motivational theories), desire as dispositions to pleasure still seems deficient. Desire seems to be intrinsically linked to motivation, such that speaking about desire elicits images of action. Desire could also be explained in terms of the guise of the good. Adopting this perspective gives us a manner of accounting for both pleasure and motivation, but is still not satisfactory. Desire as goodness is unsatisfactory because of what we have to accept as a trade-off for a theory that explains motivation and pleasure. I think that it is unnecessary for us to explain desire as goodness, or at the very least, if we describe desire in this fashion, we are referring to a rationalisation, as opposed to claiming the causal necessity of goodness in the creation of desire. Therefore, it seems as though we require a new theory of desire that explains both pleasure and motivation, given that it seems that pleasure and motivation play an important role in desiring. In the next chapter, I will discuss such a theory.

Chapter Two

Reward Theory of desire

In this chapter, I am going to explain the reward theory of desire. I am going to argue that we ought to accept this theory because it gives us a better, more holistic account of desire. The reward theory of desire accounts for all the phenomena – pleasure, motivation and goodness – which we associate with desire. The reward theory of desire states that an organism has a desire if it finds something rewarding; reward can be interpreted in terms of pleasure, motivation and goodness. The process of being rewarding, and thus being a desire, consists in two components:

incentive salience (or, learning), and motivation. Incentive salience is the type of processing that results in organisms learning which features or facets of their environment are rewarding, i.e., which facets result in spikes of dopamine. Motivation, on the other hand, refers to learning the accompanying actions which, when performed in that environment, result in a reward; as a consequence of the reward, these actions are likely to be repeated. Desire according to the reward theory of desire, therefore, refers to the repetition of actions: actions that lead to the secretion of dopamine. A reward based definition of desire may be surprising to people who associate desire with either pleasure or other phenomena. This theory, however, does provide an explanation of how reward and the other phenomena that we associate with desire intersect.

To begin this chapter in earnest I will tackle issues pertaining to the naturalization of desire. Schroeder and Butlin agree on most aspects in this regard, especially the fact that desire can be naturalised. They both accept the idea that desire emerges as a consequence of the function of the reward system. Consequently, to understand desire, one must understand how the reward system functions, and to do this, one must have a firm handle on psychology especially experimental psychology, and how it conceptualises reward. One also needs a firm handle on neuroscience, given the fact that reward is associated with the dopaminergic system, the function of which explains the results derived from experimental psychology. The merits of this type of approach to desire – a naturalized approach – were hopefully made clear in the previous chapter. When we were confronted with contending intuitions, neuroscientific evidence was introduced with the intention of helping us work out which, if any, of our intuitions were correct by seeing which if any would be borne out by the data. While Schroeder and Butlin agree on the fact that desire can be naturalised and therefore echo each other's sentiments on some issues, they also differ significantly with respect to other issues.

Schroeder and Butlin not only agree on the methodology, naturalisation, which should be used to study desire they also agree on the fact that desire can be understood as a natural kind. They differ, however, with respect to the type of natural kind desire is. Schroeder thinks that for something to count as a desire, it has to facilitate reward-based learning. Schroeder consequently thinks that the reward system provides mechanisms that facilitate reward-based learning. Butlin, on the other hand, believes that for something to count as a desire, it has to function as an input used by the goal-directed system for the purposes of tracking the reward values of outcomes. Butlin thinks that for something to count as a desire it has to facilitate reward-based learning, and that this reward-based learning should be used as an input in the goal-directed system, thus helping with action selection: learning must be used to bring about the most rewarding action.

It is interesting to note where Schroeder and Butlin differ with respect to their definition of desire's natural kind. The difference maps perfectly onto the intuitive differences that we grappled with in the previous chapter. Schroeder's natural kind definition maps onto pleasure-based accounts of desire, in that motivation (and therefore action) is not viewed as a necessary component of desire. Butlin's natural kind definition maps onto motivation-based accounts of desire because it has action selection as its ultimate goal. It is important to note that while Schroeder's account is commensurate with the pleasure thesis, Schroeder does not hold the view that the reward system realises pleasure; rather, the system has pleasure as one of its

effects. This point will become clearer when I give a more detailed account of the differences between Schroder and Butlin's accounts of desire.

Schroeder and Butlin agree that desire can be naturalised or understood as a natural kind. They also agree that in certain organisms, desire is realised by the dopamine or reward system, and they have consensus on which dopamine signals are responsible for learning and thus desire. However, where the dopamine-releasing system is concerned, Schroeder and Butlin differ with respect to which dopamine signals are responsible for the creation of motivation. Before I discuss the specificities of dopamine signals, I will discuss the dopamine-releasing system in general.

What is dopamine and what is its function? Dopamine is a neurotransmitter, more specifically, "a substance that transmits information from one nerve cell of the brain to another." (Caplin *et al.*, 2008: 663). Dopamine transmits information pertaining to reward or desire. Dopamine controls incentive salience in an organism, which refers to the assignment of appetitive value to a stimulus. A stimulus with appetitive value teaches an organism which features of its environment are rewarding and subsequently results in the initiation of actions associated with acquiring the reward in question: "Recent theoretical and experimental work on dopamine release has focused on the role this neurotransmitter plays in learning and the resulting choice behaviour." (Caplin *et al.*, 2008: 663). Both Schroeder and Butlin agree with Caplin *et al.* about the role played by dopamine where learning is concerned; however, they differ with respect to dopamine's role in affecting behaviour. To be clear both accept that dopamine plays a pivotal role in the generation of action. They differ with respect to the centrality of motivation where desire is concerned.

There are four main dopaminergic tracts or pathways in the brain: the mesocortical, mesolimbic, tuberoinfundibular, and nigrostriatal pathways (Ayano, 2016). Of these four pathways, only three comprise the reward system. The reward system is comprised of the mesolimbic, mesocortical and nigrostriatal pathways. The reward system receives dopamine from two main structures: the ventral tegmental area and the substantia nigra. The ventral tegmental area and substantia nigra are the major producers of dopamine in the brain. For now, it will suffice to just mention these pathways. At a later time, I will mention them again with specific reference to the other parts of the brain that they connect to, to create the reward pathways. For the moment I will look at dopamine and its relationship to reward.

How does dopamine relate to reward, to desire? Unlike the previous theories which attempted to give an *a priori* definition of desire (desire as either pleasure or the good), the reward-based theory of desire claims that desire can only be understood *a posteriori* where *a posteriori* refers to the function of the dopaminergic system. Put another way, in organisms which have a dopaminergic system, desire, a natural kind, is realised by the function of the dopaminergic system or the release of dopamine. When an organism receives a reward it increases the amount of dopamine in the organism. It may seem strange that I am making the claim that dopamine indicates reward, and yet receiving a reward increases the amount of dopamine in the brain and does not introduce it. This point will be cleared up when I discuss discrepancies in the firing rate of dopamine-producing neurons.

Given our inability to state *a priori* what it means to have a desire, it is difficult to specify what causes the activation of the dopaminergic system and thus the release of dopamine. The only thing that can be said in this regard is that the release of dopamine by the dopaminergic system corresponds with an organism's interaction with the environment. What about the environment causes activation of the dopaminergic system? This question is difficult to answer, or at the very least, difficult to answer without seeming circular given the fact that not all interactions with the environment result in the secretion of dopamine (only rewarding interactions result in the secretion of dopamine). One way of solving this problem, stating what is rewarding, would be by claiming that reward is associated with fitness and as such only interactions with the environment which increase or have the potential to increase fitness are desirable.

The release of dopamine, an indication of reward, is triggered by the acquisition of a rewarding object. When an organism experiences a reward for the first time, for example, gets a lump of sugar, the organism's dopaminergic system releases dopamine in response to receiving the rewarding object (in this case, a lump of sugar). Novel and random cases of reward result in the release of dopamine in response to the attainment of the rewarding object. The release of dopamine, however, changes in cases where the reward that is being received is no longer novel or random. When an organism learns that there is a cue(s) that fully and accurately predicts a reward, the dopaminergic system no longer releases dopamine in response to receiving the rewarding object, but rather in response to the predicting cue. This means that when an organism receives an indication that a reward is imminent, it releases dopamine in anticipation of that reward. Caplin gives the following report concerning this phenomenon where monkeys receive juice alongside an audio tone:

> "Initially (i.e., before the animal had learned to associate the tone with the juice), dopamine neurons fired in response to the *juice* but not the *tone*. However, once the monkey had learned that the tone predicted the arrival of juice, dopamine responded to the tone, but now did *not* respond to the juice." (Caplin *et al*, 2008: 667).

The release of dopamine by the brain results in learning. The organism first learns about the rewarding nature of its environment, namely, which objects are rewarding. The organism also learns *how* rewarding those objects are. This can be considered the first stage of learning: dopaminergic signalling in response to the acquisition of a rewarding object. At this stage, the organism learns what is rewarding and how rewarding it is. If a reward can be predicted on the basis of a cue, the organism learns how to transpose reward information from the rewarding object to the predicting cue, such that, in future encounters, it is the presentation of the *cue* that triggers the release of dopamine. When the transposition of reward information occurs, the organism has effectively learned about the rewarding nature of the cue. We can also state that the organism has learned about the predictive nature of the cue, as opposed to the cue's rewarding nature, because it is not technically the cue that rewards. This is the second stage of learning, signified by dopaminergic signalling in response to the presence of a reward-predicting cue. One could easily assume that dopamine-releasing neurons fire arbitrarily. By arbitrary, I mean that one could assume that dopaminergic neuronal fire merely *represents* reward but does not *quantify* it. However, this is not the case.

The amount of dopamine that is released codes for both the reward as well as the size of the reward. This means that when an organism responds to a cue it is not only responding to an impending reward: it is also reacting to, or anticipating, a reward of a specific size: "These estimations of value, or value functions, encode the average amount of reward." (Bayer *et al.*, 2005:129). Why does the dopaminergic system code for reward size? Coding for reward size makes sense if you view desire like Butlin. Butlin claims that desire influences behaviour. If desire influences behaviour, organisms have to be sure that they are pursuing the most rewarding desires and doing so efficiently. Secondly, the environment changes and as a consequence, the reward value of objects might change, so organisms need to constantly re-evaluate whether the objects they desire still confer reward, tailoring their behaviour appropriately. The process of re-evaluating reward value can only be undertaken if dopamine codes for both reward and reward values. But how is reward value re-evaluated?

If a reward can be predicted on the basis of a cue, the cue tokens a representation of that reward as well as its value. When the organism encounters a reward-predicting cue, the tokened representation is activated, and this results in the release of dopamine in a quantity equivalent to the previous time the cue was encountered. The organism then compares the amount of anticipatory reward, dopamine, with the actual amount of reward which is experienced. If the reward that is received matches up to the predicted reward, dopaminergic neuronal activity dissipates. If, on the other hand, the organism experiences *more* reward than anticipated, the dopamine-releasing neurons show more or stronger activity. If the reward expected is lower than the reward received, the neuronal fire of dopamine-secreting neurons is lowered. Schroeder explains:

> "VTA/SNpc neurons fire in a pattern that carries information about the difference, at time *t*, between the rewards received and expected at *t* versus those rewards the organism was predicting (at t-1) it would receive or expect at *t*. Or, to put things in a more intuitive if less precise manner, VTA/SNpc neurons signal the difference between how good the world was predicted to look at *t* and how good it, in fact, looks at *t.*" (Schroeder, 2004:50).

To understand how dopamine-releasing neurons function we have to understand prediction error. A prediction error is an estimate of some future event. When the predicted future event occurs an organism compares the prediction to the actual outcome to see if there are any discrepancies: "To understand prediction errors, we distinguish between a prediction about a future reward, or no prediction and the subsequent reward. Then we compare the reward with the prediction; the reward is either better than, equal to, or worse than its prediction." (Shultz, 2016:24). Dopamine releasing neurons function on the basis of a reward prediction error in that they represent the difference between a predicted reward – prediction based on past experience – and the reward that is actually received "The dopamine response thus reflects a reward prediction error and can be described by the simple difference between obtained and predicted reward." (Shultz, 2016:25).

Having discussed when and why dopamine is released I now turn to <u>how</u> dopamine is released. There are two types of dopamine signalling: phasic and tonic. Phasic dopamine signalling can

also be referred to as burst signalling in that this type of signalling occurs in bursts which result in a sharp increase and decrease of neuronal fire and thus dopamine in the brain: "In their phasic mode, DA neurons sharply increase or decrease their firing rates for 100–500ms, causing large changes in DA concentrations in downstream structures lasting for several seconds." (Bromberg-Martin et al, 2010:815). Burst signalling occurs as a direct response to a rewarding object or reward-predicting cue. Given the fact that reward prediction errors result in either the termination of or increase in neuronal fire one can assume, with confidence, that phasic signalling is responsible for reward error prediction and thus reward learning: "These phasic DA responses are triggered by many types of rewards and reward-related sensory and are ideally positioned to fulfill DA's roles in motivational control, including its roles as a teaching signal that underlies reinforcement learning." (Bromberg-Martin et al, 2010:815). Dopamine is also available in a more consistent manner via tonic signalling.

Tonic signalling is responsible for a systematic release of dopamine as opposed to the burst signalling of the phasic phase: "In their tonic mode, DA neurons maintain a steady, baseline level of DA in downstream neural structures that is vital for enabling the normal functions of neural circuits." (Bromberg-Martin et al, 2010:815). A tonic or systematic release of dopamine accounts for why reward and reward prediction errors are associated with spikes (bursts) in dopamine (known as phasic signalling), as opposed to the mere presence of dopamine. If burst signalling is responsible for learning, reward error prediction, and reinforcement, what is the function of tonic signalling? There is disagreement about the function of tonic signalling. Dopamine is implicated in the generation of motivation and action. This claim is supported by the fact that individuals suffering from Parkinson's disease, a disease that severely affects action, have low levels of dopamine as a consequence of the death of dopaminergic cells of the substantia nigra. This clearly illustrates that dopamine plays a crucial role in the production of action; however, it is not clear which type of dopamine signalling produces action. Schroeder and Berridge are of the view that phasic signalling is responsible for action while philosophers like Schultz (2007) and Niv *et al.* (2007) believe that tonic signalling is responsible for the initiation and control of action.

Lack of agreement about the function of tonic signalling leaves room for speculation. I am of the opinion that tonic signalling is responsible for action. The fact that low levels of dopamine are associated with Parkinson's implies that action must be supported by a slow and systematic release of dopamine. If action was supported by burst signalling which responds to the presence of rewarding objects and reward-predicting cues (reward error prediction), the dopamine required for action would be available intermittently and as a result, organisms would experience Parkinson's like symptoms between rewards, but this does not happen. Given this I think we can, and ought to, accept the fact that tonic signalling is responsible for action. Phasic signalling is responsible for incentive salience, while the rewarding actions that lead to the attainment of reward are under the purview of tonic signalling.

Reward prediction errors are necessary for learning. Reward prediction errors result in reinforcement learning: "… the dopamine neurons of the midbrain encode, in firing rate, a reward prediction error signal that could be used to drive a reinforcement learning system."

(Bayer *et al*, 2005:137). Both Schroeder and Butlin agree that reward prediction errors result in reinforcement learning; however, they differ on what is learned.

Schroeder is of the opinion that what is learnt, and therefore reinforced, is either a positive or negative representation of a cue: "For an event to be a reward for an organism is for representations of that event to tend to contribute to the production of a reinforcement signal in the organism, in the sense made clear by computational theories of what is called 'reinforcement learning'." (Schroeder, 2004:66). Schroeder has in mind learning to represent *P* in such a way that the occurrence of *P* comprises a reward. The emphasis that is placed on representation is intended to underscore the fact that motivation, conceptions of goodness, and pleasure are not constitutive of desire. This seems counterintuitive. Our intuitions tell us that desire is inherently motivational and pleasurable. Schroeder does, however, cater for the sway of our intuitions by making the claim that while motivation, conceptions of goodness, and pleasure are not constitutive of desire, they are causally connected to desire: "Desires are, in principle, independent of motivation, independent of good and bad feelings, independent of where one's attention turns or what habits one develops. Desires are independent of all these things in principle, but causally connected to them in fact, and so can explain all of these things." (Schroeder, 2006:637).

Schroeder explains the concept of a causal connection in sufficient as opposed to necessary terms by claiming that: "…if the reward theory of desire is correct, then having a powerful disposition to action is just a natural consequence of having a powerful desire, not constitutive of it." (Schroeder, 2004:139). Schroeder does this because he wants to maintain a connection between desire and the phenomenon we associate with desire without the connection being necessary. Schroeder makes this claim because he is aware that it does not follow that when an organism has a desire, that it is moved (motivated) or pleased by that desire. Frankfurt and Watson also make this point. Frankfurt makes a distinction between effective desires and what would be defined as ineffective desires [*my terminology*]. Watson, on the other hand, makes a distinction between desiring and valuing. Schroeder is of the opinion that if Frankfurt and Watson are correct, a theory of desire cashed out in motivational and pleasure terms could be viewed as incorrect. Schroeder could avoid this situation. Schroeder could claim that desire by his account results in dispositions to action, yet the mere possession of a disposition does not entail action (especially in cases where there are countervailing reasons). This claim accounts for desire in motivational terms, while making room for desires that do not result in action; it therefore handles Frankfurtian and Watsonian cases.

Like Schroeder, Butlin believes that the reward associated with a predicting cue will be reinforced. Butlin, however, believes that it is a cue *plus* the performance of an action which results in a reward. As a result, Butlin believes that reinforcement learning applies to both cue and action. Butlin is of the opinion that while a cue may predict a reward, the reward is only attained by the performance of an action. Consider a situation where a rat learns that juice is rewarding and then later learns that pressing a lever results in juice. The rat learns two things, what is rewarding, juice, and what leads to the reward, lever pressing. Both things are reinforced by the encounters in which lever pressing leads to juice and juice is rewarding.

Butlin thinks that action is an integral part of desire. Butlin however makes a distinction between actions that are under the purview of what he calls the goal-directed system and those that are under what he refers to as the habit system: "In behavioural psychology and cognitive neuroscience, a successful research programme distinguishes between two systems for action selection, which are called the habit system and the goal-directed system." (Butlin, 2017:619). Butlin makes the claim that a very specific form of motivation, the motivation that emerges from the goal-directed action selection system counts or ought to count as desire. Why must desire be identified with goal-directed action selection?

As seen, one of the ways in which Schroeder can account for Frankfurtian and Watsonian cases is by claiming that desire is a disposition to action. But also seen, claiming that desire is a disposition to action still presents some difficulties for Schroeder. Desire understood as a disposition to action implies intentional action. Intentional action does not, however, imply efficient action selection. Butlin identifies desire with efficient action selection as opposed to intentionality. To be clear actions need to be performed intentionally, however, the organism needs to determine whether the action is best-suited to bring about a reward. To clarify Bultin's position consider the difference between habitual and contingent actions. A habitual action is an action one performs spontaneously usually as a consequence of a cue. If the action is paired with a reward such that the performance of the action results in a reward, an organism will associate the action with a reward. If this association is reinforced, when an organism encounters the cue it performs the action. At some point, the action becomes overlearned meaning that once an association is formed between that action and reward the action is persistently performed irrespective of whether or not the action continues being rewarding. Overlearned actions are habitual actions. Habitual actions are clearly intentional they are performed for a particular reason, they are however inefficient because they persist even when they are no longer rewarding.

Contingent actions, on the other hand, are both intentional and efficient and thus are indicative of desire. Desire is indicated by understanding that there is a contingent relationship between action and reward. An action is performed because it is rewarding. If an action ceases to be rewarding it must be stopped. Only specific actions namely goal-directed actions in virtue of being an instantiation of a contingent relationship between reward and action count as desire. If motivation is an inessential feature of desire as Schroeder claims the contingent relationship that is assumed to exist between reward and action and is thought to cause action cannot exist. Butlin shows this by exploring experimental psychology research about outcome devaluation and contingency degradation. Butlin believes if organisms can experience outcome devaluation and contingency degradation it will be illustrative of the fact that the organism can learn about and differentiate between reward and rewarding actions.

Outcome devaluation is the process through which outcomes are devalued. Outcome devaluation research is undertaken with the intention of determining whether or not organisms perform actions habitually or whether they perform actions using the goal-directed system. To devalue an outcome one merely needs to train a group of organisms until they form an association between an action and a reward, learn that lever pressing results in juice. Once the association is formed the group is split into two. One group will have the outcome devalued by

being given a substance that induces illness such that lever pressing and the ingestion of juice will be associated with illness. The illness-inducing substance is given before and after the consumption of juice so as to strengthen the association between juice and illness. The second group is also given the illness-inducing substance but at intervals intended to weaken the association between juice ingestion and illness. If the group that has had the outcome devalued stops pressing the lever or presses it less frequently than the group for whom the outcome has not been devalued, the organisms in question will be able to select actions on the basis of goals as oppose to mere associations: "If the action is goal-directed, performance of the action trained with the now devalued outcome should be reduced in this test relative to a control condition in which the outcome is not devalued." (Klossek *et al*, 2008:40). If an organism can experience a devaluation of outcome it implies that the organism can distinguish between a rewarding and non-rewarding action.

Contingency degradation refers to a process that is similar to outcome devaluation. The difference between these two is that in contingency degradation the experimenter manipulates the relationship between the performance of an action and the reward that is received as opposed to devaluing the outcome:

The degraded contingency effect, originally reported by Rescorla (1966,1968), is observed when unsignaled presentations of an unconditioned stimulus (US) are administered during training sessions in which a conditioned stimulus (CS) and a US are paired. These added US presentations diminish responding to the CS, independent of baseline responding in the context (Witnauer *et al,* 2007:440).

What is interesting is that contiguity does not affect contingency degradation. Without a causal relationship between an action and a reward and the manipulation of this relationship, contingency degradation cannot be achieved: "The critical finding of these experiments was that responding to the CS was reduced by a manipulation that did not reduce the contiguity between the CS and the US." (Witnauer *et al,* 2007:440). Contingency degradation implies the existence of a causal relationship between an action and the ensuing outcome and therefore ensures that an action's performance is contingent on its outcome. An action is performed if and only if it results in a reward. Outcome devaluation and contingency degradation do not support the habitual performance of an action. Outcome devaluation and contingency degradation support the performance of an action that has a causal connection to an outcome, an action that results in reward: "The subjective value of the outcome and the contingent relationship between the action and the outcome are thus both important facets of goal-directed actions. They can be assessed using outcome revaluation and contingency degradation tests, respectively, thereby determining whether the behavior is goal-directed or habitual." (Jackson *et al,* 2016:3274). An organism has to believe that the action that they engage in will result in the satiation of the desire they have and if it does not the action is abandoned for another: "We are able to perform an action in the belief that it will bring about a goal or a change in the world that fulfils our needs and desires." (Klossek *et al*, 2008:39). What do differences in what organisms learn tell us about desire?

Desire is understood in intentional terms. This means that when an action is performed, it is performed with the intention of meeting a particular outcome (in this context, attaining reward). This means that the organism intending to gain a reward or satisfy a desire must know what is rewarding, know the action that needs to be performed in order to attain the reward, and also believe that the action in question will result in the reward. These conditions are consistent with performing actions for reasons, which is what we take to be important for desire. Schroeder's account of desire fails to account for these conditions. Schroeder claims that action is inessential for desire, and thus runs afoul of the assumption that desire is geared to action. If we read Schroeder more sympathetically, and we assume that he understands desire as a disposition to action, dispositions to action can lead to habitual action. But habitual action is not what we have in mind when we speak about doing something for a reason. Butlin's account is preferable because it caters for action, viewing action as essential to desire, as well as viewing reasons for action as essential to desire. Butlin's account gives us an answer to why certain actions are taken, namely that they are taken because there is a belief that they will result in a reward.

I have illustrated the fact that the reward theory of desire can account for desire. Can the reward theory of desire account for the other phenomenon that we associate with desire? In order for the reward theory of desire to account for the other phenomenon that we associate with desire, there has to be neurological connectivity and therefore functional dependence between functionally specialised areas. The reward system, specifically, the three reward pathways, have to be connected to the dopaminergic system whilst also being connected to other parts of the brain.

The mesolimbic pathway receives dopamine from the ventral tegmental area (where dopaminergic releasing neurons found in the midbrain). The ventral tegmental area has projections that deliver dopamine to the striatum; specifically the ventral striatum (Wang *et al*, 2009).The striatum is one of the nuclei in the subcortical basal ganglia of the forebrain. The striatum has a dorsal and ventral component. The ventral striatum, part of the mesolimbic pathway, consists of the nucleus accumbens and the olfactory tubercle. The ventral tegmental area also secretes dopamine to the prefrontal cortex, thus creating the mesocortical pathway. The nigrostriatal pathway, like the mesolimbic pathway, projects to the striatum with the exception being that the nigrostriatal pathway projects to the dorsal portion of the striatum as opposed to ventral striatum (Ayano, 2016). The dorsal striatum is located in the subcortical region of the forebrain and is made up of the putamen and the caudate nucleus. The dorsal striatum receives dopamine from projections that emerge from the substantia nigra as opposed to the ventral tegmental area.

The following brain regions and parts are associated with the production of pleasure: the nucleus accumbens, the ventral pallidum, and the brainstem (Kringelbach *et al.*, 2010:8). There is a neuroanatomical connection between the abovementioned hedonic mechanisms and reward mechanisms. The nucleus accumbens forms part of the mesolimbic pathway. The ventral tegmental area has projections that terminate at the nucleus accumbens. This implies that hedonic signals that are produced by the nucleus accumbens are given incentive salience by

the dopamine-secreting neurons of the ventral tegmental area, thus resulting in organisms being motivated by pleasure.

The limbic system is also implicated in the production of pleasure. The part of the limbic system most associated with pleasure is the cingulate cortex, specifically, the perigenual anterior cingulate cortex: "…the neural basis of hedonic tone has strongly implicated a structure on the inner surface of each hemisphere in the frontal lobe, in a region known as the *perigenual anterior cingulate cortex* or PGAC." (Schroeder, 2004:36). The anterior cingulate cortex is also involved in reward processing, specifically, reward-based learning: "In primates, the more dorsal aspects of ACC, particularly those regions lying in and around the cingulate sulcus (primarily area 24 but excluding perigenual area 32), have been implicated in action-outcome learning." (Jackson *et al*, 2016:3279). The function of the anterior cingulate cortex facilitates contingency based and habitual learning. This is due to the fact that the anterior cingulate cortex is responsible for detecting, monitoring, and evaluating errors pertaining to reward. So far we have established the fact that the pleasure centres of the brain are not functionally independent from the parts of the brain that produce desire, but what about action?

When dealing with motivation, it is important to note that there is a difference between the execution of an action and the selection of an action. While the motor cortex executes actions, it is the motor areas of the basal ganglia which are responsible for action selection: "However, the selection of an actual action from the range of plausible actions is not performed by the motor cortex, but by another structure deep in the brain, known as the motor center of the *basal ganglia*. The basal ganglia guide both conscious, pre-planned action and spontaneous action." (Schroeder, 2004:37). If we accept Butlin's account, it stands to reason that desire ought to be associated with action selection and as a result, we have to speak about the basal ganglia.

The basal ganglia are made up of a group of subcortical nuclei, specifically: the striatum both ventral (nucleus accumbens and olfactory tubercle) and dorsal (caudate nucleus and putamen); the globus pallidus; the ventral pallidum; the substantia nigra; and the subthalamic nucleus. I will look briefly at the dorsal striatum, more specifically, the putamen. The putamen influences many types of motor behavior, especially those that are assumed to be under the purview of the goal-directed system. The putamen is thought to be responsible for motor planning, learning and execution, motor preparation, and the selection of movement. The putamen receives dopamine from projections that emerge from the substantia nigra. The putamen and substantia nigra together form the nigrostriatal pathway. This type of connectivity is indicative of the fact that the motor system (specifically, the motor part of the basal ganglia which are responsible for goal-directed action selection and planning) are connected to the reward system. This means that stimuli that are imbued with incentive salience result in motivation.

In this chapter, I illustrated that desire ought to be understood in terms of reward and therefore the function of the reward system. I illustrated that this understanding is preferable because it places no *a priori* limitations on what could be classified as a desire, other than the fact that it must result in the release of dopamine. I compared and contrasted the two main versions of desire reward theory and endorsed one. I endorsed Butlin's account because it accounts for reason sensitive action (action that tracks reward) as opposed to habitual action. I also

illustrated that the reward system has the requisite connectivity to account for the phenomena we associate with desires, namely pleasure and motivation. The reward system is connected to the pleasure centres of the brain which means we could account for desire in terms of pleasure. The reward system is also connected to the motor centres therefore resulting in us being able to account for desire in motivational terms. In totality, the reward base theory of desire presents a better account of desire.

Chapter Three

Simple and Complex desires

In this chapter, I will defend the claim that there are two types of desires, complex and simple desires. Complex desires are desires that have a belief as their proper determinants, whereas simple desires have physiological need as their proper determinants. I take desires to be motivational in nature, and by this I mean that desires cause an action (or, more appropriately, they play a role in the process of goal setting and thus action selection). In other words, an organism will have a desire, and the desire specifies what action ought to be taken as a condition of that desire's satisfaction. Complex desires influence action selection by being responsive to belief, which is, in turn, responsive to evidence. If one desires vitamin C because one believes that vitamin C will prevent one from getting the flu, the desire for vitamin C is a complex desire because it emanates from a belief about vitamin C. If the belief about the curative properties of vitamin C were proven to be false, the desire for vitamin C ought to disappear. If, on the other hand, it is proven that vitamin C *does* have curative properties and one believes that the acquisition and ingestion of vitamin C will prevent the flu and one wants to prevent the flu, one will develop a desire for vitamin C in virtue of one's belief about its curative properties. A

person with a desire for vitamin C will form an intention to act or alternatively engage in action(s) intended to acquire vitamin C, so long as they believe that the action will result in vitamin C, or alternatively is the most efficient way of acquiring vitamin C.

Simple desires, on the other hand, are desires that are responsive to the physiological states of an organism. An organism has a simple desire when a resource it requires is nearing depletion. Simple desires can, therefore, be understood as states that emerge as a consequence of an organism moving away from homeostasis. Simple desires are triggered as a way to make an organism attend to imbalances in its internal environment. Let us take hunger as an example. An organism becomes hungry when its metabolic energy reserves reach a particular point. At that point, the organism's body releases certain chemicals and these induce hunger; consequently the organism acquires a simple desire.

This chapter is intended to argue against what is viewed as a logical consequence of accepting the view that was argued for in the previous chapter: the view that physiological need does not result in goal-directed action selection and therefore desire. Some philosophers, among them Butlin (2017) and Sterelny (2003), are of the opinion that organisms attend to their physiological needs in a manner that is not indicative of goal-directed action selection. These philosophers are of the opinion that when an organism is hungry, it acts in an habitual manner, a manner that usually results in satisfaction when in the current state. If you accept this position, as well as the position that was established in the previous chapter, you are forced to conclude that physiological need does not lead to desire. This chapter makes the argument that this is not the case. I am of the opinion that under certain circumstances physiological need results in desire.

In order to defend the claim that there are two types of desires, I will focus my attention on two arguments. The first argument comes from Sterelny. Sterelny accepts the existence of simple desires, or what he would call "internal drives." Sterelny defines internal drives as the hierarchal arrangement of motivations: "many organisms simply have a built-in motivational hierarchy" (Sterelny, 2003:81). However, he denies the existence of what I call complex desires. Sterelny's denial is predicated on the assumption that complex desires would have to be like beliefs. Complex desires would be like beliefs, according to Sterelny, if they were states that set goals, goals that could be realised by multiple actions. Sterelny believes that desires are *unlike* beliefs: desires and beliefs evolved to execute the same function (tracking) but in different environments. Function is determined by selection pressures and selection pressures are determined by the environment. If desires and beliefs evolved in different environments, they will have different evolutionary histories and therefore will instantiate the same function (tracking) in different context-dependent ways. Sterelny believes that beliefs and desires did, in fact, develop in different environments and thus under different selection pressures, and are therefore different. I do not dispute the fact that desires and beliefs have different evolutionary histories. I also do not dispute the fact that desires and beliefs emerged differently. However, I dispute the claim that they track reward differently and consequently that an account developed for beliefs cannot be generalised to desire. I will illustrate that the account developed for beliefs does, in fact, generalise to desire.

After dealing with Sterelny I will move on to an argument presented by Butlin. Unlike Sterelny, Butlin accepts the existence of complex desires but denies the existence of simple desires. Butlin states that a desire is a desire in virtue of the role it plays in the action selection process. Desires, according to Butlin, are supposed to specify the goals of an agent without specifying the action that ought to be taken to reach that goal. The action selected to satisfy a desire is determined by the belief that the action is going to result in the satisfaction of the desire. Butlin wants to convey the idea that an organism has to calculate which action of the available actions is going to result in the satiation of the desire *ceteris paribus*. Butlin believes that what I call simple desires are not in fact desires. Butlin holds this view because he believes that while physiological need motivates (i.e., results in action selection), it selects for action in a manner different from complex desire. Physiological need motivates in a manner similar, if not identical, to overlearning or habit. The idea that Butlin is trying to convey is that when in a state of physiological depletion, an organism performs an action that resulted in the alleviation of the current state at some previous time. This is not tantamount to desiring for Butlin, but rather acting out of habit. Both habitual action and desire are intentional in nature: they are both about reward. However, they differ in that, where desire is concerned, the performance of an action turns on the belief that the action it is best-suited to bring about reward, and the belief that the action is best-suited to bring about the reward turns on whether or not the action results in reward. The notion of being best-suited to bring about reward requires the abandonment of the action in question should it not lead to reward. Butlin is of the view that where physiological need is concerned, organisms continually engage in behaviours that do not result in reward if those behaviours were once rewarding. I disagree with Butlin on this point. Our bone of contention is not the claim that simple desires emerge differently; nor is it that organisms tend to reproduce actions that were once rewarding. Our bone of contention is the manner in which simple desires, drives, result in action. Butlin is of the view that drives result in habitual action, whereas I am of the view that they can result in the selection and therefore the performance of varied actions. I hold this view because I believe that desire satisfaction is determined by the structure of the environment. The environment that results in the creation of a desire may be simple (and therefore is without complexity), but the environment in which an organism acts to satisfy the desire could be complex, thus leading to complex non-habitual behaviour.

Before I give an account of Sterelny's argument it is important that I explain some terminological differences. Sterelny does not refer to desires: he refers to preferences. I believe that desires and preferences are equivalent. Consequently, I feel that what holds for preferences holds for desires. I will give an argument for why I think these two terms are equivalent, but for the moment it will suffice to make the reader aware of my assumption. Secondly, what I call simple desires Butlin and Sterelny refer to as drives. With these issues out of the way, I can begin in earnest. I will begin by looking at belief. When do we attribute belief to an organism?

We attribute belief to an organism when the organism has the ability to track and respond to its environment. While tracking and responding to the environment is necessary for belief attribution, it is ultimately insufficient according to Sterelny. For an organism to have beliefs it must also have the ability to respond to the environmental features it tracks. While

responsiveness to the environment is necessary, it too is ultimately insufficient. To be classified as having belief, an organism has to exhibit behaviour that is associated with outcome devaluation and contingency degradation: behaviour that is goal-directed but variable. Variable response to a particular representation of the world is indicative of the existence of decoupled representation(s), according to Sterelny. A decoupled representation is a representation of the environment which, when used as a factor to determine behaviour, can result in the initiation of various behaviours. The existence of decoupled representations or a mechanism whose proper function it is to create such representations results in belief. An organism can, therefore, be said to have beliefs if it has decoupled representations: "In speaking of an agent's beliefs, we are attributing decoupled representations, representations that can influence many types of action." (Sterelny, 2003:31). How do decoupled representations emerge?

Decoupled representations emerge as a consequence of environmental complexity. Sterelny argues that it would be maladaptive for an organism inhabiting a complex environment to link a representation of the world to a specific action. Environmental complexity results in the selection of complex behaviour because a simple or stereotypic response – always assuming that representation ***A*** means state of the world ***B*** and therefore the instantiation of behaviour ***C*** – would be insufficient to deal with an environment where representation ***A*** means state of the world ***B***, but implies the initiation of behaviour ***F***: "Thus environmental heterogeneity selects for variable response…" (Sterelny, 2003:11). What causes environmental complexity?

Environmental complexity is determined by the transparency or opacity of the environment that an organism inhabits. Environmental complexity is determined by the clarity of the information an organism gets from the environment it inhabits. Transparent environments are simple environments defined by unambiguous signals. Organisms that inhabit opaque environments, environments with ambiguous signals, are said to inhabit complex environments. Degrees of transparency and opacity determine response depth, and thus the formation of decoupled representations. Organisms that inhabit simple, transparent, environments do not have response depth and therefore decoupled representations. This is due to the fact that organisms that inhabit simple environments receive information that is transparent and veridical and this prompts the emergence of habitual behaviour:

> If the signal indicating the presence of a specific resource is reliable, and if the agent can use its sensory apparatus to discriminate that signal from other stimuli, then cue-driven behaviour will succeed. For with respect to that feature, the animal lives in an *informationally transparent environment*. Detection systems reliably drive adaptive behaviour in transparent environments. (Sterelny, 2003:20-21).

Consider an intersection with a traffic light as an example of a transparent environment. The intersection is an example of a transparent environment because the behaviour of every road user is, or ought to be, determined by an unambiguous cue. People behave stereotypically or habitually when presented with certain traffic cues; as such, one can predict the behaviour of a

road user if one knows the signal that will be projected to them. Informational transparency leads to automatic and habitual responsiveness, as is the case in traffic intersections.

Organisms that inhabit opaque environments, environments that are not defined by reliable information, cannot afford to response habitually. Organisms that inhabit opaque environments must have response depth, complex behaviour: "But cue-driven organisms will often struggle if ecologically relevant features of their environment – their functional world – map in complex, one to many ways onto the cues they can detect. Such organisms live in *informationally translucent environments*." (Sterelny, 2003:21). According to Sterelny only organisms that inhabit complex environments, informationally opaque environments, develop response depth and thus decoupled representations. Sterelny further says that only organisms with decoupled representations can be thought of as having beliefs or belief-like states:

> Intentional agents have *decoupled representations*. That is to say, we have internal states that track aspects of our world, but which do not have the function of controlling particular behaviors. Beliefs are representations that are relevant to many behaviors, but which do not have the biological function of directing any specific behaviour. If decoupled tracking evolves, an agent's behavior will become less stereotyped across the agent's whole behavioral repertoire. (Sterelny, 2003:29).

Let us again look at the intersection as an example. If the traffic lights are not working, the intersection turns into a four-way stop street. The rules of a four-stop street state that whoever gets to the stop sign first (or in this case, the malfunctioning traffic light) departs first. In as much as there are rules that determine how people *ought* to behave at a four-way stop, experienced drivers will tell you that you have to tailor your responses to the response of other drivers. You could get to the stop sign first and, by law, you ought to be given way; but the second person to get to the stop sign may decide not to follow the rules. In this case, if you had not yielded to assess the intentions of the other drivers you would have been in an accident. The manner in which people conduct themselves in four-way intersections is indicative of decoupled representations.

Before I give a defence of my view, another concept needs clarification. I have spoken about beliefs and belief-like states, agreeing with Sterelny when he says that they evolved as a consequence of environmental complexity. I have not yet spoken about how environmental complexity emerges. Sterelny states that the hostility of an environment results in complexity. Hostility results in complexity by changing the informational character of the environment: "Thus hostility changes the informational character of local environments, degrading the covariation between easily discriminated cues and the functional properties they signal." (Sterelny, 2003:25). When Sterelny talks about hostility he is referring to biological hostility: the existence of predators and thus the possibility of predation. Think of camouflage as an example of informational opacity. When an organism is aware that its prey can disguise itself, the manner in which it hunts (and *vice versa*) will have to change given the change in how the environment is interpreted.

Let me now return to an issue I broached earlier. Sterelny speaks about preferences and I speak about desires. I assume desires and preferences are identical. I assume that one has a preference for something that one desires *ceteris paribus*. John Elster (1983) speaks about preferences which might not be indicative of desire when he speaks about a form of preference reversal called the sour grapes phenomenon. If someone is experiencing sour grapes they desire something which they cannot acquire and as a result of being unable to acquire the object in question, prefer something less desired. Consider Aesop's the Fox and Grapes as an example. In this fable a fox wants, desires, grapes but he cannot get them and therefore claims that the grapes are sour and moves on to something less desired but available. Outside of sour grape cases, desire and preference are equivalent and as a result whatever Sterelny claims about preferences holds for desires.

As previously stated, Sterelny believes that an organism can only be said to possess beliefs if it has decoupled representations. Sterelny likens beliefs to preferences: both are intentional psychological states, and as such the criteria that determine belief attribution ought to apply to preferences. Sterelny states that in order for an organism to have preferences it would have to have decoupled representations. This is to say that preferences have to exist in an environment that is defined by informational opacity / complexity. Given the fact that complexity is a byproduct of biological hostility, the preference generating mechanism has to have evolved in an environment where its function is impinged on by the possibility of predation for it to be like belief.

Sterelny believes that the criteria that must be met in order to attribute belief to an organism cannot be met where preferences are concerned. Sterelny asserts that beliefs evolved to track an organism's external environment whilst preferences evolved to track an organism's internal environment. The external environment is defined by the existence of predators. The presence of predators leads to the evolution of decoupled representations. The internal environment is defined by the absence of predators hence the absence of decoupled representations. The presence of, and need to respond to, the complexity that is induced by predators in the external environment means that whatever applies to the external environment cannot be generalized to the internal environment: "There is a plausible evolutionary scenario that explains the evolution of decoupled representations. But that story cannot be generalized to explain the transformation of motivation: the change from motivation-based on internal signals of the physiological condition into motivation-based on representations of the external world." (Sterelny, 2003:73). In as much as Sterelny believes that we cannot move from a generalization about the external environment to a generalization about the internal world, he still needs to account for the production of behaviour.

Sterelny acknowledges that both beliefs and desires result in action and as a consequence to deny desires would be to deny that organisms engage in intentional, reward centric behaviour. Sterelny, therefore, makes a move similar to that made by Butlin. Sterelny claims that not all detection of and response to environmental features is indicative of belief. Sterelny claims that there are organisms that detect and react to certain features of the environment without having beliefs, or decoupled representations. Given this, Sterelny believes that it is also possible to act without preferences: "…just as it is possible to act without having beliefs about the external

world, so too it is possible to act without having preferences." (Sterelny, 2003:79). Where preferences are not present, behaviour can be attributed to internal drives. Behaviour motivated by internal drives is behaviour that is produced by responsiveness to physiological cues.

Internal drives produce behaviour by linking a specific representation with a specific behaviour such that the representation triggers behaviour without deliberation. The mere fact that there is an internal drive is a reliable indicator of the fact that its proper determinant is in play and thus behaviour will be initiated to satisfy the drive. George Ainslie also believes that behaviour need not be motivated by preference or deliberation: "Thus 'motivated' doesn't mean only deliberate…" (Ainslie, 2001:60).

Sterelny, Ainslie and Butlin are conveying the idea that desires are motivational; they result in action. However, not all action is indicative of desire. Sterelny argues that behaviour that emerges as a consequence of physiological cues (internal drives) is without deliberation and thus, not a candidate for desire. Sterelny argues for this conclusion based on assumptions he makes about the environment that preferences track. According to Sterelny, beliefs track the external environment. The external environment is defined by the existence of competing interests and this results in hostility, hostility in turn results in complex and deliberative behaviour. Sterelny believes that the internal environment is fundamentally different from the external environment. The internal environment is defined by a lack of hostility and hence informational transparency which results in habitual or non-deliberative behaviour. Habitual behaviour emerges as a consequence of what desires track.

What do desires track? Perhaps the better question to ask, given that Sterelny assumes that complex desires do not exist, is what would complex desires track if they existed? According to Sterelny, complex desires (if they existed) would track an organism's internal environment much like internal drives. How is the internal environment tracked? Is the internal environment hostile; does it require decoupled representation? The internal environment is without hostility according to Sterelny. Consequently, the internal environment tends to transparency as opposed to opacity: "Internal environments are transparent." (Sterelny, 2003:81). This means that the mechanisms that produce internal signals, signals that indicate the physiological state of an organism, are reliable and consistent. As a consequence of the transparency of these signals, mechanisms that have them as their proper determinants induce an internal drive (i.e., they report on an organism's physiological state) upon the detection of these signals without interrogating or deliberating on them. This subsequently results in habitual action. Sterelny claims that organisms that inhabit transparent environments, environments where information is unambiguous, have behaviour that is habitual and governed by simple cues or internal drives.

The claim that the internal environment is without ambiguity is not entirely true. If it were true, simple desires (internal drives) would only emerge as a response to their proper determinants. Let us consider hunger as an example. Sterelny considers hunger to be an internal drive and as such hunger is or ought to be experienced when an organism's metabolic energy levels decrease. Metabolic energy levels are controlled by two hormones: ghrelin and leptin. The presence of ghrelin in the blood induces hunger and therefore a desire for food, whilst the presence of leptin decreases hunger and therefore induces satiation. Ghrelin is not arbitrarily

released by the body. Ghrelin is released in response to a decrease in metabolic energy and as such, it ought to be the case that when metabolic energy is not low, the desire for food should not be felt. However, Schussler *et al.* (2012) have illustrated that this is not the case. Schussler *et al.* have illustrated that people who are satiated, people who have a low concentration of the neuropeptide ghrelin in their blood, nonetheless experience a desire for food when presented with images of food. Low levels of ghrelin are, or ought to be, an indication of satiation. The production of ghrelin ought to be triggered by a highly regimented biological process and yet, showing people images of food results in an increase in ghrelin and therefore desire for food without physiological need. If the internal environment is as unambiguous as Sterelny claims, the effect discovered by Schussler *et al.* should not be possible: evidence, however, shows that it is.

Schussler *et al.*'s research is very interesting because it can factor into issues pertaining to action selection and action selection is thought by both Sterelny and Butlin to be important when it comes to the attribution of desire. The assumption is that internal drives are not desires because of the manner in which they result in action selection. The assumption is that when one has an internal drive, the drive leads unthinkingly into action: an action that was previously rewarding under the same circumstances. This underlies the claims that internal drives are not desires. It does not, however, strike me as reasonable to assume that internal drives always lead to habitual action.

When dealing with an organism as complex as a human being, an organism that is aware that its physiological indicators may be wrong (as illustrated by Schussler *et al.*), it is likely that beliefs about satisfaction could outweigh physiological indicators. Having just eaten and feeling satisfied, it is unlikely that when presented with pictures of food which result in the activation of ghrelin, that the person in question is likely to be driven to consume more food. This person is likely to be aware that it is the pictures of food and not genuine hunger that is leading to a desire for food, and as a consequence, they are unlikely to act on that desire. This person is using beliefs about his physiological condition as a means to forestall the information he or she is receiving about his or her physiological condition. This illustrates that beliefs about one's physiological condition can override physiological reports about departures from homeostasis. This, in turn, illustrates that it is not always going to be the case that internal drives lead to unthinking and habitual to action.

We can, in the face of empirical evidence to the contrary, accept Sterelny's claim about the unambiguous nature of the internal environment. We can assert that when one feels thirst or hunger, those sensations are reliable indicators of the fact that an organism is in a state of depletion and requires replenishment: "Drinking in response to sensations of thirst, and resting in response to feelings of lassitude, are pretty good fast and frugal heuristics." (Sterelny, 2003:80). Accepting Sterelny's point does not, however, rescue his argument. Accepting the fact that internal drives are unambiguous does not entail that internal drives lead to habitual behaviour. We can accept the unambiguous nature of physiological cues and thus their reliability *ceteris paribus*. I, however, do not agree with the assertion that the reliability of physiological cues means that internal drives are not desires or simple desires.

In order to make this argument, I have to dispute biological hostility as a necessary condition for the existence of decoupled representations as argued by Sterelny. I am going to make the argument that this focus on biological hostility misses the point. Biological hostility broadly construed is nothing more than the existence of complexity and therefore competition. In as much as internal drives are reliable, they still need to compete for selection and this competition, in my opinion, does for desires what the existence of predators does for beliefs: namely, it creates complexity.

Consider what would happen if an organism has more than one internal drive. Sterelny states this situation is easy to sort out. Sterelny states that there is a motivational hierarchy and this hierarchy specifies which internal drive will be attended to first and therefore which actions will be taken. The hierarchical nature of the internal environment means that there is a lack of competition and therefore a lack of conflict amongst internal drives. Sterelny holds this view because he is under the impression that the interests of internal drives coincide with the interest of an organism, its fitness, and this is best catered for by coordination and cooperation as opposed to competition. Sterelny's assertion seems to make sense at first, but upon closer inspection, it does not hold water.

If it can be illustrated that there are environments where one would expect coordination and cooperation, yet we sometimes see competition that would illustrate the fact that it is, in principle, possible for internal drives to be in competition. Consider intragenomic conflict. Intragenomic conflict is a term that describes genes that function for their own good and to the detriment of other genes that reside in the same genome (Gardner *et al.* 2017). One would expect that an organism's genome would be comprised of genes that cooperated and coordinated their function so as to build an organism that is fit, or able to propagate the various genes that comprise it; this is, however, not always the case. Intragenomic conflict occurs when the genes that comprise a genome are in conflict with each other. Intragenomic conflict results in some genes selfishly promoting their own replication and transmission over and above the replication and transmission of the other genes that make up the genome, therefore, having a detrimental effect on the other genes that make up the genome and ultimately the organism. If a conflict can occur at the level of genes and therefore the genome, it is in principle possible for it to occur at the level of internal drives and therefore the internal environment. The possibility of intragenomic conflict creates a context for the existence of the type of competition – and therefore opacity – that results in decoupled representations in an environment that is not defined by biological hostility. I think that this can apply to internal environments as well. In principle, there is no reason why it should be the case that there is no competition in the internal environment, and therefore no possibility of the competition and complexity we see in the external environment. Sterelny is wrong when he asserts, implies, that biological hostility is the only thing that can result in opacity and therefore the creation of robust tracking mechanisms like decoupled representations.

Given the above arguments, I think we ought to accept that desires – specifically, complex desires – do exist. If we accept that complex desires do exist, half the battle is won. I now have to consider whether things like hunger are desires. I believe that hunger is a desire. I believe that hunger is a very specific type of desire: a simple desire. I already illustrated that complex

desires exist. The existence of desires was illustrated by showing that the criteria that we use to establish belief apply equally well to desires. I now argue for the acceptance of simple desires. To begin this task let me look at Butlin's work.

Like myself, Butlin believes that desires do exist. We differ, however, over things like hunger. Butlin believes that hunger is not a desire whereas I believe that it is (albeit a simple desire). I define simple desires as desires that emerge as a consequence of the physiological needs of an organism. Butlin defines hunger and related things, like thirst, as basic drives as opposed to desires: "based on evidence from psychology and neuroscience, according to which hunger, thirst, and other *basic drives* (as I call them) are not desires." (Butlin, 2017:617). However, we both accept that thirst and hunger are responses to physiological need and that when in the appropriate physiological state, one is motivated to act, and one acts in virtue of that physiological state: "Hunger is an occurrent state that we enter when our physiological need for food is relatively high, that affects action selection in the moment." (Butlin, 2017:626). We, therefore, both agree that physiological states are motivational in nature.

Butlin's beliefs about hunger are predicated on how hunger motivates action. Butlin is of the opinion that motivation or action selection is underpinned by two different systems, the habit and goal-directed system. The habit system according to Butlin prompts or promotes a form of motivation which is akin to stimulus-response. What Butlin is trying to convey is the fact that organisms learn to associate a certain stimulus with a certain response and that response with a certain reward. Therefore, when in that state, experiencing the stimulus, the organism seeks to initiate an action, response, that resulted in a reward under the same condition previously. The response is initiated on the assumption that it previously resulted in a reward and thus ought to do the same again under the current context. The goal-directed system is however different from the habit system. Like the habit system, the goal-directed system keeps track of rewarding stimuli. The goal-directed system goes a step further than the habit system in that it not only keeps track of what actions are rewarding, but it also keeps track of whether or not actions really cause or result in a reward:

> The goal-directed system, in contrast, keeps track of two relationships: the causal relationships between actions and outcomes, and the reward values associated with outcomes. It combines states representing these relationships to calculate the expected reward values of salient actions and causes the actions with the highest expected values. (Butlin, 2017:619)

Butlin believes that actions motivated or selected in virtue of the habit system should not be considered as desires. Butlin believes this because he believes that there could be a correlation between an action and a reward but the reward is not in any way caused by the accompanying action. According to Butlin, only actions motivated by the goal-directed system are desires. Butlin believes this to be the case because he defines desires as states that emerge when something acts as input for the goal-directed system: "For a state to be a desire is for it to be a member of the natural kind with the function of acting as inputs to the goal-directed system,

which track the reward values of outcomes." (Butlin, 2017:620). Butlin is trying to illustrate the fact that in order for an organism to have desires, it must not only be motivated, but must be motivated by considerations pertaining to which action is rewarding, that is to say, to which action causes or results in a reward.

This argument seems like the argument put forth by Sterelny. We can assume that when Butlin talks about the habit system he is talking about a system that produces formulaic responses to certain stimuli dependent on the reward that is associated with those stimuli. Given this, we can also assume that Butlin, like Sterelny, also assumes consistency. We can make this assumption on the basis of the fact that a habitual response can only make sense under the assumption that a particular stimulus is a reliable indicator of a certain reward. However, as previously illustrated, it does not follow from the existence of reliable stimuli and reward pairing that a formulaic response follows. Let us consider thirst for a moment. When an organism is thirsty it does not follow that it will drink. A thirsty organism can instead opt to consume food: "A dehydrated animal can drink, or it can eat since almost all food contains some water. And so forth." (Spurrett, 2015:310). To return to hunger, hunger may be a physiological condition that emerges as a consequence of specific physiological cue and thus can be considered to be a reliable indicator of an organism's internal condition. But how hunger is responded to might not be cue-or habit-driven: it could be exceptionally complex and this complexity is determined by the environment in which the desire has to be satisfied.

Consider David Spurrett's argument. Spurrett makes the claim that an organism may feel hunger and thus the need to respond to it, but how it responds can be very complicated: "A hungry animal might have more than one foraging option. Accurate information about needs does not always, then, suggests a unique 'good enough' response which would favour cue-bound control." (Spurrett, 2015:310). Consider the following example as an illustration. When I am on campus and I feel hungry, my feeling of hunger only specifics that I need to eat given my low metabolic energy reserves. Whether I eat (I could be fasting and therefore not eat), what I eat, and where I go to eat, are complexities that are not specified by my feeling hungry. My hunger may result in my acting; it may not. Even in cases where my hunger results in my acting, my actions are underdetermined. If I decide to eat, I could go to any one of many food stalls around campus – a move that would be prompted by my belief that there are food stalls and that these stalls are places where I can get food. I could also go back to my flat to have lunch – a move that could be prompted by the belief that while there are food stalls, they do not adequately cater for my dietary needs. Having a simple desire does not in any way imply the existence of habitual responses: it merely implies that some action needs to be taken, the nature of which is subject to the environment the organism finds itself in and the beliefs the organism has.

In this chapter, I have illustrated that there are two types of desires: simple and complex desires.

Chapter four

Rationality of belief and desire

In this chapter, I will discuss rationality: I will define what rationality is, and whether or not the concept of rationality can be extended to both belief and desire. I will show that both beliefs and desires are rational. I will show that beliefs and desires are rational if we understand rationality to mean responsiveness to evidence. An evidence responsive account of rationality states that, to be rational, an organism needs to respond in a way that is occasioned by the information or evidence it possesses. Where beliefs are concerned an organism has rational beliefs if the beliefs it has would change if the evidence in support of those beliefs changed. A rational organism responds appropriately (say, by running away) to the presence of a predator nearby (sights, sounds, and smells as contributing evidence). It would be irrational, in most cases, to run towards the predator where appropriate evidence exists. Where belief is concerned, it would be considered rational to believe that a predator is nearby if presented with evidence; consequently that rational belief would change should the evidence (sights, sounds, and smells) change / disappear.

Responsiveness to reason (as will be later explained) should not be equated with correctness or infallibility. To make the claim that beliefs are rational is not to claim that beliefs or the mechanism responsible for their creation is without error, as such there will be cases where an organism has a false belief but the belief is still rational. A false belief is rational if and only if it is occasioned by evidence, and as such, the acquisition of new evidence can correct the old belief if it happens to be inconsistent with the facts as we know them.

Desires, specifically complex desires, have beliefs as their proper determinants. The fact that complex desires have beliefs as their proper determinants implies that complex desires are *rational* if and only if they are occasioned by, and are thus responsive to, beliefs. The relationship that holds between complex desires and beliefs can be cashed out in different ways. False beliefs can generate rational desires. When we look at the rationality of desire, we look at whether or not a desire tracks a belief and whether that belief is justified given some evidence and not whether the evidence is good and therefore whether the belief is true or false. As such

false beliefs can result in rational desires. For example, the organism (say, a mouse) might mistake a stuffed animal for a cat, and run away. The belief that there is a predator nearby is false, and yet it has created the rational desire to run away.

There are cases where the rationality of a desire piggybacks on the rationality of a belief. In these cases, the desire is not merely rational because it follows or tracks any belief, but rather is rational because it follows a rational belief. The second example of the rationality of desire is more substantive in nature in that rational desires emerge from rational beliefs. The first account of the rationality of desire is more procedural in that it emphasizes how desires come about. The distinction I make is echoed by philosophers like Jon Elster and John Christman. John Elster distinguishes between two types of rationality: thin, and broad rationality, and this corresponds with procedural and substantive rationality. Elster defines thin rationality as consistency: "It is thin in that it leaves unexamined the beliefs and desires that form the reasons for the actions whose rationality we are assessing, with the exception that they are stipulated not to be logically inconsistent." (Elster,1983:1). Where thin rationality only specifies a formal criterion for rationality namely consistency, broad rationality specifies a more substantive criterion:

> The broad theory of individual rationality goes beyond these formal requirements. Rationality here involves more than acting consistently on consistent beliefs and desires; we also require that the beliefs and desires be rational in a more substantive sense. It is not too difficult to spell out what this means in the case of beliefs. Substantively rational beliefs are those which are grounded in the available evidence: they are closely linked to the notion of judgement. (Elster, 1983:1).

John Christman confirms the dichotomy established by Elster. Christman distinguishes between internalist and externalist notions of rationality. An internalist account of rationality is similar to thin rationality in that it specifies consistency: "Some views are 'internalist' like the claim that individuals are rational if they have consistent beliefs and desires." (Christman,1991:13). An externalist account of rationality is similar to a broad account of rationality in that both emphasise evidence (substance): "Other requirements can be considered 'externalist' and are thus more stringent: for example, the requirement that the beliefs upon which an agent's conditional desires rest are based on an (objectively speaking) adequate degree of evidence." (Christman,1991:14).

Whilst most people would argue for a more substantive account of rationality, I prefer a more procedural account of rationality. Most people would prefer a substantive account of rationality because they assume (or rather, would prefer) rationality to be equivalent to truth such that a rational proposition is equivalent to a true proposition. However, this is not necessary. A substantive account of rationality is unnecessary because a procedural account of rationality, especially if it is responsive to reason or evidence, would ultimately result in a substantive account of rationality. People may wonder why it is that I am advocating for a procedural account if it ultimately leads to a substantive account; why not just aim for a substantive

account, to begin with? A substantive account of rationality would require us to have true beliefs to begin with, and this is too tall an order given that our cognitive capacities are fallible: "According to fallibilism human beings are doomed to error. Not even the most secure, the best supported belief is free from the possibility of mistake." (Kekes, 1972:301). The idea of fallibilism and procedural rationality is reinforced by the concept of bounded rationality. Limited rationality implies rationality under constraints or limitations: "It started from the proposition that all intendedly rational behaviour is behaviour within constraints." (March, 1986:146). To be rational in a limited sense does not mean irrationality: "…human beings develop decision procedures that are sensible, given the constraints, even though they might not be sensible if the constraints were removed." (March, 1986:146). Rationality is limited by various factors, factors like time, information and cognitive resources: "The limitations were limitations of computational capability, the organization and utilization of memory, and the like." (March, 1986:146). Given these limitations, a procedural account is preferable especially if it is reason responsive.

Simple desires are rational if they are formed the right way; this is to say if they are formed as a response to departures from homeostasis. This means that simple desires are rational if they are responsive to their proper determinants. Schussler *et al.* have, however, illustrated that there are instances when simple desires are not created as a response to their proper determinants (recall the example where hunger is induced by pictures of food as opposed to levels of ghrelin in the gastrointestinal tract). Under these instances, instances where we have reason to assume that simple desires are not rational or not formed by their proper determinants, we test their rationality by looking at our beliefs: beliefs about their proper determinants and how they ought to function.

So far I have spoken about rationality as responsiveness to evidence in general terms. My version of reason responsiveness is different from the standard account. My version of responsiveness to evidence emphasizes responsiveness (procedure) as opposed to accurate response (substance). This forgoing statement, responsiveness to evidence emphasizes responsiveness (procedure) as opposed to accurate response (substance), needs clarification. Responsiveness to evidence and therefore rationality should not be understood in terms of objectively right or wrong responses. Given incomplete information, limited computational resources and a finite amount of time within which to compute beliefs, verisimilitude is not possible; nor is it necessary. An organism need only factor in evidence as and when it is available for it to be responsive to evidence and therefore rational. The thing of import is the role played by evidence in the formation of beliefs as opposed to the veracity of the beliefs that are produced. An organism is a rational believer if evidence plays a causal role in the formation of its beliefs. The mouse has no time to wait if it believes that the "cat" is close enough to pounce. At the time, the visual evidence of shape and texture is enough to form its belief, and it does so rationally.

There is an evolutionary rationale for having beliefs. Beliefs represent the environment an organism inhabits. An organism would not survive if it did not have representations of its habitat. Organisms that inhabit environments that change need to have beliefs that change. As a consequence of environmental change merely being able to represent one's environment is

insufficient. An organism inhabiting a dynamic environment requires beliefs that not only represent but also track the environment. To survive an organism must be able to update its beliefs when confronted with new evidence. It seems as if evolution requires responsiveness to evidence, especially in dynamic environments.

Understood in tracking terms it would seem as if evolution requires a version of responsiveness to evidence that is truth preserving. This seems obviously true in that the purpose of belief is to keep track of the environment so as to give the organism a true representation of its habitat. A true representation of the environment and therefore true beliefs seem, at least intuitively, to offer more survival value than false beliefs; however, this is not the case. The fact that having beliefs that are responsiveness to evidence increases an organism's chances of survival and therefore chances of reproduction does not imply that responsiveness to evidence ought to be understood in terms of objectively true beliefs. Early humans may have believed that dead bodies were cursed or evil and as such should be disposed of. This belief is false, but in terms of survival, limited contact with decomposing bodies is essential.

Evolution has as its primary objective increased fitness as opposed to truth preservation. True beliefs are important in so far as they aid survival and fitness. If survival and fitness can be increased by a system that has limits and therefore a system that occasionally churns out incorrect answers, that system has as good a chance of being selected as the accurate system. Consequently, a system that errs could be selected for in preference to a more accurate system if the error-prone system imposes lesser costs on the organism than the more accurate system.

If we understand fitness, direct fitness, as the number of offspring an organism produces and if we understand survival as a prerequisite for reproduction, we can conclude that an organism needs to survive in order to increase its probability of reproducing and thereby increasing its fitness. Consequently, whatever increases an organism's chance of survival also increases the organism's fitness *ceteris paribus*. As previously stated, an organism's chances of survival are linked to the accuracy of its beliefs. There is a good reason why this should be the case. If an organism believed that a predator was not present when one was actually present that organism would be killed.

While it is true that true beliefs have a positive impact on an organism's survival it does not follow that only true beliefs can play this role. More importantly, it does not follow that while true beliefs increase survival they are better, comparatively speaking when compared to false beliefs. The mere fact that accurate beliefs are linked to survival does not mean that a system that errs will automatically result in a maladaptive and therefore less fit or unfit organism (Stich, 1993). In an ideal world where organisms have perfect information (time to compute and computational resources large enough to handle the requisite computations), having true beliefs would be better. In a world where information is imperfect, time to compute as well computational resources are limited, an error-prone system can fare better than a system that produces true, error-free beliefs especially if the accurate system imposes massive costs on the organism.

Evolution is more concerned with the cost an error imposes on an organism as opposed to the mere existence of error and the plurality thereof. Evolution can select a system that is error-prone if the cost of error is lower than the cost incurred by the organism when it commits additional resources to producing veridical beliefs (Stich, 1993). Consider predator detection as an illustration. When an organism receives information that indicates that a predator is around, the information could be false or it could be true. Irrespective of the truth or falsity of the information an organism needs to decide how to act. An organism has three possible moves. The organism could initiate predator avoidance behaviour or the organism could seek more information. Seeking more information would be a move initiated in response to the ambiguity of the information at hand. More information would be sought to verify the information at hand so as to form and act on true beliefs. Finally, the organism could disregard the information. Disregarding the information at hand entails either assuming that the information is false or alternatively that the information is ambiguous and as a consequence of the ambiguity cannot be relied upon.

If the organism in question initiates predator avoidance behaviour and it happens that a predator is around, the organism survives. In this case, the information indicating the presence of a predator or information interpreted as indicating the presence of a predator truly does represent the presence of a predator and therefore results in a true and therefore rational belief and therefore rational action. If the organism happens to initiate predator avoidance behaviour and there is no predator around, if the organism was feeding at the time it got this information, the initiation of predator avoidance behaviour results in the organism forgoing its meal. The organism could discount, disregard, the information at hand because it is ambiguous. The ambiguity does not, however, imply that a predator is not around. If a predator is in fact around disregarding information on the basis of its ambiguity could be catastrophic.

The organism could, as a consequence of the ambiguous information, go in search of more and clearer information so as to disambiguate between threat and non-threat. If the organism in question engages in this type of behaviour and there indeed is a predator at large, the organism could be killed. The first example that I spoke about is trivial. If the information an organism has is true and the organism initiates avoidance behaviour in virtue of that information, the ensuring belief is not only true: it is also rational. Beyond being true, the belief and ensuing behaviour have a good payoff: survival. Survival does come at the expense of nutrition in that the organism forgoes a meal; however, a meal that is forgone can be replaced if one is alive to hunt or forage at some later stage.

The second and third examples are more interesting. An organism can initiate predator avoidance behaviour without a predator present. If this happens, the organism suffers a loss / incurs a cost. If it turns out that a predator is not present, the initiation of avoidance behaviour results in the organism losing out on food. This loss in food, an unnecessary cost imposed on the organism, might prompt some to think that it is worthwhile to verify the information at hand so as to form and act on true beliefs; this is, however, not the case. It is self-evident why it ought to be the case that when an organism is sure that a predator is around that it ought to initiate predator avoidance behaviour. It might not be clear why this ought to be the case in a

situation where the presence of a predator is unconfirmed, especially if the initiation of this behaviour imposes a cost.

If we understand the above scenarios in terms of costs incurred by the organism and we consider those costs as a variable that influences the type of system likely to be selected by evolution, it becomes clear why responding to a false alarm is not only right but can be understood as rational. Responding to a false alarm definitely imposes a cost, but it is a lesser cost on the organism when compared to the cost incurred if it happens that a predator is around and the organism fails to respond appropriately. Responding to a false alarm could be better, in terms of cost incurred by the organism, than disregarding information or the pursuit of additional information: information intended to establish the veracity of the beliefs an organism has if a predator is around.

The fact that we are dealing with incomplete information, information that is often ambiguous (opaque environments), it makes more sense, is adaptive and therefore rational, to assume that information that points towards the presence of a predator is accurate and therefore to initiate avoidance behaviour. If the information is false, because it happens that there is no predator, initiating avoidance behaviour imposes a cost on the organism, but the cost is significantly lower than the cost imposed by discounting the information or seeking more information when a predator is around.

A false-positive or type one error, assuming that there is a predator when one is not around, has survival value because it results in a rational belief which in turn results in the initiation of the most rational action, action that imposes the least cost, given the uncertainty. A system that errs (produces false-positives), may, in fact, have more survival value, and therefore potentially confer better fitness, than a system that produces true beliefs *ceteris paribus*. Assuming that the information that indicates the presence of a predator is true and thus forming beliefs based on it, is rational given the potential costs.

A false-negative or type two error, assuming that a predator is not present when one is present, also increases fitness, albeit indirect fitness. Indirect fitness refers to fitness as it is applied to an organism's kin. Indirect fitness deals with the survival of the unit of selection, the gene, and not a specific instantiation of it, namely an organism's direct offspring. As such, if I died protecting individuals that share my genetic heritage, my sacrifice can be understood as fitness-enhancing.

If an organism makes a type two error as a consequence of discounting the evidence at its disposal, or by seeking more information and ultimately getting killed, that response can be fitness-enhancing for the kin of the organism in question. If the death of the organism in question results in its kin acquiring new information, e.g., learning about a new predator, the death of the organism in question can be viewed as adaptive given that it will increases the chances of survival of the kin. When an organism makes a mistake that its kin can learn from, the mistake in question increases the probability of survival and thus the fitness of the kin: "Baldwin argues that adaptive plasticity allows sub-optimal individuals to acquire higher

fitness. Hence, learning improves the survival of the population of such individuals and thus it facilitates that the genetic evolution may proceed." (Sznajder *et al.*, 2012:301).

As a means of illustrating this point consider the following example. A group of foraging animals comes across a new type of fruit. This fruit has no telltale signs of being poisonous. These animals could forgo this fruit on the assumption that it is poisonous; however, forgoing the fruit when it is *not* poisonous could result in forgoing a valuable nutritional resource. Given this, the leader of the group decides that he will taste the fruit, therefore determining whether or not it is poisonous. If the fruit is indeed poisonous, and the leader of the group dies, the remaining animals will learn that that particular fruit is poisonous and thus ought to be avoided. The death of the leader has a net positive effect in that his group, which can be comprised of his direct decedents (kin) or unrelated conspecifics, will increase their probability of survival and thus the probability of reproducing by simply learning from the consequences of their leader's behaviour.

It naturally seems more adaptive to commit a type one error, false positive, than it would be to make a type two error. Both types of error can, however, be fitness conferring. Looking at the discussion pertaining to type one and two errors it ought to be evident that rationality does not and should not be understood as implying truth. Rationality should be understood as implying responsiveness to reason.

As human beings, we often impose this type of rationality on each other. This is indicated by the fact that claims about the rationality or irrationality of a person, specifically their actions, beliefs, and desires, often take the form of counterfactual claims. Counterfactual reasoning is a type of reason responsiveness, in that thinking counterfactually entails inverting a particular conclusion by imagining that the reasons that entail a conclusion are different and as a consequence, the conclusion is or ought to be different. Counterfactuals can be used to establish causality in the form of counterfactual theories of causation. Counterfactual theories of causation take the form, if ***A*** had not occurred, ***C*** would not have occurred. ***A*** is the antecedent and ***C*** is the consequent such that had it not been for ***A***, ***C*** would not have happened.

Counterfactuals can also be used to assess rationality, rationality understood as reason responsiveness. This is due to the fact that rationality, like counterfactual thinking, requires having a sensitivity to the causal influence of reasons. Consequently if one is able to employ counterfactual thinking one must be rational or at the very least possess the capacity for rationality (understood as reason responsiveness) because they understand the causal nature of reasons. Assessing somebody's rationality using a counterfactual implies assessing whether a person understands that the current situation could have been different had it been the case that the facts at hand were different.

When we want to determine whether or not an action and/or belief is rational/reasonable, we often ask whether or not the person who engaged in that action or holds that belief would have done or believed the same under different circumstances. We also establish rationality by asking whether another individual, a rational individual, would have acted or believed in the same way given the same information. Questioning whether or not another individual would

have had the same beliefs or acted in the same way as the person in question, is an attempt to determine whether or not the beliefs of the person in question are responsive to reason and therefore rational, this is the reasonable man test.

This way of testing reason responsiveness or rationality also emphasizes a point I made earlier. I earlier made the claim that responsiveness is more important than responding correctly. At times we often say that a person responded incorrectly; however, given the situation, i.e., the reasons at play, the response makes sense and is therefore rational given the context. Utterances of this sort indicate that rationality is measured using the appropriateness of a response to the information at hand as opposed to its objective truth. A rational person or organism is one who forms and alters his or her beliefs on the basis of the evidence or reasons that they have.

Some might find it difficult to accept that organisms as complex as human beings are rational with respect to how they respond, and some might, therefore, find it difficult to accept rationality as responsiveness to evidence (as I have defined it). I, however, think that my version of responsiveness to evidence can account for the rationality attributed to human beings because it can result in beliefs that are true. If you recall, I made the claim that beliefs are formed on the basis of evidence. I also made the claim that beliefs are maintained on the basis of evidence. In responding to evidence, an organism could end up with a false belief; this belief would be false but nonetheless responsive to evidence in that it was occasioned by some evidence. Once this belief is formed, its maintenance is dependent on whether or not the acquisition of new or additional information justifies it. If the organism acquires new evidence and the evidence undermines the belief in question, it ought to be discarded. If, on the other hand, new evidence *confirms* the belief in question, it ought to be maintained. Given the fact that organisms are always acquiring new information, the beliefs an organism has are always subject to review and therefore falsification. Beliefs which pass falsification can be taken to be true or at the very least rational. My version of responsiveness to reason can, therefore, result in true beliefs. My version of responsiveness does not presuppose the truth of a belief as a starting point but it can end up with the truth as a by-product because it insists on sensitivity to evidence as a precondition for both the creation and maintenance of beliefs.

What about desires? Are desires rational? The above question should be divided into two. The first question should be: are complex desires rational? The second: are simple desires rational? I will start my discussion by looking at complex desires. Are complex desires rational?

This question has partly been answered in both this and the previous chapter; as a result, much of the following exposition will be somewhat repetitive. Complex desires have beliefs as their proper determinants. This was illustrated in the case discussed in the previous chapter. In the previous chapter, I referred to a desire for vitamin C. The desire for vitamin C was predicated on the belief that vitamin C has curative properties. On this assumption, a desire for vitamin C is based on the belief that vitamin C cures or prevents the flu and a corresponding desire to not get the flu. The underlying assumption is that if one does not have a belief that vitamin C cures the flu, then one will not have a corresponding desire to take vitamin C.

If one does, however, have a desire, a desire for vitamin C, we can expect the person who has the desire to have it because of a reason: namely, a belief. The belief could be that vitamin C has curative properties; it could be the fact that vitamin C increases intelligence – whatever the belief, a complex desire has to have a belief as its proper determinant. Given the fact that I have articulated an account of rationality that is reason responsive, complex desires are rational if and only if they are responsive to their proper determinants. Complex desires are therefore rational when they are responsive to beliefs; this is to say complex desires are rational if they change in response to the beliefs that engender them. In the vitamin C case, if a person who believes that vitamin C cures the flu were to read in the journal *Nature* where scientists have illustrated that vitamin C neither cures nor prevents the flu, that person's desire, if it is engendered by the belief that vitamin C has curative properties and is therefore rational, ought to change. If the desire persists and its persistence is predicated on the grounds that the person with the desire *still* believes the claim that vitamin C has curative properties in spite of the evidence, the desire would be irrational. The desire would be irrational because the belief that engenders it would be irrational, but not because it (the desire) is wrong.

Richard Brandt (1969) can be read as espousing a definition of rationality that is based on counterfactual reasoning and therefore an account that is reason-responsive. Brandt claims that one is rational or has rational desires if the desires one has can survive or withstand logical criticism and being subjected to additional information: "But if it is true that intrinsic desires can be extinguished or modified by bringing knowledge-facts and logic-into contact with them, I do not think it is unhelpful to use the word 'rational' to designate the status of the desires which survive this confrontation." (Brandt, 1969:44).

Julian Savulescu (1994) also asserts this point. Savulescu, however, does not speak about counterfactual reasoning: he speaks about vivid imagination. For a person to be rational according to Savulescu, the person has to be able to vividly imagine alternative states of affairs that could result as a consequence of the information they have at hand: "…***P*** knows what ***A*** and ***B*** are like, but she must also imagine what ***A*** and ***B*** would be like for her. I call this vivid imagination." (Savulescu, 1994:196). The ability to vividly imagine differing states of affairs requires being able to change reasons or any other casual factors so as to be able to conjure up different imaginations. The ability to vividly imagine requires reason-responsiveness and as such, I think Savulescu's account of rationality is reason-responsive much like mine.

Savulescu and I might be read as making similar (albeit different) points in that he emphasises the lack of logical mistakes in processing information as a necessary condition for rationality. Some people might be inclined to argue that my favouring *responsiveness* as opposed to *correct responding* to be in contravention of the lack of logical mistake criterion espoused by Savulescu; this is, however, incorrect. Correctly processing information (i.e., a lack of logical mistakes) does not imply objectively correct, error-free, outcomes. We can separate issues pertaining to the accurate processing of the available information from issues pertaining to whether or not the inputs are objectively correct, and therefore the object correctness of outcomes.

Savulescu can be read has making a claim about the veracity of inputs, beliefs in this case, or a claim about rules that govern how beliefs inform desires. I tend to view him as making the second claim: a claim that is substantially weaker than the first. If my interpretation is correct, Savulescu is more concerned with how beliefs give rise to desires. According to my interpretation, Savulescu is more concerned with the manner in which desires are formed, as opposed to their objective correctness. Savulescu is more concerned with desires logically emanating from beliefs, i.e., desires being entailed by beliefs as opposed to the veracity / truth of beliefs that give rise to desires.

Some might be inclined to rebut this claim by saying that desires are not responsive to reason by illustrating that there are cases where a desire is maintained even in the presence of countervailing reasons. Take as an example a person who fails to heed his doctor's advice about his drinking habits. If the doctor says that the person in question ought to stop drinking but the person still has a desire for and still continues consuming alcohol, that could be viewed as an illustration of the fact that desires are not responsive to reason and are therefore irrational. I do not agree with this sentiment. The mere fact that some desires are irrational does not mean that desires, in general, are irrational. Some desires will be rational whilst some will be irrational. I am making the claim that it is the proper function of desire-making mechanisms to result in rational desires, but it does not necessarily mean that all desires will be rational, because desire-making mechanisms can malfunction.

In the second chapter, I argued for the adoption of a reward-based theory of desire. I argued for a reward-based theory of desire because I claimed that it accounts for all the phenomena that we associated with desire. I argued that while this theory accounts for desire in terms of reward, it explains phenomena like pleasure and motivation and their link to desire, phenomena which remain unexplained in a theory that strictly identifies desire with either pleasure or motivation. The desire-as-reward theory explains pleasure and motivation, but can it explain the rationality of desire? If my endorsement of the desire-as-reward theory is to be taken seriously, it also has to account for the rationality of desire.

I think that the desire-as-reward theory does account for rationality. The desire-as-reward theory champions a reason- or evidence-responsive account of desire at the level of both desire formation and desire maintenance. The fact that reason-responsiveness results in rationality implies that any system that shows sensitivity to reason(s) in its function can be assumed to be rational. The dopaminergic or reward system is reason-responsive and is therefore rational.

How does the dopaminergic system exemplify rationality or responsiveness to reason in the manner in which it functions? An organism discovers or learns about what is rewarding and therefore what ought to be pursued when it engages in activities that result in a reward. This means that reward and therefore desire is formed in response to evidence. If an organism does not experience reward, the organism in question does not develop desire.

The maintenance of desire is also subject to reason-responsiveness and is therefore rational. Once a desire is acquired, if it can be predicted using some cue, the organism learns to associate the reward with the cue such that if the cue is seen, it triggers the desire. The triggering of the

desire merely means that the organism releases an amount of dopamine it anticipates it will get based on past experience. When this anticipatory dopamine is released on the basis of a predicting cue, it sets up a baseline which will be measured against the release of dopamine (reward), which will be initiated by the desired object or state of affairs. The organism then compares the reward it got in the past with the reward it is currently receiving to determine whether or not there is a change in the reward value associated with the object or state of affairs. This process results in the maintenance or abandonment of desire. The abovementioned process is self-evidently reason-responsive: it explains whether something retains its status of being a reward and thus being desired is based on evidence, and is therefore rational.

What about simple desires? Are simple desires rational? Simple desires are rational; they are rational in two ways. Simple desires are rational in virtue of their responsiveness to their proper determinants. There are, however, instances where simple desires are non-responsive to their proper determinants in which case beliefs enter into the equation. In these instances, a simple desire is rational if and only if it corresponds to certain beliefs that we have. If I have recently eaten and feel hungry due to viewing images of food, I can rule that my current instance of hunger is not *genuine* hunger, but rather a response to the images of food, and is therefore irrational.

The preceding chapter illustrated that desires, simple and complex, are rational if we view rationality as responsiveness to reason. I argued that responsiveness to reason results in rationality because rationality is nothing more than believing and desiring on the basis of evidence.

Conclusion

In this paper, I set out to examine whether or not desires can be viewed as rational. In order to answer this question, the paper first grappled with issues pertaining to what desire is. In this paper, I proposed the re-imagining of desire. By re-imagination, I proposed the naturalisation of desire, the empirical study of desire and the other parts of the brain that it is connected to. I proposed that desire, reward, is produced by the function of the dopamine system. In suggesting that desire be studied using neuroscientific methods, I suggested that what is desirable and therefore rewarding can only be determined *a posteriori*, on the basis of how it activates the dopaminergic system.

Given the *a posteriori* nature of desire, organisms are not born with intrinsic desires, organisms learn what is desirable depending on how rewarding its behaviours are under certain environmental conditions. The rewarding behaviours are likely to be repeated and therefore reinforced, hence the assumption that having a desire is the process of undergoing reinforcement learning.

This account of desire is more amenable than the previous accounts of desire because it gives us the ability to account for the various phenomenon that we associate with desire. One can, for example, explain why people associate desire with pleasure by exploring whether or not there are anatomical connections between the dopamine system and the pleasure centres of the brain and the manner in which they are connected.

I moved on to proposing the existence of two types of desires, complex and simple desires. Complex desires have beliefs as their proper determinants while simple desires have the organism's physiological state at a given time as their proper determinants. I suggested that both complex and simple desires are rational. By rational, I have in mind responsiveness to proper determinants, reason, which I take not to imply correct responsiveness. I illustrated this point by showing that under the right circumstance, evolutionary circumstance, both type one and type two errors ought to and can be viewed as rational, if this is the case a system can exhibit rationality whilst also being prone to error so long as it is responsive to its proper determinants and the errors it commits impose a lower cost on the organism than correct answers.

I then argued that the function of the dopamine system is reason responsive and therefore rational. By reason responsive, I mean that desires are not formed arbitrarily but rather in response to the reward experienced by the organism when it performs certain actions under certain circumstances. I then conclude that if the dopamine-releasing system is responsible for the creation of desire and its functionality is reason responsive and reason responsiveness implies rationality, desires must also be rational in both how they are formed and maintained.

Bibliography

1. Ainslie, G. (2001), *Break-down of Will*, Cambridge University Press.
2. Aristotle, Ross, W. D., and Brown, L. (2009), *The Nicomachean ethics*. Oxford: Oxford University Press.
3. Ayano, G. (2016), Dopamine: Receptors, Functions, Synthesis, Pathways, Locations and Mental Disorders: Review of Literatures. *Journal of Mental Disorders and Treatment.* Vol, 2. No, 2: 1-4.
4. Bayer, H.M. and Glimcher, P. W. (2005), Midbrain dopamine neurons encode a quantitative reward prediction error signal. *Neuron*. Vol, 47. No, 1: 129-141.
5. Brandt, R.B. (1969-70), "Rational Desires," *Proceedings of the American Philosophical Association 43*: 43-64.
6. Bromberg-Martin, E.S., Matsumoto, M., and Hikosaka, O. (2010), Dopamine in motivational control: rewarding, aversive, and alerting. *Neuron*. Vol, 68. No, 5: 815-834.
7. Butlin, P. (2017), Why Hunger is not a Desire. *Review of Philosophy and Psychology*. Vol, 8:617-635.
8. Darwall, S. (2003), Desires, Reasons, and Causes. *Philosophy and Phenomenological Research.* Vol, 67. No, 2:436-444.
9. Davidson, D. (1963), Actions, Reasons, and Causes. *The Journal of Philosophy.* Vol, 60. No, 23:685-700.
10. Caplin, A. and Mark, D. (2008), Dopamine, Reward Prediction Error, and Economics. *Quarterly Journal of Economics*. Vol, 123. No, 2: 663–702.
11. Castro, D.C and Berridge, K.C. (2014), Opioid Hedonic Hotspot in Nucleus Accumbens Shell: Mu, Delta, and Kappa Maps for Enhancement of Sweetness "Liking" and "Wanting". *The Journal of Neuroscience*. Vol, 12. No, 34: 4239–4250.
12. Chouinard, P.E., Paus, T. (2006), The Primary Motor and Premotor Areas of the Human Cerebral Cortex. *Neuroscientist*. Vol, 12. No, 2: 143–152.
13. Christman, J. (1991), Autonomy and Personal History. *Canadian Journal of Philosophy.* Vol, 21: 1-24.
14. Elster, J. (1983), *Sour Grapes: Studies in the Subversion of Rationality.*
15. Frankfurt, H.G. (1971), Freedom of the Will and the Concept of a Person. *The Journal of Philosophy*. Vol, 68. No, 1: 5-20.

16. Friedrich, D. (2014), Belief and Motivation. *Theoria.* Vol, 80:255-268.

17. Gardner, A. and Ubeda, F. (2017), The meaning of Intragenomic Conflict. *Nature Ecology and Evolution.* Vol, 1.

18. Jackson, S.A.W., Horst, N.K., Pears, A., Robbins, T.W., and Roberts, A.C. (2016), Role of the Perigenual Anterior Cingulate and Orbitofrontal Cortex in Contingency Learning in the Marmoset. *Cerebral Cortex*. Vol, 23. No, 7: 3273–3284.

19. Kekes, J. (1972), Fallibilism and Rationality. *American Philosophical Quarterly.* Vol, 9. No, 4:301-309.

20. Klossek, U.M.H., Russell, J., and Dickinson, A. (2008), The control of instrumental action following outcome devaluation in young children aged between 1 and 4 years. *Journal of Experimental Psychology*. Vol, 137. No, 1:39-51.

21. Kringelbach, M. L. and Berridge, K. C. (2010), The Functional Neuroanatomy of Pleasure and Happiness. *Social Research*. Vol, 77. No, 2: 659-678.

22. Leon, M. (2001), The willing addict: Actor or (helpless) bystander? *Philosophia*. Vol, 28: 437-443.

23. March, J.G. (1986), Bounded Rationality, Ambiguity, and the Engineering of Choice. In: Jon Elster (ed.). *Rational Choice*. Oxford: Basil Blackwell, Chapter 6-142.

24. Matyas, F., Sreenivasan, V., Marbach, F., Wacongne, C., Barsy, B., Mateo, C., Aronoff, R. and Petersen, C. C. H. (2010), Motor Control by Sensory Cortex. *Science, New Series*. Vol. 330, No. 6008:1240-1243.

25. Millikan, R. (1983), *Language, Thought, and Other Biological Categories: New Foundations for Realism*, Massachusetts Institute of Technology Press.

26. Niv, Y., N.D. Daw, D. Joel, and P. Dayan. (2007), Tonic dopamine: opportunity costs and the control of response vigor. *Psychopharmacology* 191 (3): 507–520.

27. Raz, J. (2008), On the Guise of the Good. *Legal research paper series*. No, 43.

28. Savulescu, J. (1994), Rational desires and the limitation of life-sustaining treatment. *Bioethics*. Vol, 8. No, 3: 191-222.

29. Schroeder, T. (2004), *Three Faces of Desire,* Oxford University Press.

30. Schroeder, T. (2006), Desire. *Philosophy Compass*. Vol, 1. No, 6: 631-639.

31. Schultz, W. (2007), Multiple dopamine functions at different time courses. *Annual Review of Neuroscience* 30: 259–288.

32. Schultz, W. (2016), Dopamine reward prediction error coding. *Dialogues in Clinical Neuroscience*. Vol, 18. No, 1: 23-32.

33. Schussler, P., Kluge, M., Yassouridis, A., Dresler, M., Uhr, M., and Steiger, A. (2012), Ghrelin levels increase after pictures showing food. *Obesity.* Vol, 20. No, 6: 1212-1217.

34. Smith, M. (1987), The Humean Theory of Motivation. Mind. Vol, 96. No, 381: 36-69.

35. Spurrett, D. (2015), The Natural History of Desire. *South African Journal of Philosophy*. Vol, 34. No, 3: 304-313.

36. Sterelny, K. (2003), *Thought in a Hostile World: The Evolution of Human Cognition*, Wiley-Blackwell.

37. Stich, S. (1993), *The Fragmentation of Reason*, MIT Press.

38. Strawson, G. (1994), Mental Reality, MIT Press.

39. Sznajder, B., Sabelis, M.W., and Egas, M. (2012), How adaptive learning affects evolution: reviewing theory on the Baldwin effect. *Evolutionary Biology*. Vol, 39. No, 3: 301-301.

40. Velleman, D. (1992), Guise of the Good. *Nous.* Vol, 26. No, 1: 3-26.

41. Wang, G., Volkow, N.D., Thanos, P.K., and Fowler, J.S. (2009), Imaging of Brain Dopamine Pathways. *Journal of Addiction Medicine*. Vol, 3. No. 1: 8-18.

42. Watson, G. (1975), Free Agency. *The Journal of Philosophy.* Vol, 72. No, 8: 205-220.

43. Witnauer, J.E., and Miller, R.R. (2007), Degraded contingency revisited: Posttraining extinction of a cover stimulus attenuates a target cue's behavioral control. *Journal of Experimental Psychology: Animal Behavior Processes,* Vol, 33. No, 4: 440-450.