What does it Feel Like to Be a Bat? Searching for Ways to Investigate Conscious Emotion in Non-human Animals

by

Walter Sánchez Suárez

A Thesis

presented to

The University of Guelph

In partial fulfilment of the requirements

for the degree of

Doctor of Philosophy

in

Animal and Poultry Science & Neuroscience

Guelph, Ontario, Canada

Abstract

What does it feel like to be a bat? Searching for ways to investigate conscious emotion in non-human animals

Walter Sánchez Suárez
University of Guelph, 2015

Advisor:
Professor Georgia Mason

The capacity for conscious emotion is the crucial attribute for deserving moral consideration. Unfortunately, the impossibility of knowing what it is like to be another being from *her own* point of view – knowing "what it is like to be a bat", as Thomas Nagel famously put it – implies that we can only make inferences regarding whether she can experience conscious emotion. In humans we take advantage of verbal self-reports for coming up with strong inferences in this regard. In the absence of verbal reports, however, the tools we have for investigating which beings are capable of conscious emotion are less conclusive. In this work I examine the relationship between consciousness and emotion, the limitations of the tools we currently have for inferring primary conscious emotion in non-verbal beings, and propose promising new strategies for making progress with this task. The second part of this thesis provides evidence from verbal humans suggesting that at high levels of discrimination performance, the use of internal states as discriminative stimuli in an operant paradigm is highly likely to depend on conscious processing of this information. I therefore argue that this paradigm is a behavioural marker of states of conscious emotion in the absence of verbal self-reports, and thus, if emotional states are used as discriminative stimuli, a potential valid tool for coming up with stronger inferences when investigating conscious emotion in non-verbal animals.

Acknowledgements

I would like to thank many human and non-human friends for helping me to navigate this difficult challenge I decided to pursue. You know who you are, and I feel extremely fortunate for sharing my life with you. Some do not physically exist anymore, but they remain alive in my conscious mind.

Thanks Georgia, for making this formative experience possible. I have learnt a lot through these years, and my view of the world will never be the same. Thanks to the wonderful members of my advisory committee – Ian Duncan, Melissa Bateson, Liz Paul and Boyer Winters –, and to the examiners at my comprehensive exams and thesis defence – Kristin Andrews, Mark Fenske, Lee Neil and Flavio Schenkel. Some people have contributed with valuable comments to this work, thanks to them too: Colin Allen, Lucas Arinero, Zoltan Dienes, Dora Duka, Carole Fureix, James Grau, Óscar Horta, Becky Meagher, Adam Shriver, and William W. Stoops.

I would also like to show my gratitude to the experts on drug discrimination who provided me with comments and data: Warren K. Bickel, Joshua A. Lile, Kenneth A. Perkins, Alison Oliveto, Kenzie L. Preston, Craig R. Rush.

I feel very lucky to have shared my time in Guelph with my friends and colleagues at The Mason Lab, and from The Animal Behaviour & Welfare Group of the University of Guelph.

Finally, thanks to the Fundación la Caixa for awarding me with one of their scholarships and making this experience possible.

Table of contents

List of tables

List of figures

Chapter 1

What does it feel like to be a bat?

"What is it like to be a bat?" In this influential work, Thomas Nagel formulated an important argument with regard to the study of consciousness: the impossibility of knowing what it is like to be another being from *her own* point of view. Nagel is obviously right in that, at present, it is not possible to experience what it is like to be another (conscious) being from her own perspective (see Nagel, 1974; see also Hyslop's 'The problem of other minds' in the The Stanford Encyclopedia of Philosophy, 2015). Fortunately, studying what it is like to be that subject from our perspective, however, is a different question. Making valid inferences with regard to what 'it is like to be that being' in this sense is likely a tractable task. As Kristin Andrews put it, Nagel himself argued that "the best we can do is to imagine what it would belike for *us* to be bat-like" (Andrews, 2015, p.5).

It has been proposed that the 'hard problem of consciousness' – the theoretical dualist *explanatory gap* between the conscious state we experience and the related physical process we can study (see Chalmers, 1995) – implies that consciousness "cannot currently be studied by the usual methods of science" (Dawkins, 2015). Scientists studying consciousness in verbal humans, however, are making valuable progress understanding this biological process. With regard to this 'hard problem', the cognitive neuroscientist Stanislas Dehaene recently wrote "once we clarify how any piece of sensory information can gain access to our mind and become reportable, then the insurmountable problem of our ineffable experiences will disappear" (Dehaene, 2014, p.10). In Chapter 2 we will see that by taking advantage of human verbal self-reports, even though it is not possible to experience what it is like to be that subject from her own point of view, we can come up with valuable inferences with regard to the conscious states of verbal humans.

But, can we investigate these issues in non-verbal beings? Or, more precisely, since this thesis deals with the question of conscious emotion, can we ever make valid inferences regarding what it *feels like* to be a non-verbal being? Many think that we

can (see, for instance, DeGrazia, 1996; Allen et al., 2005; Shriver, 2006; Edelman & Seth, 2009; Mason, 2011; Varner, 2012). In this thesis, I will argue that by taking advantage of human verbal self-reports for validating processes that depend on conscious processing of emotional information, it is possible to make progress with this question – one that has crucial ethical and legal implications.[1] Thus, in Chapter 2 we will see that empirical studies in humans are starting to show that certain processes (e.g., behaviour) cannot be dissociated from the conscious processing of some of their related information. These are obviously assessed by verbal self-reports, and, as I will argue, we can take advantage of this approach for validating conscious-dependent processes in humans related to emotion that, in turn, can be used for investigating conscious emotion in the absence of verbal reports. There, I will underline that although this approach can be used across the animal phyla, the inferences we can make are strongly determined by the degree of phylogenetic proximity between verbal humans (our gold standard) and subjects of other species. In Chapter 3, I will propose that the use of emotion (and other interoceptive states) as discriminative stimuli in an operant paradigm qualifies as behaviour potentially dependent on conscious processing of those states; and in Chapter 4, I will take advantage of human verbal self-reports for assessing whether this behaviour is a valid tool for investigating conscious emotion in the absence of verbal reports. Finally, in Chapter 5 I will present my conclusions and future directions for this approach.

[1] According to the most relevant ethical theories (e.g., Singer, 1993; Persson 1993; Pluhar, 1995; Rowlands, 1998; Regan, 2004), the capacity for consciousness in general, and for suffering and having pleasurable experiences in particular, are crucial for deserving moral consideration, regardless of the species (see Horta, 2010). Beings in possession of this capacity can be harmed in ways that matter to them, and their

Chapter 2

How can we assess conscious emotion in non-human animals?

Table 1 shows the acronyms used in these chapter:

Table 1 – List of acronyms

AC	Access consciousness
CR	Conditioned response
CS	Conditioned stimulus
CNS	Central nervous system
DS	Discriminative stimulus
EEG	Electroencephalographic signals
EMG	Electromyographic responses
GW	Global workspace
GWT	Global workspace theory
HOC	Higher order consciousness
HOT	Higher order theory
IT	Inferior temporal cortex
PET	Positron emission tomography
EEG	Electroencephalographic signals
ACC	Anterior cingulate cortex
2AFC	Two alternative forced choice operant task
DS	Discriminative stimulus
IT	Inferior temporal cortex
PC	Phenomenal consciousness
PrC	Primary consciousness
UR	Unconditioned response
US	Unconditioned stimulus
2AFC	Two alternative forced choice operant task

1. Introduction

Emotions are states with valence, that is, they can be positive or negative. Zajonc (1998), for instance, defined emotion as "[internal] stimuli, processes, or responses that involve…the property of being good/bad". These states have often been regarded as intrinsically conscious. William James, for example, argued in 1884 that emotions are pleasurable or "displeasurable" felt states typically induced by bodily reactions to external or internal (i.e., from within the body) stimuli. In James' own words, "the bodily changes follow directly the perception of the exciting fact, and…our *feeling* of the same changes as they occur is the emotion". This, and similar views common among psychologists, implies that emotional states are always 'felt', i.e. conscious and available for report in subjects capable of language (see Clore, 1994; Ellsworth, 1994; Frijda, 1999).

Here, however, I will consider other definitions that do not entail these states being necessarily conscious (see Section 3), and can therefore be applied to both humans (who can verbally report conscious emotional feelings) and non-human animals (who cannot). For instance, Edmund T. Rolls (2014, p. 14) provides an operational definition by which "emotions are states elicited by rewards and punishers, that is, by instrumental reinforcers". According to Rolls (2014, p 13, see also pp. 63-66), emotions "have the evolutionary utility of specifying the goals for actions, and are states in their own right, which specify what actions may hope to achieve, but in which the action is arbitrary, that is, not specified by the emotional states".[2] Other views, however, do not necessarily require novel arbitrary operants. Anderson and Adolphs (2014), for example, define emotion as "an internal, central (as in central nervous system) state, which is triggered by specific stimuli (extrinsic or intrinsic to the organism). This state is encoded by the activity of particular neural circuits that give rise, in a causal sense, to externally observable behavior, as well as to associated cognitive, somatic, and physiological responses". According to this latter view, emotions are typically associated with behavioural responses related to rewards or punishers (e.g., bees' proboscis extensions when finding nectar [Bateson et

[2] From Rolls' viewpoint, emotions can *only* become conscious states if they are further processed within a language system, such as the human verbal language (see Rolls, 2014).

al., 2011]; and rats' limb withdrawal responses when given electric shocks to their paws – a punisher, [see Grau et al., 1998]), and also to physiological and neurophysiological changes (think, for instance, of the increased heart rate you would experience if someone tried to rob you in a dark street). As inferred from self-reports in humans, emotions are often processed consciously, therefore giving rise to the conscious feeling of emotion (e.g., a feeling of fear), but as we will review in this chapter, this is not always so.

The assessment of the conscious feeling of emotion is of the uppermost ethical importance. As argued in Chapter 1, according to the most relevant normative ethical theories, the capability for suffering and experiencing pleasure – i.e., processing emotional information consciously – is both necessary and sufficient for attributing a being with moral consideration. This knowledge, in turn, is necessary for issuing laws regarding these subjects' wellbeing (see Chapter 1). Relevant to this, Section 2 will deal with the potential role of consciousness in behaviour. Yet, in Section 3 we will see that identifying which subjects are capable of conscious emotions in the absence of verbal reports is a very complicated task. There we will see that many behavioural, physiological and neurophysiological pieces of evidence related to emotion are often not very informative of whether or not states of emotion are processed consciously. In Section 4, several strategies for making progress in the assessment of conscious emotion in non-verbal beings will be proposed, and it will be argued that their usefulness is dependent on evolutionary convergence, that is, whether or not subjects' traits related to states of emotion are homologous to those of verbal humans. Finally, Section 5 will summarize the main points in this chapter.

2. Consciousness and its role in behaviour

2.a. Does consciousness play a causal role in behaviour?

Here I will briefly review some terminological issues to do with consciousness, and then discuss its potential role in behaviour. I will use 'blindsight' and similar pieces of evidence, trace and delay conditioning, and priming effects as key test cases.

Consciousness is an umbrella concept for different sorts of subjective states. There are several ontological theories of consciousness, ranging from radical dualist approaches based on the idea that conscious states or experiences are not part of the physical world, through to fully physicalist ones. As an example of the latter, theories such as 'the type-type identity' propose that conscious states are not just generated by the activity of some animals' central nervous systems (CNSs), but that the conscious qualitative property and the neural property are identical – just as are the properties of being water and being composed of H_2O molecules (for an in depth review of the different ontological theories of consciousness see Robert Van Gulick's excellent entry in The Stanford Encyclopedia of Philosophy, 2014). The current scientific physicalist approach to the study of consciousness typically considers three different concepts: 1) vigilance[3] – the state of wakefulness, which varies when we fall asleep or wake up; 2) attention – the focusing of our mental resources onto a specific piece of information; and 3) *conscious access* – the fact that some of the attended information enters our awareness and – at least in verbal individuals – becomes reportable to others (Dehaene, 2014, p. 8). This chapter will mainly focus on the latter notion of consciousness and its role in behaviour.

Do conscious experiences, as conscious access (i.e., information processed consciously) play a causal role in behaviour? In 1874, Thomas H. Huxley disparagingly compared the conscious events of humans and other animals to "a steam whistle that contributes nothing to the work of a locomotive". Consciousness, he was arguing, is simply an epiphenomenon of the workings of the brain. In contrast, more than a century later, the majority of consciousness researchers support

[3] Not to be confused with anti-predatory responses.

the hypothesis that consciousness does in fact play an active role in behaviour. According to the neuroscientist Stanislas Dehaene (2014, p. 101), "the capacity to synthesize information over time, space, and modalities of knowledge, and to rethink it at any time in the future, is a fundamental component of the conscious mind, one that seems likely to have been positively selected for during evolution". This capability seems to enable humans (and possibly other animals) to predict and imagine potential scenarios, and eliminate hopeless ones before even trying them. Richard Dawkins formulated this idea by arguing that "consciousness should be favoured by natural selection, simply because…the substitution of imagined or symbolic or vicarious behavior for real trials which, if erroneous, may have fatal consequences" (see Dawkins, 1976; and also Popper, 1978). In Dehaene's view (2014) "consciousness supports a number of specific operations that cannot unfold unconsciously. Subliminal information[4] is evanescent but conscious information is stable – we can hang on to it for as long as we wish". This is consistent with Seth and colleagues' (2005) assertion that "there is very little evidence for long-term learning of unconscious input. In contrast, the evidence of learning of conscious episodes is overwhelming. Even implicit learning[5] requires conscious attention to the stimuli from which implicit regularities are (unconsciously) inferred". Additionally, "the capacity of consciousness at any given moment seems limited to one consistent scene" (Seth et al., 2005). In the words of Dehaene (2014, p. 89) "consciousness also compresses the incoming information, reducing an immense stream of sense data to a small set of carefully selected bite-size symbols. The sampled information can be

[4] Subliminal information is information that subjects do not consciously access, as in priming experiments, in which subjects are typically exposed to a very briefly show visual stimulus (the 'prime' stimulus) and that influences their response to a second supraliminal (i.e., consciously accessed) stimulus, the 'target' stimulus (see the entry on priming in The Oxford Companion to Consciousness, Bayne et al., 2009).

[5] Skills and knowledge can be acquired either implicitly, where the learning takes place without the learner being consciously aware of it, or explicitly, where the individual is conscious of what is being learned. In other words, implicit learning is caracterized by learing that takes place largely independent of conscious awareness both of the process and of the products of acquisition (see the entry on explicit versus implicit learning in The Oxford Companion to Consciousness, Bayne et al., 2009).

then routed to another processing stage allowing us to perform carefully controlled chains of operations, much like a serial computer. This broadcasting function of consciousness is essential. In humans, it is greatly enhanced by language, which lets us distribute our conscious thoughts across the social network". According to these viewpoints, which, as we will see below, are supported by empirical findings, conscious states are thus adaptive tools. As Gerald Edelman (2003) put it, "the capacity to distinguish among very large numbers of inputs, while integrating them in ways related to the past history of an individual, provides an adaptive advantage not possessed by animals without such [conscious] systems".

Thus, as seen above, the capability of processing information consciously appears to be relevant for increasing the probability of overcoming current and future life-threatening situations, such as adapting to novel scenarios by taking advantage of relevant knowledge synthesized through previous conscious experiences – which, in turn, may trigger responses in a faster and non-conscious manner.[6] In words of Crick and Koch (2003), "it seems to be a great evolutionary advantage to have zombie modes that respond rapidly, in a stereotyped manner, together with a slightly slower [conscious] system that allows time for thinking and planning more complex behavior". Additionally, below we will see that, at least in verbal humans, consciousness seems necessary for attributing mental states to others and thereby predicting their potential behaviour; for introspection;[7] and also for reaching

[6] This is not surprising, since animals provided with CNSs need to constantly process huge amounts of data, and "in addition to the capacity-limited processing of consciousness, another level of information processing seem necessary to explain how organisms can process the vast input of external and internal stimuli. For the system to function efficiently, this other level had to process information in an automatic, parallel mode" (Wiens & Öhman, 2007). Thus, "nonconscious self-regulatory processes may be of help…in difficult circumstances because nonconscious processes are not subject to the same set of limitations as are conscious processes" (Williams et al., 2009).

[7] Introspection is generally regarded as a process by which we learn about our own currently ongoing, or very recently past, mental states or processes (see Schwitzgebel, 2014).

collective decisions (see also Chapter 3 'What is consciousness good for', in Dehaene's 'Consciousness and the brain', 2014).

2.b. Notions of consciousness relevant for understanding its potential role in behaviour

When assessing the relationship between conscious experiences and behaviour, several notions of consciousness have been typically used for understanding how this type of processing may shape behaviour. Ned Block (1995) refined the term 'phenomenal consciousness' (PC) in an attempt to underline the 'subjectiveness' of any conscious experience. This concept is often used synonymously with *sentienc*e, since it "refers to the qualitative, subjective, experiential, or phenomenological aspects of conscious experience, sometimes identified with *qualia*"[8] (Allen & Trestman, 2015).

A related concept is that of 'access consciousness' (AC) (not to be confused with 'conscious access', see p. 3). In Block's view, some information processed by the brain is broadcast "in a 'global workspace',[9] and thereby made available for

[8] In philosophy, the concept of qualia typically refers to the "what it is like" aspect that is intrinsic to any conscious experience. A classic example is "the redness of the color red", since there is something it is like for you subjectively to undergo the experience of seeing this colour (see Tye, 2013).

[9] The Global Workspace Theory (GWT) is a theoretical framework for conscious and unconscious brain events that has tree basic constructs. Bernard Baars defines the GW as a momentary memory that can be accessed by numerous *input assemblies*, such as the active neuronal assemblies involved in visual experiences and motor control, which provides the seeds of specific contents of consciousness. The second construct is a very large set of receiving assemblies, both cortical and subcortical; third are contexts, defined as coalitions of neuronal assemblies, which can select, evoke, and shape the contents of the GW without themselves becoming conscious. Conscious experiences are proposed to arise in the interaction between GW contents, the subjective self – a non-conscious dominant coalition of contexts also called *dominant context* –, and receiving assemblies. According to Baars, the GWT can be thought of in terms of a theatre metaphor, with conscious contents corresponding to

cognitive processing tasks such as categorization, reasoning, planning, and voluntary direction of attention" (Block, 2005; Allen, 2013). According to the Global Workspace Theory, broadcasting information from this global workspace to the rest of the brain is what renders it conscious (Koch, 2014). The relationship between PC and AC, however, is still a matter of debate. A consideration relevant to this chapter is, as put by Bayne and colleagues (2008), whether PC is – or must be – operationalized in terms of AC. According to these researchers, "although certain theorists regard verbal report as the gold standard of consciousness, many (most?) of us are willing to ascribe PC on the basis of something akin to AC (minus verbal report). For example, we are inclined to think that infants and other non-linguistic animals are conscious of their bodies and environment, despite the fact that they may be unable to produce any form of metacognitive commentary on the objects of their experience. Arguably, we are inclined to regard such animals as conscious because they have representations whose contents can be freely deployed for use in reasoning and the rational control of action. In other words, we do seem to take AC – or something very much like it – as an intuitive marker of PC." Although it remains theoretically possible that AC and PC could be dissociated – as in David Chalmer's famous zombies (i.e., human-like creatures who are capable of access consciousness but no phenomenal consciousness[10]), below we will see that we do not know of any examples of humans accessing information that spontaneously shapes behaviour without them, crucially, being able to report phenomenal experiences regarding the conscious construct (i.e. state) in which it is integrated. From this point of view, and in the absence of better evidence, AC as defined by Block does not fully qualify as conscious processing unless it is related to a phenomenal experience. Thus, throughout this chapter I will maintain that consciousness and conscious processing

---

the information presented in a brightly lit spot on a dark theatre stage, communicating with a large unconscious audience, and scripted by in turn by a behind the scenes stage crew, director, and scriptwrter. See Baars' entry in The Oxford Companion to Consciousness (Bayne et al., 2009), and also Baars, 2001.

[10] See David Chalmers' 'Facing Up to the Problem of Consciousness.' (1995) , and Robert Kirk's entry on "Zombies" in The Stanford Encyclopedia of Philosophy (2015).

equate to *conscious access*, which as defined by Dehaene (2014) always entails both AC and PC.

Another two relevant concepts related to consciousness are those of 'primary' (PrC) and 'higher order consciousness' (HOC). According to Seth and colleagues (2005), "consciousness is often differentiated into 'primary consciousness', which refers to the presence of a reportable multimodal scene composed of perceptual and motor events, and 'higher-order consciousness', which involves referral of the contents of primary consciousness to interpretative semantics, including a sense of self and, in more advanced forms, the ability to explicitly construct past and future scenes" (see also Edelman, 1989. Both PrC – a first order process – and HOC are forms of conscious access, and therefore entail AC and PC. Clearly, primary consciousness, as much as higher order consciousness, is relevant to the suffering and well-being of an individual (animal or human) and is of the utmost ethical importance. Therefore it is primary consciousness that is the focus of this thesis.

2.c. Processes dependent and not dependent on conscious access: three hypothetical roles of consciounsess in behaviour

Next I will describe different forms of behaviour which are known either to require conscious access or not to require it. A classic behavioural example that illustrates what can or cannot be done in the absence of conscious access is illustrated in research differentiating conscious vision from 'blindsight', i.e. non-conscious visual perception (see Weizkrantz, 1996). Seth and colleagues (2008) define blindsight operationally "as the capability of some individuals with visual cortical damage to perform visually guided behaviours even though they report the absence of any associated conscious content". This capability is typically assessed through forced-choice tasks, in which subjects with blindsight are presented with visual stimuli in the area of their visual fields related to the cortical lesions and asked to discriminate amongst them. As argued by Seth and colleagues (2008), "two properties of blindsight indicate intuitively that the seeing is not conscious. First, blindsight patients do not spontaneously attempt to use the information practically or inferentially". For example, a glass of water in the blindfield would not be spontaneously picked up by a thirsty blindsight patient (Zoltan Dienes, personal

communication). Second, "blindsight patients themselves think they cannot see" (Seth et al., 2008; see also Chapter 3 in Kristin Andrews' "The Animal Mind", 2015). Another related example is that of 'blindsmell'. In this case, non-consciously perceived air-borne chemicals still affect the behaviour of humans with lesions in areas normally involved in the processing of this sort of olfactory information. This is demonstrated by subjects' ability to discriminate odours above chance without reporting any change in smell perception (Sobel et al., 1999; Li et al., 2010). Similarly, Tsuchiya and Adolphs (2007) found that a patient who had lost the ability to perceive taste – due to extensive bilateral lesions of the insula – described solutions of sugar, saline or lime juice as all tasting 'like pop', drinking them indiscriminately. However, when given a choice between contemporaneously presented beverages, he showed a strong preference for the sugar solution. Thus, despite no apparent conscious experience of the unpleasant salty or sour tastes, the patient showed strong motivational preferences based on the (normal) affective value of the sugar solution when provided with the opportunity for a comparison. These three examples from humans with cortical lesions illustrate how stimuli can potentially be processed non-consciously and still induce changes in behaviour.

Finally, two forms of conditioning, trace and delay, may also provide an example of behavioural processes dependent and not dependent on conscious access of related information. In Pavlovian or classical conditioning, subjects (or a subject, e.g., dog) are presented with a neutral stimulus (the conditioned stimulus – CS, e.g., the sound of a bell) before a second stimulus (the unconditioned stimulus – US; e.g., food) that typically elicits a behavioural response (the unconditioned response – UR; dog's salivation). After several presentations (pairings) of the CS and the US, the CS 'acquires' the property of eliciting a learned or conditioned response (the conditioned response – CR) in advance of the US. In the 'delay' version of classical conditioning, the CS is presented and remains present until the US is presented – with the two stimuli overlapping and then co-terminating. Evidence from self-report in humans suggests that consciously accessing (in the sense of understanding) the relationship between the CS and the US is not necessary for conditioning in this latter case (see Clark et al., 2001; Wiens & Öhman, 2002 & Manns et al, 2002), although this is not always the case (Lovibond & Shanks, 2002 & Kattner et al., 2012). In trace conditioning, however, a silent stimulus-free 'trace' interval follows the CS before the presentation of the US, and in this case only subjects who become consciously aware

of the temporal contingencies of the conditioning stimuli seem to successfully acquire the conditioning (i.e., emit CRs after being presented with CSs) (see Clark et al., 2001 & 2002; Clark, 2011; Bekinschtain et al., 2011; and Raybuck & Lattal, 2014 for self-report-based and neurophysiological evidence relevant to this hypothesis. Thus, trace conditioning may therefore qualify as a *type-c process*, i.e., a process in which awareness of the relationship between the discriminative stimulus (DS) and its consequences is necessary for an intentional action (see Jack & Shallice, 2001).

In the last decades researchers have progressively discovered that many other tasks previously thought to be dependent on the conscious processing of information can be performed non-consciously. For instance, by using experimental techniques of "subliminal priming", it has been found that humans with intact brains can non-consciously process different types of information that then 'primes' or affects their behaviour. Various experiments have shown that words, numbers and pictures subliminally presented to subjects are processed, and eventually integrated – e.g., as a word and its meaning, instead of as several characters presented together – in ways that influence subjects' performance in tasks for which the subliminal information is relevant (see Kouider & Dehaene, 2007 for a review of visual masking studies). As one interesting example, it has been found that expert chess players – but not novice ones – are capable of extracting information from chess configurations subliminally presented to them, *only if* they are relevant for a target (supraliminal) chess configuration presented right afterwards; in particular, it was found that this subliminal information influenced expert players' response times when evaluating whether the target configuration entailed a checking configuration (participants being instructed to indicate whether the target displayed a checking or a nonchecking configuration by pressing a left key or a right key). The researchers concluded that long-term practice prompts the acquisition of visual memories of chess configurations with integrated form-location conjunctions, and that these 'perceptual chunks' enable complex visual processing outside of conscious awareness (Kiesel et al., 2009). This example illustrates the complex effectiveness of some non-conscious processes. In agreement with these findings, Custers and Aarts' (2010) argued that "setting, pursuing, and realizing goals can occur without conscious interventions", since "the mind…continuously and largely unconsciously processes behavioral-relevant information to readily "tell" its owner what she wants and should do to deal with the opportunities and challenges presented by the environment". In fact, by using

subliminal priming and retrospective self-report questionnaires for assessing whether subjects consciously accessed the priming stimuli, research on humans "has demonstrated effects of subliminal stimulation on goal pursuit, such as increased task performance after priming of achievement-related words, enhanced fluid consumption in a taste task after priming of drinking-related words, and an increase in instrumental behavior leading to specific goals (such as helping another person by providing useful comments) after priming of names of significant others (such as a good friend) or occupations (such as nurse) associated with these goals" (Custers & Aarts, 2010).

These findings are in line with previous work suggesting that mental processes such as perception and learning can take place without conscious awareness (as we saw for delay conditioning, for instance; see Kihlstrom et al, 2000 for an in depth review of these non-conscious processes). However, this is just a part of the whole picture, and, probably not all information processed by the brain can be integrated non-consciously. We know that even cross-modal information (i.e., that gathered by distinct sensory modalities) can be integrated in the absence of consciousness processing. For example, the illusion called the "McGurk Effect" (see McGurk & MacDonald, 1976; and also Dehaene, 2014, p. 62) shows that visual and auditory information is integrated before we consciously access it. In this well-known case, a sound ('ba') is paired with the visual component of another sound (a mouth saying 'ga'), leading to the conscious perception of a third sound ('da'). However, as Dehaene (2014, p. 63) puts it, that may be the case for some overlearned information (i.e., that processed many times before), but non-routine bindings – "those that require the *de novo* creation of unforeseen combinations" – do seem to require conscious processing. For instance, the expert chess players above had to consciously process information relevant to their games many times while practicing in order to acquire visual memories of chess configurations that could be processed outside of conscious awareness; however, non-experts without this practice did not have their unconscious abilities. Another very familiar example is that of learning to ride bicycles. Riding a bicycle for the first time seems to depend on conscious integration of the information required to learn this skill. Similarly, learning to read, and to drive a car seem to also depend on consciously accessing the information relevant to these abilities. Thus, as seen above, "consciousness is an elaborate functional property and as such is likely to have been selected, across millions of years of Darwinian evolution, because it fulfills a particular operational role" (Dehaene, 2014). The evidence suggests this role has to

do with integrating information in order to 1) generate representations of the world; 2) predict and imagine potential scenarios, and eliminate hopeless ones before even trying them; 3) and learn novel contingencies – create novel bindings: add new information acquired through conscious integration to the catalogue of behavioural responses that can then eventually be triggered without conscious processing (in a faster, more efficient manner).

2.d. Self-consciousness and its role in behaviour

To complete the picture, self-consciousness is a concept related to HOC also relevant for understanding the relationship between the conscious processing of information and certain behaviour. Philosophers such as Owen Flanagan have argued that all conscious experiences involve at least a weak from of self-consciousness. This view stems from the idea that conscious beings experience the "there is something it is like" intrinsic to any conscious state as 'theirs'" (see Flanagan, 1992, and also Gallagher and colleagues' entry in The Stanford Encyclopedia of Philosophy on the different "Phenomenological Approaches to Self-Consciousness", 2015).[11] However, in contrast, self-consciousness is more often typically conceived as a specific and often quite complex version of conscious processing. This concept is typically used to describe different capabilities, such as "awareness of one's body as a physical object, or as the medium of one's own perception and action (i.e. bodily self-awareness); awareness of one's own mental states (i.e. mental or experiential self-awareness); awareness of oneself as perceived by others, or as a member of a social group such as a family, team, or institution (i.e. social self-awareness); awareness of one-self as a persistent character in the narratives told by oneself and others (i.e. narrative self-awareness)" (Allen & Trestman, 2015). It has been argued that this type of self-consciousness is necessary for certain capabilities, such as introspection,

[11] The question of what types of conscious experiences a being should experience to deserve moral consideration is, as mentioned in the introduction, crucial from an ethical viewpoint. Unfortunately it is beyond the scope of this review, and I will not tackle this question further here.

the possession of a 'theory of mind',[12] and in turn, for experiencing states akin to what humans know as conscious empathy (see, for instance, Carruthers, 2003 & Rolls, 2014, as examples of a higher order theories (HOTs) consistent with this position; see also DeGrazia, 2102 for a different view on the HOTs).

Not surprisingly, self-consciousness has been closely linked to metacognition, i.e, "the ability to monitor one's cognitive processes, the ability to take action to control those processes, and general knowledge about how one's cognitive processes function" (Flavell, 1979; Nelson, 1996; Basile et al., 2014; see also Andrews, 2015, pp. 73–78). In its simplest version, metacognition consists of a cognitive event and a monitoring process on it. For example, "you have likely had the experience of reading something, realizing that you have no memory for the last paragraph, and then having to re-read the passage" (Basile et al., 2014).

Self-consciousness is thus a kind of processing of conscious information relating to the self. Importantly, by gathering behavioural and neurophysiological types of evidence related to it, it is possible to build a solid hypothesis on which beings may experience this fascinating capability for which, again, Stanislas Dehaene (2014, p. 24) provides insightful words: "attending a concert or watching a gorgeous sunset can put me in a heightened state of consciousness without requiring that I constantly remind myself that *'I am in the act of enjoying myself.'* My body and self remain in the background, like recurrent songs or backdrop illumination: they are potential topics for my attention, lying outside my awareness, that I can attend and bring into focus whenever needed. In my view, self-consciousness is much like consciousness of color or sound. Becoming conscious of some aspect of myself could just be another form of conscious access in which the information being accessed is not sensory in nature but concerns one of the various mental representations of "me"– my body, my behaviour, my feelings, or my thoughts".

To sum up, in this section we have seen different but overlapping concepts of consciousness, the relationship between them, and some ways in which they may or do shape behaviour. The rest of this chapter will build on this, to deal with strategies

[12] The 'theory of mind' refers to the cognitive mechanisms by which it is posible to respresent others' mental states (see the entry on theory of mind and consciousness in The Oxford Companion to Consciousness [Bayne et al., 2009]).

for inferring conscious access of first order states related to emotion, by taking advantage of different types of evidence.

3. Consciousness and emotional responses: changes in emotion are not necessarily accompanied by related conscious feelings

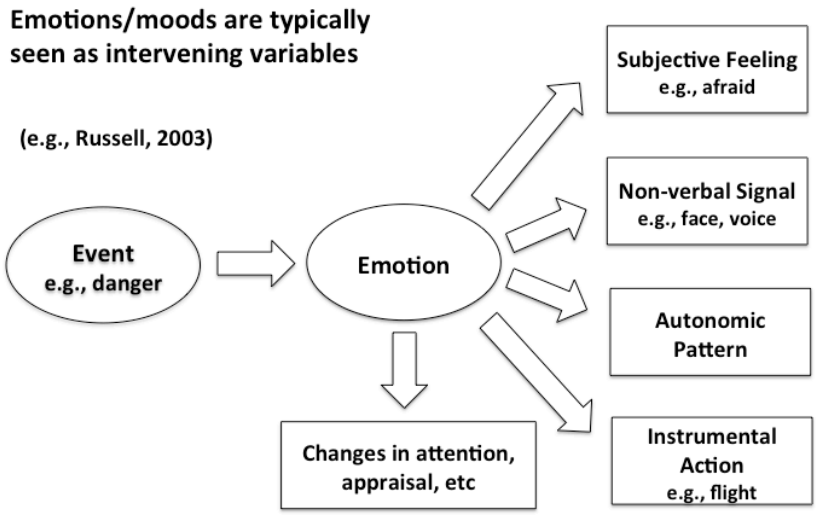
As we saw in Section 1, emotions are valenced states that, at least in verbal humans, can be experienced consciously. In the first part of this section we will briefly review several experimental findings consistent with the hypothesis that emotions can also be processed non-consciously. Importantly, this hypothesis fits into a broader view of consciousness that different types of information or responses often thought to depend on conscious processing can actually be non-conscious (see Section 2). In the second part of Section 3, we review cortical areas that are likely to play a very important role in conscious emotion, at least in adult verbal humans. In the third part of Section 3 we then review several specific types of evidence related to states of emotion, and their potential usefulness (or lack of usefulness) for inferring conscious emotions.

3.a. Empirical evidence for non-conscious emotions: an introduction

In Berridge and Winkielman's influential paper 'What is an unconscious emotion? (The case for unconscious liking)' (2003) these researchers review their own empirical findings to conclude: "positive and negative affective reactions can be elicited subliminally, while a person is completely unaware of any affective reaction at all (in addition to being unaware of the causal stimulus)". Thus in their view, the term "unconscious emotion" can refer to two quite distinct processes: emotional responses that occur with no conscious awareness of the eliciting stimuli (e.g. no awareness of the relevant reinforcer, if we follow Roll's view that emotions are states associated with reinforcement); or emotional responses that occur without a self-reportable 'felt' component. It is this last type of process that we focus on here, since the presence or absence of this subjective component of emotion is our topic of interest, and also the ethically relevant issue for animals. Rusell (2003) also portrayed

the dissociation between feelings and other aspects of emotional responses in a useful diagram, redrawn as Fig. 1 (below).

Figure 1 – The dissociation between feelings and other aspects of emotional responses (Russell, 2003)



Below we will see that the earliest evidence that responses to aversive, harmful stimuli may not require emotional feelings came from studies of nociception versus pain, especially revealing (although brutal) animal studies by Charles Sherrington on decerebrate cats, conducted in the late 19th and early 20th centuries (see Sherrington, 1906). Nociception is typically defined as the detection of noxious (i.e., potentially harmful and therefore negative) stimuli, and subsequent responses to them. These aversive stimuli are detected by specialized neurons (specifically types A, C, and silent) known as nociceptors. Nociceptors initially process and relay this type of information to other structures of the animal's nervous system, which further contribute to processing and eventually elicit a behavioural response (see Marchand, 2008; National Academies, ILAR, Pain Working Group, 2009). In verbal humans, and possibly other animals, the CNS typically processes this nociceptive information in ways that generate the conscious experience known as pain (see, Allen, 2004;; National Academies, ILAR, Pain Working Group, 2009): "an unpleasant sensory and emotional experience associated with actual or potential tissue damage, or described

in terms of such damage" (IASP, 1979). Although nociception and pain are very closely related in conscious beings, they can be dissociated (see Shriver, 2016b). Thus, in some cases humans can feel pain without nociception, i.e. in the absence of potentially noxious stimuli, as in for instance, the case of 'neurogenic pain'[13] (e.g. phantom limb pain, see Marchand, 2008; Ramachandran & Hirstein, 1998). Conversely, verbal humans can process nociceptive information without experiencing it consciously: a dissociation between nociception and pain that can be found even in humans with intact CNSs. For example, a withdrawal response to a noxious thermal stimulus – like getting burned by grabbing a very hot cup in the kitchen – is a reflex that occurs non-consciously, typically so rapidly that it *precedes* the conscious feeling of pain (in both its sensory and affective nature) (James Grau, personal communication).[14] Human and nonhuman animals with damaged brains or spinal cords can also emit nociceptive responses, even when information cannot be normally processed by their CNS (something that – as we will see below – is assumed essential for consciousness), including, according to the ILAR Committee on Recognition and Alleviation of Pain in Laboratory Animals (National Academies, ILAR, Pain Working Group, 2009), "the withdrawal of body parts (e.g., limbs, tails) from noxious stimuli occurs in decerebrate cats (Sherrington 1906), and spinally-transected cats and rats in which connections to the brain are severed (e.g., Grau et al. 1998)". Thus, we can see that some simple avoidance responses elicited by harmful stimuli do not require, or even need to correlate with activity in parts of the body that we normally associate with the capacity for conscious experience (an intact brain, cerebral cortex or neocortex).

Much more recently, several studies have yielded results consistent with other states of non-conscious emotion in verbal humans, including non-conscious *positive* emotions. Among these studies, Piotr Winkielman and colleagues have yielded one of the most interesting experimental bodies of work (e.g. Berridge & Winkielman,

[13] Neurogenic pain is defined by the IASP as pain arising as a direct consequence of diseases affecting the somatosensory system (see Treede et al., 2008).

[14] Interestingly, "research…has shown that the [conscious] affective dimensions of pain can be manipulated independently of the [conscious] sensory dimension, and that two distinct activation patterns during brain imaging correspond to the different dimensions" (Shriver, 2016b).

2003). They suggest that "unconscious liking" can be subliminally induced in humans, with this non-conscious state then influencing subjects' subsequent behaviour. For example, they found that when thirsty humans were given a pitcher of a fruit-flavoured drink immediately after being exposed to subliminal cues of different valence (pictures of human happy, neutral, or angry facial expressions), they would pour and drink about 50% more after being presented with the positive cues than after the neutral ones. In contrast, when exposed to subliminal angry pictures, subjects would pour and drink less than when presented with neutral pictures (Winkielman et al., 2000). Crucially, subjects did not report changes in their emotional states (assessed immediately after the subliminal exposure, using a self-report hedonic scale). In a second experiment (Winkielman et al., 2000), the researchers further found that when thirsty subjects were presented with the same subliminal cues as in the first study, immediately after being given a sip of the fruit beverage they were willing to pay almost double for a can of this drink if they had been exposed to the subliminal happy expressions compared to if exposed to subliminal angry ones. Moreover, subjects gave higher ratings to the question "how much of this drink do you want to drink right now?" (despite no changes in self-reported mood).

As a final example, work by Zemack-Rugar and colleagues (2007) showed that non-consciously perceived adjectives also affect behaviour even when no subjective emotional change is reported by participants. Subjects were subliminally presented with negatively valenced but qualitatively different emotional cues – those of guilt (e.g. blameworthy, culpable), and sadness (e.g. sad, miserable) – which affected subjects' behaviour in 'emotion-specific' fashions'. This study was inspired by research showing that when these emotions are conscious, guilty and sad people tend to adopt different behaviours, with respect, for example, to self-indulgence and helping others in unpleasant tasks: individuals experiencing sadness tend to be attracted to immediately gratifying or tempting stimuli, increasing their consumption of a host of indulgent products, but they avoid unpleasant helping tasks. In contrast, individuals experiencing guilt "tend to avoid indulging, as it is incongruent with the experience of blame or fault and can be perceived as a self-reward…and seek to punish or deprive themselves" (Zemack-Rugar et al., 2007). Consistent with this, Zemack-Rugar and colleagues found that participants subliminally presented with adjectives related to guilt subsequently showed lower indulgence and more helping

behaviour than did participants subliminally primed with adjectives related to sadness, even though they reported no differences in their subjective emotional ratings.

These three types of example, in which reflexes are induced, or motivated behaviour is influenced, by validated positive or negative stimuli despite no changes in the subjective component of emotion, all took advantage of the capability of humans to self-report to evaluate whether these hypothetical states of emotion were conscious or non-conscious. Their results epitomise the nature of unconscious or 'non-conscious emotion', as defined in complementary ways by different researchers: "an affective reaction of which one was simply not aware, even upon introspection" Kihlstrom (1999); the unintentional, automatic, and relatively effortless control of one's exposure to, processing of, and response to emotionally evocative events driven by goal pursuit (Williams et al., 2009); and Berridge and Winkielman's own criteria for non-conscious emotions: "people must not be able to report their emotional reaction at the moment it is caused. Yet there must be clear evidence of the emotional reaction either in their behaviour, or physiological response, or subsequent subjective impressions of an affect-laden event".

Despite not being felt, these unconscious emotions do still involve the brain. Berridge and Winkielman (2003) argued that "subcortical brain circuits are truly affective, but only as unconscious core processes contained within ordinary emotion". Recent work by Marco Tamietto and Beatrice de Gelder (2010) reviews current findings on the neural bases of the non-conscious perception of emotional signals in verbal humans to draw somewhat similar conclusions. According to these researchers, "many emotional stimuli are processed without being consciously perceived", and "subcortical structures have a substantial role in this processing". These conclusions are based on neuroimaging studies using the techniques of backward masking[15] and binocular rivalry[16], showing that non-consciously perceived

[15] According to the technique of 'backward masking', a subject is presented with a stimulus (e.g., a visual stimulus) for a short time, which is followed by the presentation of a masking stimulus (e.g., another visual stimulus) for a longer time. This latter masking stimulus typically prevents subjects from consciously processing the target stimulus (see Wiens & Öhman, 2007).

[16] Binocular rivalry refers to a perceptual phenomenon that occurs when very different visual patterns are presented to each eye simultaneosly. In normal vision,

emotional stimuli elicit activity in the amygdala, superior colliculus, basal ganglia and pulvinar, and that activity in such subcortical structures is often enhanced compared with activity in response to consciously perceived stimuli. These researchers also suggest that because the non-conscious processing of emotional stimuli is accompanied by characteristic neurophysiological correlates that are often qualitatively and quantitatively different from those associated with conscious perception, the non-conscious perception of emotional stimuli should not be seen as a degraded counterpart of conscious perception but rather as a different mode of processing visual signals (rather as discussed in Section 2).

The rest of Section 3 will assess whether conscious emotion, in contrast, has a well-understood neurological basis (sub-section 3.b). Sub-section 3.c will then critically review some non-verbal types of evidence related to emotion, and yet which may not reveal whether states of emotion are felt.

3.b. The brain and conscious emotion

It has been argued that the "neural correlates of human consciousness include the presence of thalamocortical signaling, fast, irregular, low-amplitude electroencephalographic (EEG) signals, and widespread cortical activity correlated with conscious contents" (Edelman & Seth, 2009; see also Edelman, 2003). But, have we identified specific neurophysiological markers of conscious emotion? Tamietto & de Gelder (2010)'s review of neuroimaging studies of humans processing emotional stimuli consciously or unconsciously show that activity in neocortical areas such as the occipitotemporal, frontal or cingulate cortex is typically higher in response to emotional stimuli that are consciously perceived. This activity can even be suppressed under conditions of visual unawareness. They conclude that a major

---

the two eyes receive corresponding views of the world from slightly different perspectives, yet the visual system succesfully interprets and synthesizes them into a coherent, stable preceptual experience. Under a binocular rivalry paradigm, the brain proves incapable of arriving at a stable interpretation of the retinal input, with one eye's view dominating for a few seconds before being replaced by its rival from the other eye (see John R. Searle entry on Binocular Rivalry in the entry on priming in The Oxford Companion to Consciousness, Bayne et al., 2009).

difference between conscious and non-conscious types of processing in verbal humans may be the combined involvement of cortical areas and of cortico-subcortical interactions in the former.

Other works also indicate distinct patterns of brain activation under potential states of emotion: those that verbal humans report to feel. For example when we feel pain, the anterior cingulate cortex (ACC) activates (reviewed by e.g. Allen et al., 2005, Farah, 2008): evidence suggests that this structure of the mammalian neocortex "does play a central role in the experience of unpleasantness of pain in humans", and also "in learning to avoid noxious stimuli". Thus "patients…indicate [that] following ACC lesions that the intensity of pain remains, but that it is less bothersome" (Allen et al., 2005; see also National Academies, ILAR, Pain Working Group, 2009)[17]. However, as these researchers also stress, " the ACC is…just one part of a very complex system" and that "it is important not to place too much importance on any single chunk of neural tissue".[18]

In conclusion, current evidence suggests that conscious emotion in verbal humans and in other vertebrates is likely generated within the forebrain, and that in mammals, the cortex (or, potentially, analogous structures, such as the pallium in birds – see Seth et al., 2005; Butler and Cotterill, 2006; Edelman and Seth, 2009) plays a crucial role in processing this type of information.[19] It is important to

[17] Interestingly, lesions in the rat ACC are found to decrease the aversiveness of potential states of pain, despite the animals still responding to these noxious stimuli, a finding that may represent a behavioural version of the verbal reports from humans indicating that that the intensity of pain following ACC lesions "is less bothersome" (see Allen et al., 2006).

[18] Although some forebrain areas have been proposed as potential candidates for processing information consciously, no single brain structure or area has been identified as solely responsible for integrating information consciously (see, for instance, Crick & Koch's hypothesis on the role of the claustrum in the mammalian brain, 2005).

[19] In agreement with Colin Allen, however, it is important to bear in mind that "to say that these mammalian structures are required for pain is, of course, to beg an important question. Even if neocortical structures are required for mammalian pain experiences, it does not follow that they are required for fish" (see Allen, 2013).

underline, however, that the relationship between different patterns of brain activity and the self-reported (conscious) states of emotion must be better understood in order to reach more solid conclusions. As Adam Shriver puts it for the case of pain, "there is ample evidence from lesion studies – bolstered by single unit recording, direct stimulation, and fMRI – that the primary somatosensory cortex, the anterior cingulate cortex, and the insula cortex all play a central role in humans' typical experiences of pain and that interfering with their functioning will selectively impair aspects of painful experience…but lesion studies [in humans] indicate that there are no cortical areas that are always necessary for the conscious experience of pain" (Shriver, 2016a; see also Shriver 2016b). Investigations of patterns of brain activity related to different states of conscious emotion – the neurophysiological correlates of conscious emotion – are ongoing and, ultimately, they potentially could yield tools for assessing these states in absence of verbal reports (see Section 4).

3.c. Non-verbal types of evidence related to emotion and their potential usefulness for inferring conscious emotion

To parse out the correlates of conscious versus unconscious emotion, we will look at data from four distinct sources. The first and second of these build on the examples used in Section 3.a and 3b. These review respectively the emotion-like responses of intact verbal humans exposed to validated positive and negative stimuli that are influential despite inducing no reportable changes in subjective mood or emotion; and then the emotion-like responses of body parts that are disconnected, by some lesion, from a functional forebrain. The third set of data comes from subjects rendered unconscious (in the sense of losing the state of wakefulness, see Section 2) by anaesthesia; while the last comes from subjects with cortical lesions, or even without any cerebrum at all, exposed to positive and negative stimuli.

3.c.i. Correlates of unconscious emotion in verbal adult humans

Section 3.a introduced examples from masking studies and similar, in which the motivated behaviour (e.g. drinking; helping) of normal human subjects was influenced, by validated positive or negative stimuli, despite no changes in the subjective component of emotion. In other cases, stated attitudes and preferences toward neutral stimuli could be shifted towards more positive or negative depending on whether the neutral stimuli were accompanied by or paired with non-consciously processed stimuli (e.g. Tamietto & Gelder, 2010). Rather similarly, healthy subjects with 'virtual' cortical lesions generated by trans-cranial magnetic stimulation over their visual cortex, and clinical patients affected by so-called affective blindsight[20] are capable of discriminating emotional visual stimuli and correctly guessing the emotion displayed in images of facial expressions, even when these are presented in the blind portion of their visual field, and which importantly, they report they cannot see (de Gelder et al., 1999; Jolij & Lamme, 2005). It is relevant to note, however, that blindsight patients still do not spontaneously attempt to use this information practically or inferentially (see Section 2).

Startle reflexes are another behavioural pattern affected by states of emotion. In a typical startle procedure, subjects are exposed to positive, neutral, and negative stimuli – for instance, pleasant, neutral, and unpleasant pictures– and then to a sudden aversive stimulus such as a noise burst that elicits a startle response, such as an eyeblink (typically quantified via electromyographic responses (EMG) in the orbicularis oculi muscles; see, for instance, Vrana et al., 1988 & Lang et al., 1990). Negative stimuli (e.g. unpleasant pictures) typically potentiate this eyeblink startle reflex elicited by the sound or shock, while positive stimuli (e.g., pleasant pictures) inhibit it (see e.g. Lang, 1995, Reagh & Knight, 2013). Recent studies have investigated whether this even occurs when emotional stimuli are not consciously perceived. Data for positive stimuli are quite complex. Some findings unexpectedly

[20] Affective blindsight refers to the residual visual ability of patients with damage to the primary visual cortex (V1, striate cortex) to react reliably to the emotional valence of stimuli presented to their blind visual fields, and yet whose presence and properties they are unable to report (see de Gelder & Tamietto, 2007, Scholarpedia, 2(10): 3555).

suggest that non-conscious positive stimuli *increase* the startle response (Ruiz-Padial et al., 2011). However, all seem to suggest that positive emotional stimuli do not diminish the startle response unless subjects process them consciously (Ruiz-Padial & Vila, 2007; Reagh & Knight, 2013). In contrast, subliminal negatively valenced stimuli have been shown to significantly increase the eyeblink response just like supraliminal ones (Ruiz-Padial & Vila, 2007; Ruiz-Padial et al., 2011; Reagh & Knight, 2013). Thus, although the potentiation of startle by negative stimuli seems not to require conscious emotion, the obverse may be true, and more research is needed in order to elucidate whether this is the case. Finally, a recent study on startle responses found that humans can acquire fear learning elicited by an US (strong dyspnoea) when presented immediately after a previous CS (mild breathlessness, see delay conditioning in Section 2), even when they report being unaware of the relationship between these stimuli (Pappens et al., 2015). The finding that the startle tone used during the experiment elicited increases in startle eyeblink responses, was interpreted by the authors as suggesting that "startle eyeblink potentiation may reflect subcortical emotional learning".

Emotional facial expressions can also occur in humans with unconscious emotions: changes in facial expression are not necessarily accompanied by awareness of the states of emotion related to this behaviour, even in normal humans with intact brains. For instance, it has been found that people prevented from consciously perceiving pictures of validated happy, neutral, and angry faces – immediately followed and masked by neutral faces – react with distinct facial muscle reactions that correspond to the happy and angry stimulus faces, and crucially, rate them as neutral instead of happy or angry in a self-report questionnaire (Dimberg et al., 2000).

Other data from humans further suggest that even quite complex learned instrumental responses to achieve positive rewards can occur without subjects reporting any associated emotion. For instance, opioid dependents will self-administer very small doses of heroin by pressing a lever in an operant paradigm, even though they report not being able to feel these doses (Lamb et al., 1991; Comer et al., 2008).

Finally, turning to physiological responses, "in humans, specific types of incision or analgesic regime that reduce reported subjective feelings of post-operative pain, often leave patients' post-operative physiological stress responses undiminished" (reviewed by Mason, 2011). For instance, it is known that "in adult

humans, postoperative cortisol output is undiminished by analgesics that successfully treat the reported pain" (see Desborough 2000; National Academies, ILAR, Pain Working Group, 2009),

3.c.ii. Emotional responses that occur despite disconnection from the brain

As was described in Section 3.a., nociception and pain can be dissociated in humans. Furthermore, human subjects with spinal lesions that prevent nociceptive information from reaching the brain will still respond to noxious stimuli. For example, World War II veterans with a total transection of the spinal cord who had no sensation whatsoever in their lower limbs still responded with extension reflexes to noxious stimuli applied to their legs (Kuhn, 1951; see also Marshal, 1954; Hagbarth 1960; Macphail; 1998; Shriver, 2006). Likewise, in spinally transected cats, pinching or clamping the tail promotes stepping movements of the hindlimbs (Lovely et al., 1986), as though simple locomotory escape movements can occur even without pain (National Academies, ILAR, Pain Working Group, 2009). Additional research reveals that even the instrumental learning of avoidance responses is possible without involvement of the brain: spinally transected rats learn to keep their limbs withdrawn from a shock for longer periods of time if doing so will terminate the insult (Grau et al., 1998). If conscious emotion requires a brain, then all these somatic responses reveal behaviour patterns that do not require conscious emotion.

3.c.iii. Emotional responses in anaesthetised subjects

Subjects under anaesthesia can also emit responses of an apparently emotional nature. Thus, "physiological stress responses to tissue damage in humans and other mammals occur during surgery, despite subjects being under deep anaesthesia" (see Mason, 2011). For example, "fully anaesthetized rats can also learn some associations, for instance the pairing of a tone with electric shock" (reviewed by Mason, 2011; see also Pang et al., 1996). For instance, Weinberger and colleagues (1984) discovered that rats are able to learn to withdraw one of their hindlimbs in the presence of a tone when previously paired with an electric shock applied to that limb under anaesthesia. The finding that these stimuli did not elicit changes in brain

activity or consistent changes in heart rate was interpreted as suggesting that these subjects remained in deep anaesthesia. Interestingly, once awakened, pairing the tone with their water consumption times induced conditioned suppression of their drinking behaviour (Weinberger et al., 1984). As another example of classical conditioning during anaesthesia, mammals – such as rats and sheep – can acquire 'conditioned taste aversion'[21] even when under deep anaesthesia (Burešová & Bureš, 1979; Provenza et al., 1994). Although anaesthetized, these animals can still learn to associate palatable food given before undergoing anaesthesia with the effects of a poisoning solution (as demonstrated by the reduction of food intake once awakened). As Mason (2011) put it, "if fully-anaesthetized vertebrates can be assumed not to be conscious, we can use their reactions to this type of stimulus to identify responses that do not require conscious feelings". Thus, these types of responses *per se* are not sufficient for inferring states of conscious emotion.

3.c.iv. Emotional responses that occur despite a lack of cortices or even entire cerebra

As we saw in Section 3.a., which introduced the concept of unconscious emotion, decerebrate cats will withdraw their body parts (e.g., limbs, tails) from noxious stimuli. Sympathetic responses such as hypertension, and pupil dilation are often used to infer emotion, but these also occur in response to noxious stimuli even in decerebrate rats and dogs, revealing that they are not likely to require conscious emotion (Sherrington, 1906, see also National Academies, ILAR, Pain Working Group, 2009). Decerebrate rats and dogs also display tachycardia when presented with noxious stimuli (National Academies, ILAR, Pain Working Group, 2009); likewise, different vocalizations related to distress can be evoked in decerebrate and decorticate mammals when presented with tactile nociceptive stimuli (Newman, 1988); thus these responses are not likely to depend on conscious processing either.

[21] Conditioned taste aversion occurs when an animal associates the taste of a certain food with symptoms caused by a toxic, spoiled, or poisonous substance. Generally, taste aversion is developed after ingestion of food that causes nausea, sickness, or vomiting.

Crucially, however, this may not be the case for every type of vocalization. For instance, isolation calls – given by the infants of mammals when separated from their littermates or parents – seem to require an intact forebrain (Newman, 1988). Therefore, more research is needed with regard to these types of evidence.

Furthermore, anencephalic and hydroanencephalic infants (Luyendijk & Treffers, 1992; Steiner, 1979) are capable of changes in facial expression, often induced by being presented with emotional cues, such as sweet, sour, or bitter stimuli. Although it has been argued that these behavioural changes are evidence for conscious emotions (see Merker, 2007; Panksepp, 2005 for arguments consistent with this view), according to the evidence presented in Sections 3.a and 3.b, and to the main theories of consciousness (see, for instance, Baars, 2001; Edelman 2003; Dehaene, 2012), these humans lacking the cortex and often "possessing only a brain stem" (Winkielman & Berridge, 2004) may therefore not be capable of processing information consciously. Again, however, caution applies when interpreting these results: the neuronal footprints of states of conscious emotion must be better studied in order to reach more solid conclusions in this regard. Future studies on verbal humans undergoing conscious emotional versus non-conscious emotional tasks, and, as argued above, more brain-lesion cases in which subjects report deficits in feeling but not in emotion will probably shed additional light on this question.

3.d. Conclusions

We have just reviewed data supporting the hypothesis that emotions do not need to be conscious states. This view is in agreement with findings showing that much information processed by the human CNS is not experienced consciously. This is not surprising, considering the very limited capacity for processing information at a conscious level and, on the other hand, the vast amount of data that animals' CNSs have to deal with in order to keep homeostasis and enhance the chances of survival. Thus, as other types of information, conscious processing may not always be necessary when dealing with emotion related information that can be processed in a more automatized and faster way (see Section 2).

We have also seen behavioural, physiological and neurophysiological responses which are closely related to emotion, and yet do not require, or are not

likely to require conscious processing for their emission or production. Overall they reveal that many physiological responses to positive or negative stimuli, and even some behavioural ones, cannot be used to demonstrate the presence of first order conscious processes related to emotion.

4. How can we get solid evidence on whether any non-human animal experiences conscious emotion?

If a wide variety of commonly-measured emotional responses are not likely to require conscious emotion, and so cannot be used to identify sentient beings, what evidence can we use?

It is not known for sure. However, as for any other conscious process, having conscious emotions is highly likely to have functional effects that represent survival advantages for those conscious animals compared to non-conscious ones. In Section 2 we saw that empirical evidence from humans suggest that the functional role of consciousness possibly has to do with 1) generating integrated representations of the world; 2) predicting potential scenarios, and eliminating hopeless ones before even trying them; 3) making "novel bindings", and adding this information acquired through conscious integration to the catalogue of behavioural responses, that eventually can be triggered in absence of conscious processing. Crucially, we saw Dehaene's conclusion that this "*de novo* creation of unforeseen combinations" of information seems to require conscious processing. In the light of these potential functions, integrating cross-modal information (e.g., visual, interoceptive…) with components of emotion that, as we saw in Section 1, are related to scenarios of reward and punishment, probably represents an extremely adaptive tool. In agreement with this view, Williams and colleagues (2009) argued that the "conscious processing of emotions might be adaptive since by implementing the catalogue of emotion related responses, animals increase the probability of survival". Writing about pain, Patrick Bateson argued similarly in 1991 that a function of pain might be that of inducing long-term memory coupled with learning, in order to avoid situations that gave rise to the original pain experience. Reviewing several recent studies, Mason (2011) also argued that "the function of conscious affective states in homeotherms is controversial, but many argue that they mediate certain forms of flexible behavioural

decision-making, such as motivational effects (modifying the strengths of competing appetitive behaviours according to their relative costs and benefits), and types of goal-seeking/harm-avoidance that require innovation or planning". Allen et al. (2005) likewise suggested that specific kinds of learning "seem to be closely correlated to conscious awareness, for example…more sophisticated kinds of operant learning as opposed to simpler forms of instrumental learning". Thus, these researchers think that conscious emotions must have functional effects, and these effects in turn must be measurable.

In the light of these considerations, below we propose several strategies that might provide useful tools for identifying states of conscious emotion in the absence of verbal reports. These tools must be validated as processes dependent on, or correlating with conscious processing in humans, before designing studies adapted to each species' characteristics in order to look for similar pieces of evidence. An important consideration here is that, even if we could find such biomarkers in verbal humans – i.e., gold standards – validating these tools necessarily depends on verbal self-reports (a form of self-consciousness that requires the use of symbolic language, see Section 2). However, this does not necessarily imply that these pieces of evidence would *only* indicate conscious higher order processes, but that humans have to process the emotional information in this way in order to verbally report it (an issue I will return to in Chapter 5). Finding this evidence in nonverbal beings would *at least* reveal conscious access in the form of primary consciousness to states of emotion.

A second relevant consideration has to do with evolutionary homologies and analogies. When trying to infer states of conscious emotion in the absence of self-reports, the degree of usefulness of any non-self-report based type of evidence is firstly determined by the degree of phylogenetic proximity between humans and subjects of other species. Here, the concepts of divergent and convergent evolution, and homology and analogy play a crucial role. When evaluating the cases of non-verbal beings, the closer in phylogeny to humans a subject is, typically the better inferences we can make on whether he or she is experiencing a conscious emotion. Animals of species close to humans in phylogeny typically share anatomical, physiological and behavioural traits, which enable us to make the best possible inferences of whether they may also process their states of emotions consciously. Thus, certain behaviour and patterns of brain activation related to states of emotion

can be more easily compared to those observed in verbal humans when experiencing similar states that, crucially, they report to feel.

As we consider the case of animals progressively more distant in phylogeny from humans, we find fewer homologies and more analogies. This is a consequence of convergent evolution, a process in which two distinct lineages can evolve similar characteristics independently of one another, as a consequence of facing similar environmental challenges and selective pressures. That is, traces not inherited from a common ancestor. Even though vertebrate and invertebrate animals diverged more than 500 million years ago, during the Cambrian Explosion, they were able to evolve analogous anatomical and functional characteristics, nd according to some theories invertebrates may also undergo states of emotion (Adolphs & Anderson; 2014; Rolls, 2014). However, making hypotheses on whether they have conscious emotions is much more difficult in these cases. These animals' nervous systems, physiology, and behaviour are often very different from those of humans and other vertebrates, and thus comparing the types of evidence associated to conscious emotions in verbal humans to those observed in these animals when undergoing states of emotion is much less informative. This implies that the hypotheses we can make on whether these animals are capable of conscious experiences, and eventually conscious emotions, are typically weaker than in the case of vertebrates. Thus, the degree of usefulness of any type of evidence potentially dependent on conscious processing in verbal humans is strongly determined by the question of homology and analogy. With these considerations in mind, next we will see some strategies that may provide us with these pieces of evidence related to conscious emotion.

4.a. Using of symbols to self-report states of emotion

Since verbal self-report in humans depend on the conscious processing of information, the capacity for accurate report has been proposed as a method for assessing conscious experiences (Seth et al., 2005; see also Rolls, 2014). It has been argued that "the capacity for some form of vocal learning is shared by at least six animal groups, including cetaceans, bats, parrots, songbirds, hummingbirds, elephants, and possibly even mice and some other rodents" (Edelman & Seth, 2009). There are many examples showing that chimpanzees can be trained to use symbols

related to human language (Gardner et al., 1989; Savage-Rumbaugh et al., 1994; see also Rivas, 2005 for a different viewpoint), and even animals belonging to species phylogenetically more distant from humans also seem capable of similar tasks. A well-known example is that of Alex, an African grey parrot who "was able to name objects, having acquired vocabularies roughly equivalent to those of some language-trained chimpanzees (albeit after years of training and reinforcement)" (Edelman & Seth, 2009). Alex named objects in categorization paradigms, and produced accurate reports for discriminations he made. For instance, "when presented with an altered array of objects, seemed able to make a judgment to the effect that – 'I know that something in this perceptual scene has changed, and here is what has changed'" (Edelman & Seth, 2009). Thus, some animals could be trained to associate symbols to different emotions, and potentially use them to self-report about these states.

It has been correctly pointed out, however, that these types of evidence could be the result of combinations of multiple operant associations between those symbols and the related outcomes during training, without requiring any conscious processing (see, for instance, Sara Shettleworth's Cognition, Evolution and Behavior, 1998). As argued by Shettleworth (1998), "no other species (than humans) communicates naturally in nearly such an elaborate way", a consideration related to the question of to which degree these nonhuman animals "actually understand all that they can produce and vice versa". Still, even if non-animals could use these symbols in the way adult humans do, what we now know regarding conscious versus non-conscious learning from humans (see Section 2) suggests that just learning to use these novel symbols accurately may require access consciousness. Again, as Seth and colleagues put it (2005), in humans "there is very little evidence for long-term learning of unconscious input. In contrast, the evidence of learning of conscious episodes is overwhelming. Even implicit learning requires conscious attention to the stimuli from which implicit regularities are (unconsciously) inferred". Thus, this approach may not tell us that much about higher order conscious capabilities, but would represent a potential way to assess states of emotions related to primary consciousness. Importantly, below we will see that this paradigm can be more easily implemented by using operants. As well as not conflicting with the controversial question of whether non-human animals can acquire any aspects of human language – perhaps being capable of human-like forms of higher order consciousness – this instrumental approach is a more practical tool for at least investigating forms of

primary consciousness (a first order process) related to emotion in a much wider range of species. I will argue that if humans cannot discriminate and generalise between similar emotional states induced by different means unless they report being consciously aware of these states, we would then have a consciousness-dependent behavioural process for investigating states of conscious emotion in non-verbal beings. But first I will comment on the potential use of the 'cognitive bias technique' as a biomarker of states of conscious emotion

4.b. Cognitive bias tasks, and other uses of discriminative stimuli

The technique of "cognitive bias" is based on findings showing that humans have increased expectations of bad or good outcomes if in respectively negative or positive affective states (Harding et al., 2004; Mendl et al., 2009). In the past years it has been used in studies on nonhuman animals by first training subjects to discriminate stimuli on the same modality (e.g., two different odours) associated with either positive or negative outcomes in a 'two alternative forced choice' (2AFC) operant task,[22] and afterwards presenting them with ambiguous stimuli (e.g., different

[22] Experiments on cognitive biases are typically based on operant 'two alternative forced choice' paradigms (2AFC). A 2AFC typically sets a decision making condition in which a choice must be made between two responses based on limited information about which is correct – that is, which will be rewarded (see Bogacz et al., 2006). Unlike classical and simpler forms of instrumental conditioning (see Section 3), we do not have evidence that humans can display this type of behaviour when the relevant information is not consciously processed by the brain; moreover, decerebrate or spinally transected mammals can only perform simpler types of operants. Importantly, if one of the functions of consciousness is to blend novel information into a conscious state related to punishments or rewards, learning to succeed in any 2AFC paradigm potentially qualifies as a process dependent on conscious processing. Obviously, this is not to say that 2AFCs can be only performed if information is consciously processed, since absence of evidence is not evidence of absence, but the possibility that these types of operants (or "purposeful behaviour", as argued by Allen et al., 2005) depend on conscious processing remains to be investigated. Thus, finding out whether humans must be aware of their states of

mixtures of both odours) that must be discriminated according to the previous training (see Section 2.d. in Chapter 3 for an in depth explanation). Importantly, prior to this second phase researchers can manipulate subjects' states (e.g., inducing a potential negative emotion in bees by shaking them vigorously to simulate the state produced by a predatory attack on a colony, see Bateson et al., 2011) to therefore predict animals' interpretation of the ambiguous stimuli according to their hypothesised emotional states. Interestingly, these studies discovered that subjects of different species – dogs (Mendl et al., 2010); rats (Brydges et al., 2011); starlings (Bateson & Mateson, 2007); and honeybees (Bateson et al., 2011) – display operant choices consistent with cognitive biases when discriminating the ambiguous stimuli. Cognitive biases are thus a potential behavioural tool for inferring emotions, but do they tell us something about whether subjects are aware of these emotions? It has been argued that these responses do not necessarily reflect a change in emotion and may instead reflect a change in attention (see Giurfa, 2013 for the case of honeybees). Importantly, attention does not always imply conscious processing, even in verbal humans (Koch & Tsuchiya, 2006). Thus, these biases *per se* are not sufficient for inferring conscious emotions in these animals. As Mendl and colleagues (2011) put it, "biased decision-making under ambiguity may…reliably reflect the valence of an animal's affective state, but the question of whether (and which) animals have an actual awareness of such states (conscious emotions) remains open". Future studies on verbal humans may shed light on this question by assessing whether these biases are dependent on the subjects' conscious awareness of the emotions relates to this behaviour. If this were the case, we could then use these behavioural markers of conscious emotions to better interpret whether non-verbal beings displaying cognitive biases are aware of their emotional states. In other words, if verbal humans need to be aware of the emotions related to these cognitive biases in order to display them, we would have good reasons for supporting the hypothesis that any animal exhibiting these biases must be capable processing emotional information consciously.

Even if humans could display cognitive biases without consciously processing the relevant states of emotion, this hypothetical finding could still shed some light on

---

emotion in order to use them as discriminative stimuli (DS) in operant 2AFC paradigms may shed additional light on this question.

whether non-verbal beings are capable of conscious emotion. This could be investigated by 1) assessing whether the cognitive biases (responses) displayed by verbal humans differ in strength according to whether or not they report being aware of the states of emotion related to these biases, and 2) by additionally assessing whether masked emotional stimuli induce no biases at all (i.e., the stimuli are not shown long enough for subjects to attend them). If this were the case, similar studies could be performed in nonverbal animals in order to evaluate whether these beings display similar levels of biases according to whether the emotion-inducing cues are presented for longer or shorter periods of time. Although these results may be explained by different levels of attention which may not require any conscious processing, results from humans would strongly suggest that at least one of these levels would depend on the conscious processing of related states of emotion.

4.c. Discrimination and generalisation of states of emotion

We have just seen that investigating whether emotions can be only used as DSs in an operant 2AFC paradigm if subjects are consciously aware of these states may be a way for finding behaviour related to emotion dependent on conscious processing. However, even if using emotion as DSs may not necessarily require conscious processing, generalising from these cues to similar states of emotion induced by different means may depend on conscious processing (see Chapter 3 for an in-depth review of these paradigms). It is well known that both humans and many other animals are capable of these operant discriminations and generalisations. Furthermore, humans and animals such as rats, mice, and pigeons, perform similarly when discriminating the effects of different drugs – some of them related to emotion – in operant discrimination paradigms, and generalising them to drugs that induce similar effects but not necessarily through similar mechanisms of action (see, for instance Lal and Yaden, 1985; and Chapter 3 for an in depth review of these experiments). Additionally, some studies suggest that animals such as rats and pigs are capable of generalising the effects of drugs to similar effects or states not induced by drugs (see, for instance Gauvin & Holloway, 1991; Carey & Fry, 1993; and Chapter 3). Thus, finding out whether humans only generalise from emotions used as DSs to similar states if they are consciously aware of these states may provide us with

a potential window into the world of non-verbal animals' emotions. If these operant tasks require conscious processing of emotions in verbal humans, we would then have good reasons for thinking that this may also be the case in non-verbal beings yielding similar results in discrimination and generalisation tasks, thus representing a self-report paradigm on conscious emotion not necessarily dependent on any kind of symbolic language (see Chapter 3, Section 4).

4.d Patterns of brain activity as biomarkers of conscious emotion: using mobile brain imaging devices

An alternative approach considers, not only the likely cognitive and behavioural abilities supported by conscious emotion, but also the neurological machinery related to these states (see Dawkins, 2015). As we have seen in Section 3.b., neurophysiological studies in verbal humans are starting to elucidate patterns of brain activation related to different states of conscious emotion. The identification of these patterns of activity represents a potential tool for the assessment of similar states of emotion in absence of verbal reports. Crucially, the concurrent use of *non-invasive* mobile brain imaging devices with behavioural paradigms aimed at investigating conscious emotion may shed more light on this question.[23] As example of this approach – although not related to emotion – is that of complementing the binocular rivalry paradigm with brain imaging techniques (Edelman & Seth, 2009). In these experiments, monkeys were trained to press a lever to report perceived stimuli in a binocular rivalry paradigm, "neurons in macaque inferior temporal (IT) cortex showed activity correlated with the reported percept, whereas neurons in the visual area V1 instead responded to the visual signal". Edelman and Seth (2009) argued that "this suggests a critical role for IT in visual consciousness", and that "these observations are consistent with evidence from humans subjected to binocular rivalry", concluding that "this correspondence between monkeys and humans

[23] See, for instance, the 'Rat cap' non-invasive positron emission tomography (PET) in rats, Woody et al., colleagues 2004; and also novel mobile and non-invasive EEG devices (see De Vos et al., 2014 & Norton et al., 2015) that may be used on non-verbal humans and adapted to nonhuman animals for gathering neurophysiological information related to states of conscious emotion.

provides an example of how benchmark comparisons across humans and animal species can be made." Thus, this same approach could be applied in experiments related to emotion, such as the instrumental one described above, in order to make progress in finding non-verbal correlates of conscious emotion. Evidence showing that "the same brain structures are implicated in affective reactions for both humans and other animals [mammals]", such as "orbitofrontal and other prefrontal cortex, cingulate cortex, amygdala, nucleus accumbens, ventral pallidum, mesolimbic dopamine and opioid systems, hypothalamus, midbrain, and brain stem sites" (Berridge, 2003), suggest that this approach may be particularly useful for assessing potential states of conscious emotion in nonhuman mammals, and in other vertebrates with similar structures and neurophysiological functions. Furthermore, by identifying patterns of electrical activity correlating with conscious processing of emotion in verbal humans – e.g., the *P3 wave* for the case of visual information (see Dehaene, 2014, pp. 123-125) – we may, potentially, validate useful tools for identifying states of conscious emotion not just in subjects with homologous CNSs, but also in animals with very different nervous systems.

5. Conclusions

In this chapter I have argued that despite the capability of experiencing conscious emotion (such as states of suffering or pleasure) being a crucial trait for any being to be attributed with moral consideration, the assessment of this capability in the absence of verbal reports is not a straightforward task.

In Section 1 I introduced the concept of emotion. In Section 2 I reviewed different concepts of consciousness in both human and nonhuman animals with the aim of clarifying how they relate to each other, and their relevance when investigating the relationship between primary consciousness and behaviour. In Section 3 we saw that many non-verbal types of evidence typically used to assess states of emotion are actually not informative as to whether these states are processed consciously or not. There we argued that the main problem with these tools is that they are not solely dependent on conscious processing, and therefore are not sensitive enough for inferring states of conscious emotion.

Finally, in Section 4 I argued that, by taking advantage of validation through human self-report, several lines of research could potentially provide us with evidence of first order conscious processing of emotion. I underlined, however, that the usefulness of these strategies is strongly determined by the degree of phylogenetic proximity between humans and subjects of other species. I agreed with Edelman and Seth in that experiments assessing accurate report could yield this type of evidence. However, there I argued that taking advantage of the potential capabilities of some nonhuman animals for some aspects of human language (verbal report by interpretation and use of symbols-words) might be problematic. Although these studies may potentially yield evidence on forms of higher order consciousness related to emotion, the question of whether these subjects use symbols in ways homologous to human language is still controversial. Alternatively, I proposed that the use of operant paradigms avoids this question, while potentially informing us about the capability for primary consciousness of emotional contents. Furthermore, I argued that this technique could be implemented in an extremely wide range of animal species. Thus, I proposed that experiments on 1) cognitive biases, and 2) discrimination and generalisation between different states of emotion – which typically use operant discrimination paradigms – may be potential type-c processes related to emotion. We also saw that the validation of patterns of brain activity related to different states of emotion in humans represents a potential tool for making progress on this topic. Furthermore, I argued that by investing these pieces of neurophysiological evidence together with behavioural paradigms that are likely to depend on conscious information processing we could come up with better tools for identifying states of conscious emotion in the absence of verbal reports.

Chapter 3

Conscious processing of discriminative stimuli

1. Introduction and overview

In this chapter I will first introduce discrimination learning and related concepts, and present relevant terminology from the operant literature. Focusing first on external cues, or 'exteroceptive stimuli', I will briefly outline the importance of discrimination learning in animals' everyday lives and in training animals to perform useful tasks for humans, before covering discrimination training and generalisation as valuable research tools for biologists and psychologists interested in animals' perceptual and cognitive worlds. I will then show how interoceptive stimuli, i.e. cues from within the body, can also act as discriminative stimuli (DSs). There are many examples of this. Interesting from our perspective are those involving affective states, such as pain or anxiety (which are highly relevant to ethics and animal welfare). However, better studied and more revealing are those involving human subjects, and recreational drugs or potential drugs of abuse, because such studies often involve the assessment of subjects' awareness of the drugs used as DSs. Finally, I will discuss whether, in order to use states like pain or anxiety as DSs, animals must process those affective states consciously.

2. Discrimination learning and operant behaviour

2.a. Operant studies of discrimination learning and some relevant terminology

In the psychology of animal learning, discrimination training is a technique in which subjects – either human or non-human animals – are trained to give a particular response in the presence of one stimulus or set of stimuli, and a different response, or the withholding of a response, in the presence of different stimuli. Discrimination experiments thus generally use operant or instrumental techniques. Operant behaviour can be defined as any behaviour whose frequency is determined primarily by its history of consequences, with these affecting the frequency of similar responses emitted in the future under similar conditions (Glennon and Young, 2011). In operant terminology, during 'discrimination training' subjects are reinforced to do a particular action by being provided with something valuable each time they perform this action in the presence of a particular predictive cue. This valuable outcome, known as a 'reinforcer', is either a reward ('positive reinforcer', e.g., food they like, money, etc), or the avoidance or removal of an aversive event ('negative reinforcer', e.g., termination of an electric shock). As a result of discrimination training, subjects will be more likely to perform the action in the presence of a stimulus (termed S+) signalling reinforcement than in the presence of a stimulus (termed S-) signalling the absence of a reward, a lesser reward, or a punishment[24]. The specific S+ and S- stimuli that subjects identify in order to make the required response are called 'discriminative stimuli' (DSs). For example, a hungry rat can be trained to discriminate between two different odours by operating a lever in an operant chamber. If when presented with the first odour (S+) the rat receives food each time she presses the lever, she will gradually learn to associate these events, i.e. that performing this action in the presence of the odour will provide her with food. If the rat is not provided with any kind of reinforcement in the presence of the second odour (S-), even when pressing the lever, she will learn to discriminate between these two

[24] A punishment is an event that reduces the likelihood of the operant (targeted behaviour).

stimuli, and that pushing the lever will yield reward only when the first odour is present (see Morse & Skinner, 1958 for an example in pigeons). Occasionally subjects can also be presented with two different S+s, and reinforced to perform a particular behaviour each time one of these two DSs is present (see, for instance, Uchida & Mainen, 2003 for an odour discrimination task in rats). In this case, if the rat of the previous example were provided with two levers in the operant chamber, she would be rewarded for pressing one lever in the presence of one odour, and also reinforced for pressing the other lever when presented with the second odour. Gradually she would learn to press the correct lever in the presence of each S+. Operant responses can take a wide range of forms. For example, the hungry rats of my example had to press a lever to activate a food dispenser, but they could also be reinforced for turning in circles, rearing (i.e., standing on the hind paws), or scratching. Basically, any response can be used and be acceptable, as long as the animal can perform it and make the association.

An important term related to discrimination learning is 'stimulus control' (see Morse & Skinner, 1958). A stimulus is said to gain control over the instrumental response when this response occurs only in the presence of the DS. Therefore, DSs are those stimuli that gain stimulus control. This control is acquired through association with events (reinforcement) that occur immediately after response. Stimulus control can be thus defined as the degree to which responding occurs in the presence of a specific stimulus and does not occur in the absence of this discriminative stimulus. Think, for instance, of a pigeon in a Skinner or operant box being trained to learn that pecking the response key when lit in green is followed by a delivery of grain as reinforcement. Once the pigeon learns this relationship, it is said that the instrumental response – pecking, in this case – comes under control of the discriminative stimulus (the green light in the response key).

'Context specificity' is another concept related to discrimination training. Behaviour learned during the discrimination training is said to be context specific or context dependent. That is, the stimuli with which animals are presented become discriminative stimuli in the particular context the training takes place, but have less effect in other contexts (Gazit et al., 2005). A good example is a puppy that gives appropriate responses to cues – discriminative stimuli – in the training school where she was trained, but not in other contexts such as when out on a walk (McGreevy & Boakes, 2007). Another nice example comes from a 'cognitive bias' study (see

Cahpter 2 and Section 2.d.) on starlings. The birds were trained on a discrimination task with differential rewards, in which the shade of the background (light or dark) determined which of two covered dishes contained a food reward (S+) (Brilot et al., 2010). These concepts should help in understanding the next section, in which I will explain how animals spontaneously learn DSs within their environments.

2.b. The spontaneous learning of discriminative stimuli

Animals – both human and non-human – are constantly responding to DSs in their natural environments. This ability to learn to identify stimuli that indicate that performing a particular action will probably yield a desirable reward is a valuable tool for survival. Think, for example, of a racoon who learns that opening the lid of an organic waste bin (S+) in front of a certain house – so context specific to this place – will provide her with food. This animal has not been specifically trained to learn this behaviour. However she was able to learn to associate the action of opening the lid of that particular bin – an S+ with some characteristics such as odour, location, shape… – with obtaining food. She might also try to open the recycling bin placed beside the organic waste one, but as soon as her repeated efforts yield no reward – i.e. are not reinforced – this animal will lose interest in this particular bin. Not surprisingly the cats I live with – Talulah and Bruno – are also very good at spontaneously learning to use DSs. For instance, they recently started to associate having dinner with "Walter passing close to the kitchen at nightfall". When I returned to Canada, after two months without seeing them, their evening feeding time coincided by chance with nightfall and the cats spontaneously made this association without training. Since it is now autumn as I write, the unfortunate result is that every day they ask for their dinner a few minutes earlier. In instrumental terminology, the action "Walter passes close to the kitchen" is the S+ and their dinner is the reinforcement here. Since they ask for their dinner when I pass close to the kitchen when it is getting dark, but not when it is still light, we can infer that the DS "Walter walking close to the kitchen" is context specific to nightfall. This stimulus takes control over the cats' instrumental response "asking for dinner". Another very well known example illustrating how animals spontaneously learn DSs is that of Clever Hans (Der Kluge Hans): a 19th century horse once believed to have human-like skills in arithmetic, among other intellectual abilities. Clever Hans' apparent abilities had nothing to do with

mathematical skills but with his ability to interpret his interrogator's bodily posture and use it as a DS. For instance, when a question was posed to the horse, both interrogator and audience leaned forward very slightly, and when Clever Hans' reached the correct number of foot taps, they tended to straighten themselves very slightly. Hans had been able to learn to discriminate between these and other subtle changes in bodily postures, in order to emit an answer that yielded him a sugar cube as reward. Consistent with this, when the interrogator or audience did not know the answers, Hans was unable to answer even the simplest question (see Pfungst, 1911). Similar effects occur in humans. Think, for example, of an Olympic 400 hurdles runner already positioned in the stadium track. This athlete will surely interpret the starting shot as the S+ in order to start running. Here the reward is, of course, winning an Olympic medal. However, an identical sound probably will not elicit the same operant response in other contexts, so this behaviour is context specific to the stadium track. These are just a few "everyday" examples that illustrate how humans and other animals are constantly responding to different stimuli that they interpret as indicators of the likelihood that certain actions will yield rewards.

2.c. Training animals to work for humans by reinforcing particular DSs

Since nonhuman animals can be trained to use different cues as DSs, humans often take advantage of animals' capabilities to make them perform tasks that humans cannot or do not want to perform themselves. This is the case for 'working' and 'service animals'. For instance, since animals can detect some stimuli that humans cannot detect themselves some animals are often trained to use these cues as DSs and thereby perform a useful task for humans. To illustrate, dogs can be trained to emit an operant response – a particular vocal signal – that alerts their human companions that they are going to suffer an epileptic seizure. These animals undergo discrimination training in order to learn to identify seizure related cues in humans associated to human pre-seizure states (S+), and seem to be capable of predicting seizures even 45 minutes before they happen. Barking the alert signal in the presence of human pre-seizure associated DSs is reinforced with food, while these operant responses are never reinforced during non pre-seizure states (S-) (Strong et al., 1999 & Kirton et al., 2008). Similarly, it has been reported that dogs can also be trained to detect certain S+s associated with states of different sorts of tumours, such as odours

from human skin cells associated with melanoma (Pickel et al., 2004), cell-derived volatile organic compounds (VOCs) in urine related to prostate cancer (Cornu et al., 2011), or in exhaled breath of breast and lung cancer patients (McCulloch et al., 2006 & Ehmann et al., 2012). Thereby behaviour trained by using operant techniques may be of great value in helping with early diagnoses of these kinds of cancer.

More broadly, for thousands of years animals have been forced to work for humans by using methods based on instrumental discrimination training. Training horses to be ridden, for instance, often requires the use of operant conditioning. Horses are thus trained to discriminate between riding commands such as the action of pulling the reins (S+s) versus no commands (S-s), in order to perform an action or operant – to stop in this case. Traditionally these operants have been reinforced by the negative reinforcement when horses are presented with the S+s, and by inflicting some kind of punishment – more or less harmful – whenever S+s are not followed by the expected action. However, we have seen above that reinforcement can also consist in providing animals with a reward instead of with a negative reinforcement.[25] Animals used to take part in circus acts are also frequently trained by using discrimination training. Thus, individuals from many species, mostly mammals and birds, are trained by being presented with orders of different nature – verbal commands, physical movements, etc – (S+s) while forced to perform the desired behaviour (see McGreevy & Boakes, 2007). As in the previous example, these animals are also reinforced by negative reinforcement or by being provided with positive reinforcement (e.g. food) each time they perform the behaviour trainers want them to display.

[25] Relevant to this, it has been recently shown that positive reinforcement by using food rewards leads horses to display learned behaviour better than when negative are used (Sankey et al., 2010).

2.d. Using discriminative training to investigate how animals process exteroceptive stimuli

Investigating whether and how animals use certain DSs – i.e. how these stimuli take control over their operants or actions – is a valuable tool for understanding animals' perceptual worlds. In particular, by designing experiments based on discrimination training techniques we can study animals' sensory abilities, and investigate how they categorize stimuli. Since animals must distinguish between stimuli in order to use them as DSs, these experiments enhance our understanding of animals' worlds by providing us with information regarding which stimuli subjects can successfully tell apart. These experiments typically consist of two or three sequential steps: 1) *training*; 2) *test of acquisition*; and potentially 3) a *generalisation phase*.

During the training phase subjects are initially trained to acquire the discrimination of one given stimulus (S+) vs. another one (S-) as already described. At the end of this phase, subjects undergo a test of acquisition in order to evaluate whether they have properly learnt the discrimination (see Section 3 for an explanation of the acquisition and generalisation criteria using the operant drug discrimination paradigm as example). As one example, Kendrick and colleagues (2007) showed that sheep in a Y-maze are able to discriminate between pictures showing frontal views of 25 pairs of sheep faces, by associating one picture of each face pair (S+) with a food reward. Moreover, when they tested the sheep for retention of discrimination performance, it took up to 800 days for their performance to decline enough to become significantly poorer than levels at criterion. Thus, by discovering that sheep can discriminate between pictures of faces of different conspecifics, their impressive capacities for visual recognition were revealed.

As another example, octopuses showed themselves able to discriminate between different planes of polarized light (Moody and Parriss, 1961). The octopuses were successfully trained (using sardines as reinforcer) to attack when a watertight torch (with a polaroid filter) introduced into their tank emitted polarized light with a positive direction of the electric vector (S+), but not when the vector direction was negative (S-). Two groups (A and B) were trained to discriminate between horizontal and vertical directions of the electric vector and two groups (C and D) between

oblique directions at 45º or 135º from the horizontal, and to attack whichever vector was to be the positive for the group to which they belonged. These results suggest that octopuses' retinas have specialized receptors to detect polarized light, and thus are of great help in inferring how octopuses perceive their worlds. In a final example, revealing cognitive rather than sensory abilities, a border collie called Rico was found to be able to retrieve over 200 items upon request (mostly children's toys and balls): verbal labels he learnt to use as S+s via reinforcement with food or play whenever he brought the correct item (operant). Furthermore, if a novel object was placed among seven familiar items, Rico would retrieve it correctly when asked to fetch an item using a novel name, indicating that he could deduce the referent of a new word on the basis of exclusion (Kaminski et al., 2004).

Subjects who succeeded in acquiring such discriminations may also undergo a third step called a *test of stimulus generalisation or substitution*. This allows researchers to study the similarity of the DSs used during training, to novel stimuli presented to the subjects. The degree to which these novel DSs take control of the operant previously associated with the training S+s varies from strong to weak, depending on how similar subjects find these two different stimuli. Those novel stimuli that are typically treated as if they were one of the training DSs are considered to be similar from the subjects' point of view (Section 3.b.ii. for an explanation on generalisation criteria in drug discrimination experiments). This does not necessarily imply that these stimuli are identical: just that subjects interpret the characteristics of a particular stimulus as resembling one stimulus used during training more than the other. Thus, by assessing the degree of stimulus control subjects show for a new DS we can calculate the degree of generalization between training and novel pairs of DSs, and therefore assess how similar subjects find these two stimuli.

This generalisation from trained DSs to other stimuli gives us an additional way to explore animals' perceptual and cognitive worlds. For example, in a classic work by Herrnstein, pigeons were trained to discriminate between pairs of images containing or lacking certain elements, such as humans, trees, bodies of water, or even a particular person (Herrnstein & Loveland, 1964, Herrnstein et al., 1976). When these pigeons were then presented for the first time with novel pictures, they were able to perform almost as well as when classifying pictures used in the training. These findings indicate that these birds can construct categories – perhaps even preverbal concepts – from the different visual stimuli they process. In similar

experiments on honeybees, subjects were trained to discriminate between a set of impressionist paintings by Monet, and a set of cubist paintings by Picasso, by being rewarded with sugar each time they used one correctly as a DS. They could learn these tasks well, even with greyscale versions of these images. Furthermore, they could generalise reasonably well to novel exemplars: trained bees shown colour and greyscale versions of novel Monet-Picasso painting pairs never encountered before tend to perform the correct operant, suggesting that they had learned about the categorical structure of these two artists' paintings rather than the specific cues of single exemplars (Wu et al., 2012).

More recently, as described aleady in Chapter 2, a novel animal welfare assessment tool has been developed based on how animals generalise from S+s and S-s they have been trained to. This measure is called 'cognitive bias', and is based on findings that humans show increased expectations of bad or good outcomes if in respectively negative or positive affective states (Harding et al., 2004). Animal subjects are thus trained to discriminate between two DSs in the same modality and usually on some physical continuum (frequency, percentage mix of odours, etc). For example, honeybees can be trained to extend their mouthparts to a two-component odour mixture (S+) predicting a reward (e.g., sucrose), and to withhold their mouthparts from another odour mixture (S-) predicting either punishment or a less valuable reward (e.g., quinine solution). Once this discrimination is acquired, they are presented with new stimuli whose characteristics are part-way between the two DSs used in training – e.g. exposing bees to three novel odours composed of ratios intermediate between the two learned mixtures (Bateson et al., 2011). Animals undergoing negative affective states are expected to display tendencies (biases) to classify these ambiguous novel stimuli as more similar to the non-rewarded S- DS, while positive states are expected to cause animals to classify these ambiguous novel stimuli as more similar to the rewarded S+ DS: classifications inferred from which operant is controlled by these novel ambiguous stimuli. Data from different species including dogs (Mendl et al., 2010), rats (Brydges et al., 2011), starlings (Bateson & Mateson, 2007) and honeybees (Bateson et al., 2011) are consistent with this, revealing that the 'optimistic' generalisation from S+s to similar stimuli may indicate good animal welfare, and 'negative' generalisation (from S-s to similar stimuli), poor animal welfare. These are just a few examples from a vast literature that illustrates

how animals, by discriminating DSs, and generalizing from them, can potentially 'self-report' to us about their sensory, cognitive or affective worlds.

3. Interoceptive discriminative stimuli and the behavioural self-report of subjective internal states

3.a. Interoceptive discriminative stimuli

Section 2 illustrated how animals are capable of using *exteroceptive* cues – stimuli that come from animals' external environments – as DSs. Animals can also use *interoceptive* cues – stimuli that originate inside the body, such as hunger, dizziness, anxiety or even bladder pressure – as DSs. Lal (1979) defines these types of stimuli as biological events within the body that support discriminative responding. Just as for exteroceptive DSs, this learning may occur spontaneously as animals go about their lives, or it may be successfully trained in the laboratory by researchers. As one early example of the latter, Slucki and colleagues (1965) trained a group of five hungry rhesus macaques to use the states generated by a non-aversive inflation and deflation of a small latex balloon within the jejunum as DSs. These monkeys, on a daily 23-hour deprivation-feeding cycle, were able to learn that by pressing a lever when the latex balloon was rhythmically inflated (S+) they would be provided with sugar pills. Despite their state of hunger, they also learned that operating the lever in the absence of rhythmic inflation of the balloon (S–) would not yield any reward. In a similar study, Soldoff and Slucki (1974) trained a female rhesus monkey to operate a lever only when saline perfused into her urinary bladder reached a certain volume and generated a certain pressure (S+). She was maintained on a daily 22-hour food deprivation schedule, receiving a restricted regimen after the end of each experimental session, and was reinforced with food pellets each time she pressed the lever when S+ was present. Thus, as the researchers argued, under appropriate reinforcement contingencies, visceral events can come to serve as DSs. In another species frequently used in the lab, Bárdos and Ádám (1978) found that rats can also use states generated by electric stimulation of their gastrointestinal mucosa as S+s in an operant discrimination paradigm. Rats water deprived for 23-hour first learnt that

pressing the one lever in the Skinner box after a stimulation (S+) at intensities that generated visible effects on rats' behaviour – such as overt startle reactions and muscular contractions – would provide them with water. Interestingly, lower intensities initially only acted as DSs when paired with an exteroceptive light stimulus, but the acquired discrimination then persisted even when the light stimulation was withdrawn. In order to assess whether real visceroceptive and not exteroceptive stimuli were used as S+s in this latter case, two associated control experiments were performed. In one of these experiments the wiring between the intestinal electrode and the stimulating apparatus was disconnected, but the whole experiment was left otherwise unaltered. In a second one, the visceral loop was rinsed with a solution of 1% procaine in order to anaesthetize the mucosal receptors. The discrimination behaviour of the rats was sharply impaired in both cases, showing that rats were indeed using the states generated by the internal electrical stimulation as DSs.

Interoceptive states such as hunger can also be used as DSs. For instance, Corwin and colleagues (1990) trained rats in a two-lever food-reinforced discrimination paradigm, to press one lever when deprived of food for just 3 hours and the other when deprived of food for 22 hours. After establishing the discrimination, a generalisation phase followed in which drugs known to reduce food intake were tested for their ability to engender responding similar to that of recent food ingestion. If a food deprived rat responded as if food-sated after an intraperitoneal injection of cholecystokinin (CCK), d-amphetamine or dl-fenfluramine, it would suggest that the interoceptive stimuli associated with an injection of any of these drugs was similar to the interoceptive stimulus of recent food ingestion. Interestingly, CCK doses within the range reported to reduce food intake in 24-hour food deprived rats consistently induced responding like that seen after 3-hour of food deprivation, even in rats that had been deprived of food for 22 hours. In contrast, neither d-amphetamine nor dl-fenfluramine consistently led to responding on the 3-hour deprivation lever in these 24-hour deprived rats. The researchers concluded that rats can use states of satiety as DSs, and that interoceptive stimuli induced by intraperitoneal administration of CCK seem similar to those produced by food in the gut, supporting the hypothesis that CCK plays a role in the regulation of food intake by affecting feelings of satiety. On the other hand, these results also suggest that the discriminative stimulus effects of d-amphetamine and dl-fenfluramine

in food deprived rats are not similar to those of food in the gut, and that these drugs reduce food intake by mechanisms other than the production of interoceptive stimuli similar to satiety. Benoit and Davidson (1996) used a similar technique and reached similar conclusions for 2-deoxy-D-glucose (350 mg/kg), which was found to produce interoceptive sensory cues similar to the states following 24-hour food deprivation in rats.

Changes in blood pressure can also be used as DSs. Both healthy (Bennett & Lal., 1982a & 1982b) and hypertensive rats (Lal and Yaden., 1985) were successfully trained to discriminate the antihypertensive drug clonidine from placebo, and to generalise its effects to other antihypertensive drugs sharing the same mechanism of action (all hypotensive drugs that act through alpha-2 adrenoceptor mechanisms). However, only hypertensive rats could generalise the effects of clonidine to other clinically used antihypertensive drugs' DSs with different mechanism of action. These drugs represented four distinct classes of antihypertensive agents according to their mechanisms of action, and the subjects generalised them to clonidine in a dose-dependent manner. In the light of these results, and after comparing them to the findings on healthy subjects, and considering that clonidine effects are not only hypertensive, Lal and Yaden (1985) concluded that only hypertensive rats seem capable of using the state generated by reduced blood pressure as DS in a drug discrimination paradigm.

Finally, studies on 'metacognition' – the ability to monitor one's cognitive processes – (See Chapter 2, Section 2) can also be designed as discrimination paradigms using interoceptive stimuli as DSs, as recently argued by Basile and Hampton (2014). In these researchers' words, the sensitivity to internal signals, such as memory, can be evaluated in the same way as to external signals, such as a light or tone. Thus, when designing experiments on metamemory (a form of metacognition consisting of the ability to monitor one's memory processes), they suggest 1) creating a primary task in which memory varies; 2) setting-up a secondary task with contingencies that encourage discrimination: subjects are reinforced for responding in one way if the memory is strong and in another way if it is weak – the DSs in this paradigm; 3) to evaluate plausible alternative cues that might control behaviour. Basile and Hampton thus maintain that it is possible to infer whether subjects can monitor their memory in the primary task if they can use it as a discriminative cue in the secondary task.

Just as for exteroceptive DSs, we can thus see that studies using interoceptive stimuli as DSs allow us to identify internal stimuli that animals can tell apart. Via generalisation effects they also potentially allow us to identify stimuli that animals classify as similar to the DSs that they were trained on. Arguably the largest body of work along these lines focuses on psychoactive drugs as DSs (see Section 3.b), while some fascinating animal studies instead use negative affective states as DSs (see Section 3.c).

3.b. Psychoactive drugs as DSs

Drugs have been extensively used as DSs in discrimination paradigms. A typical drug discrimination (DD) paradigm involves a discriminative training procedure in which psychoactive drugs are used as DSs. Here, subjects are typically trained to discriminate between the effects induced by a particular dose of a drug, the "training drug", (S+), versus an alternative, reinforced vehicle or placebo condition (a second S+). Not discriminating accurately is never rewarded in this paradigm. Eventually, subjects can also be trained to discriminate different doses of the same psychoactive agent, and between different drugs in a two choice procedure. This paradigm can even be implemented using three or more choices, in which subjects are trained to discriminate between three or more different drug conditions.

Historically, these DD procedures were first employed in nonhuman animals: they were developed in the 1960s for studying the DS properties of drugs using rats, mice, gerbils, rhesus monkeys, squirrel monkeys and pigeons (Kamien et al., 1993), principally to characterize these agents' pharmacological actions. For assessing subjects' performance in each discrimination trial, a Fixed Ratio (FR) procedure was normally used. FR indicates the number of correct responses (e.g. lever-presses) required within the training regime for the subject to obtain a unit of the reinforcer (e.g., a food pellet). The logic behind this procedure is that requiring more than one correct response per reinforce increases the sensitivity of the discrimination procedure (over a FR1 schedule, where only one correct response is required to obtain a reinforce). The FR value is variable (i.e. different experiments have used different FRs), but most drug discrimination studies use FR10 schedules (i.e., one unit of

reinforcement is delivered after every tenth consecutive correct responses)[26] (see Solinas et al., 2005, for a review of the animal DD paradigm).

In the late 1970s several groups of researchers adapted this procedure for human subjects, to categorize substances according to the effects they induce, and to analyse their abuse potential (see Altman et al., 1977; Chait et al., 1984). Instead of the operant responses typically required of animals, the behavioural task that humans are typically set involves distributing points or tokens among different drug codes presented to represent the different drug options. The subjects allocate these depending upon how certain they are of the drug's identity, and are typically monetarily reinforced for each correct choice.

A verbal or written subjective self-report component (SR) is normally added, aimed at identifying which specific experienced effects of the drug parallel or are even relied upon during the learning and/or performance of the behavioural task (see Appendix Table 1). As we will see in Section 4, these self-report questionnaires provide researchers with a better understanding of the complex relationships between the drugs subjects discriminate and the interoceptive effects they induce. For instance, some studies show that subjects report different subjective effects while discriminating different doses of the same drug. This is the case in work by Duka et al. (1998a), which revealed that subjects' most reported subjective effect when discriminating ethanol from placebo was "light-headedness" at lower ethanol doses, but "changes in taste" at the highest dose[27]. (See Section 4.a for several studies investigating the relationships between operant DD and SR discrimination).

According to several researchers (e.g. Díaz & Velázquez 2000; Oliveto et al., 2002), results from operant DD tasks in human studies are very similar to those using nonhuman animals, validating the use of animals for generating human-relevant data.

---

[26] According to Solinas and colleagues (2005), this value is high enough that random or involuntary responses do not interfere with training and testing, but low enough that subjects can quickly produce the number of responses required. To reduce the number of incorrect responses, researchers usually require that incorrect choices reset the number of accrued correct choices, thus subjects must make 10 consecutive correct choices in order to achieve the reinforcement.

[27] The researchers interpreted this outcome as suggesting that the more salient stimulus at task detracts from the importance of the other cues.

Thus both animals and humans can only discriminate drugs that humans report to induce subjective effects (i.e., psychoactive drugs), and animals' performance in drug discrimination experiments is very similar to that of humans (see Díaz & Velázquez 2000, Oliveto et al., 2002). This theme of cross-species similarity will be revisited in both section 3.b.i and 3.b.ii, but to give some first illustrations here, drugs with similar DS properties in laboratory animals, i.e. drugs that animals generalise between (e.g., between different stimulants, or between different opioids) also tend to produce similar subjective effects in humans (e.g. Chait et al., 1984). Thus as Gauvin and colleagues put it (1992), "novel test drugs from pharmacological classes outside the training drug stimulus class have generally engendered saline appropriate responding" (in drug versus saline two choice paradigms). For example, Kamien and colleagues (1993) found that results from amphetamine studies show similar patterns of generalisation to other psychomotor stimulants in humans, pigeons, rats and rhesus monkeys**.** Furthermore, with the exception of monkeys, the 'potency ratios'[28] remained constant regardless of procedure or species. Generalisation patterns and potency ratios were also assessed for other drugs, such as opioids, benzodiazepines, caffeine, ethanol, THC, and nicotine, and remained constant regardless of procedure or species (Kamien et al., 1993). These researchers therefore concluded that psychoactive drugs seem to exhibit DS effects that are remarkably stable across species and procedures. To give just two more of many additional similar examples, the influence of the training dose on nicotine discrimination in humans is "strikingly similar" to findings in rats, as are the ways that subjects trained on low doses of nicotine generalise to high doses (Perkins et al., 1996); while the DSs effects of d-amphetamine in both humans and rats are similarly attenuated by the antipsychotic risperidone (Rush et al., 2003). As I will argue later (see Section 4), such parallels suggest that data from human subjects in such tasks can validly be used to help understand the DD responses of non-humans.

Regardless of species, some of these DD studies focus on acquisition (see Section 3.b.i), to identify which specific drugs, or which specific doses, can be used as DSs, and also what factors either prevent successful training from occurring or

[28] Potency ratios are calculated by dividing the lowest dose of generalisation test drug that caused >75% training drug-appropriate responding by the training dose – see next section on drug generalisation studies.

block successful utilisation of the discrimination once it has been acquired. Other studies also focus on post-acquisition tests of generalisation (see 3.b.ii), the aims here typically being to find out which drugs or doses will be treated as similar to the specific drug and/or dose used as the training DS. Examples of both types of experiment follow.

3.b.i. Psychoactive drugs as DSs: Studies of response acquisition and performance

The initial phase of any DD paradigm, acquisition, enables researchers to find out whether subjects can discriminate between the effects generated by different psychoactive drugs. Subjects are typically trained to discriminate the effects induced by a drug versus a placebo or another drug. Afterwards, subjects are tested to assess whether they acquired this discrimination (i.e. whether they are capable of successfully discriminating the different drug conditions). During acquisition (as well as generalisation – covered next), human subjects also typically undergo self-report tasks. As previously explained (see Section 3.b.), these questionnaire-based tasks try to capture which specific subjective effects of the drug parallel or may be even relied upon during the learning and/or performance of the behavioural task (see Appendix Table 1).

In the acquisition phase of both human and animal DD studies, subjects are typically tested at individual level. Humans are normally required to reach a discrimination accuracy of at least 80% when either making discrete choices between letter/icons representing the drug/placebo conditions or distributing points or tokens between these options. Furthermore, the different conditions are interspersed across testing trials, and subjects must typically succeed in a number of consecutive trials. For instance, in work by Kelly and colleagues (1997), subjects' DD accurate responding in a point-distribution task had to be greater than 80% in five consecutive sessions (with conditions presented in randomized order, and the same condition never administered on more than three consecutive trials). In a study by Lile and colleagues (2012) subjects had to press two buttons on a computer screen (each representing one of the two conditions tested), and were likewise required to reach at least 80% correct responding, for in each of four consecutive sessions (with the

conditions again administered in random fashion, and each presented at least twice every four sessions). As one last example of many, in research by Perkins and colleagues (1996), subjects again had to be correct in at least 80% of their choices of letter codes representing the drug or placebo condition, in at least six trials (three per dose) attained within ten or fewer testing trials (five for each dose).

In animal studies, the acquisition criteria are less standardized, and often more stringent than in humans. Subjects must again reach a given percentage of accurate responses per session in a number of sessions (with different conditions also interspersed). For instance, rats in one drug versus placebo DD paradigm were required to reach at least 90% of correct lever choice (FR10) per session in 9/10 consecutive sessions (Vivian et al. 1994), while in another (Gauvin and Holloway 1991), the acquisition criteria were again set at 90% or more correct choices, but per session over six consecutive sessions. Mantsch and Goeders (1998) required rats in a FR20 drug versus placebo paradigm to reach at least 85% of correct choices per session in ten consecutive sessions. Less stringent criteria were favoured by Shearman and Lal (1979): here, rats discriminating two conditions in a FR10 paradigm had to reach 10/14 (71%) presses of the correct choice lever per trial, and do so in four consecutive trials. As a final, unusual example showing rather weak acquisition criteria, rats in a FR10 two condition DD paradigm were required to display 10/20 presses at the correct choice lever per session for four consecutive sessions (Wood et al., 1989; more examples are given in Table 2).

To give some examples of successful animal DD acquisition, mice have been shown capable of discriminating the effects of 3,4-methylenedioxymethamphetamine (MDMA) – a drug of abuse with mixed stimulant, hallucinogen and affective effects – from placebo or saline (Fantegrossi et al., 2009); gerbils can discriminate between chlordiazepoxide, chlormethiazole, ethanol, and pentobarbital, despite these drugs belonging to a broad pharmacological class of CNS sedatives/hypnotics that share some similar stimulus effects (Järbe & Swedberg, 1998); and pigeons can readily learn to discriminate between d-amphetamine, pentobarbital and saline (Leberer & Fowler, 1977). Some interesting cases perhaps suggest that animals use the subjective effects of drugs as DSs. For example, Lal and colleagues (1978) discovered that rats can rapidly learn to discriminate small doses of haloperidol – an antipsychotic dopamine antagonist that does not normally induce subjective effects in humans – from saline, but only when pre-treated with amphetamine, a stimulant

dopamine agonist. This result suggests that these rats used as DSs the effects of haloperidol drug on the state induced by amphetamine. Similarly, naloxone is a pure opioid antagonist that, when unmanipulated, subjects find hard to discriminate. However, rats *can* discriminate this drug from saline if they are pretreated with morphine 24 hours earlier. If the morphine treatment is not given, in contrast, naloxone treated rats choose the saline lever. This suggests that it was only the state induced by naloxone on the effects previously induced by morphine that these animals could use as a DS (Lal et al., 1978).

Related to this, sometimes researchers perform 'tests of stimulus antagonism', with the aim of finding out whether the stimulus effects of the training drug can be blocked. Here doses of a specific receptor antagonist are administered in combination with the training drug. Newly synthesized agents can also be tested in order to determine whether they are antagonists of the training drug (Glennon & Young, 2011). For example, Colpaert and colleagues (1978) found that rats' ability to discriminate cocaine and d-amphetamine – both stimulants – was antagonized by the effects of the antipsychotic drug haloperidol. In this study, one group of rats was trained to discriminate injected cocaine from saline, while a second group was trained to discriminate injected d-amphetamine from saline. After acquiring this discrimination, haloperidol – a dopamine receptor blocking agent – was shown to be equally effective in antagonizing the cues induced by both in the first group, and d-amphetamine in the second group. These findings support the hypothesis that d-amphetamine and cocaine share at least some stimulus properties in rats. Similar procedures have been used in humans, and reveal that similarly, the antipsychotic drug aripiprazole antagonizes the DS effects of d-amphetamine (Lile et al., 2005).

As another example of stimulus antagonism, this time in birds, Herling and Winger (1981) trained pigeons to discriminate injections of pentobarbital (5 mg/kg) from saline. The barbiturate administered alone consistently produced greater than 90% of pentobarbital-appropriate responding. However the concomitant administration of pentobarbital and increasing doses of the anxiogenics [29] pentylenetetrazol (PTZ) or bemegride resulted in a dose-related decrease in

[29] Anxiogenic drug: a drug that induces anxiety.

pentobarbital-appropriate responding.[30] Again, similar procedures have been used in humans, showing that the DS effects of the stimulant methylxanthine caffeine can be blocked by the anxiolytic drug[31] triazolam, although not by another anxiolytic, buspirone (Oliveto et al., 1997).

As a final modification to typical acquisition studies, subjects may sometimes undergo several different training acquisition phases in which progressively lower doses of the active drug or drugs are tested. This is in order to assess the smallest dose that subjects can readily discriminate from the placebo condition (a question that other researchers instead tackle using generalisation techniques, as we will review in the next section). As one example of this progressively lower training dose method, Preston and Bigelow (1998) found that humans trained to discriminate 20 mg of the opioid hydromorphone versus placebo, and for which the training dose was progressively reduced, were capable of discriminating down to a 3.5 mg dose.

Thus, together with stimulus generalisation studies (see next section), these tests of antagonism, and of dose-related acquisition, have proved very useful for characterizing the pharmacological characteristics of drugs. They can show striking parallels between human and animal responses. They also help to reveal some affective properties of certain drugs – something considered in more detail in Section 3.c. Sometimes these pharmacological characteristics may involve apparent increases or decreases in anxiety: one of many examples we will review in Section 3.c. on how affective states may act as DSs.

3.b.ii. Psychoactive drugs as DSs: Studies of stimulus generalisation

In most DD studies, once subjects acquire the discrimination between two conditions in the acquisition phase, a generalisation phase follows (with those subjects that did not reach discrimination criteria being excluded from further study; see for instance, Smith & Bickel [2001] on humans, and Vellucci et al. [1998] on rats). In this generalisation phase, researchers evaluate the similarity of the stimulus effects of a training drug to either the effects produced by other doses of the same

[30] See Section 3.c. for more examples of animal studies using anxiety and other negative affective states as DSs.

[31] Anxiolytic drug: a drug that relieves anxiety.

agent, or the effects produced by different agents. The degree of generalisation or 'substitution' is assessed from the subjects' operant choices (typically in a single trial per condition), and, in general, those doses or drugs that lead to training drug - appropriate responding have behavioural and pharmacological profiles similar to the training drug (e.g. Oliveto et al., 1997).

In human studies, a subject is deemed to have generalised a novel drug condition to the drug condition she was previously trained to discriminate (e.g., Drug A), when she significantly chooses this option over the other option/s (e.g., Drug B) in the new operant DD task. The same generalisation criterion is typically applied at the group level, but in this case subjects' scores are pooled and analysed together (see, for instance, Rush et al., 2004 on subjects discriminating d-amphetamine from placebo). In general, most studies typically assess significant changes in performance in both the DD and SR (subjective report) tasks during the generalisation phase. Additionally, if the novel drug or dose produces predominantly training-drug-appropriate responses (usually at 80% or greater), the novel drug is said to substitute completely and, thus, produce stimuli similar to those of the training drug (see Preston & Bigelow, 2000; Preston et al., 2007; Prus et al., 2005 & Appel et al., 1999;).

Other criteria may be used however. According to Gauvin and colleagues (1992), lower levels of training drug responding can still be considered accurate behavioural "barometers" for the qualitative and quantitative similarities between the new condition and the training one. If the novel drug produces responses predominantly on the placebo or no drug-appropriate lever, the novel drug is characterized as *not* substituting for the training drug. Intermediate degrees of responding are characterized as partial substitution (see Preston et al., 2007). In general, DD performance statistically significant above chance (50%) but below 80% of training drug appropriate responses can be considered as strong partial generalisation (Preston et al., 2007; Prus et al., 2005 & Appel et al., 1999). Researchers also commonly assess whether choices of the 'Drug A' option are significantly different from responses made to this option when the subject was on placebo (e.g. Stoops et al., 2005 & Rush et al., 2004). Drug appropriate responses significantly different from placebo may thus be considered of interest even if they are below chance; these would be considered instances of weak partial generalization.

With regard to subjective effects experienced during the generalisation phase, researchers typically evaluate which items of various self-report questionnaires are

significantly different from the placebo condition. As for the operant discrimination task, these effects can be evaluated both at the subject and at the group level.

As in humans, generalisation performance in animal studies can also be assessed at the within-subject (i.e., individual) level (see for instance, Carey and Fry, 1995 on pigs), and/or at the group level (see, for example, Wood & Lal, 1989 on rats). The criteria for generalisation, however, are even less standardized than in human DD studies, and often highlight the few generalising individuals, whereas human studies are more likely to look at what whole groups of people are likely to do. In drug versus placebo paradigms, for instance, some researchers assess whether the responses allocated to the drug option (drug lever) when on the testing condition are statistically indistinguishable from the responses emitted to that drug lever during the acquisition of the drug condition (see, for example, Gauvin & Holloway, 1991; additional examples will also be presented in Table 1 – Section 3.c.). Such approaches are all focused on full generalisation or substitutability. In other studies, researchers instead evaluate whether the responses on the drug lever during the generalisation tests are allocated significantly more often than chance (where chance is 50%; see, for example, Wood et al., 1989). Other researchers, in contrast, have weaker criteria still, treating generalisation as any allocation of responses to the drug lever that is significantly different from what was seen on saline (even if these levels are really low, see Mantsch & Goeders, 1998).

In humans, DD procedures are characterized as having *specificity* within the different pharmacological classes according to the effects they induce: subjects usually generalise to drugs that induce similar subjective effects to the training drug, and not to drugs of different classes (Chait et al., 1984). Such agents need not necessarily share an identical mechanism of action[32]: just be able to produce similar pharmacological effects in the subjects. For example, generalisation studies in humans show that diazepam generalises completely to other benzodiazepines, and also to barbiturates, but not to drugs belonging to other pharmacological classes such as stimulants, opiates, and antihistamines (Johanson, 1991). As mentioned in Section

[32] In pharmacology, the term "mechanism of action" refers to the specific biochemical interaction through which a drug substance produces its pharmacological effect. A mechanism of action usually includes mention of the specific molecular targets to which the drug binds, such as an enzyme or receptor.

3.b., animal data also show striking similarities with these human patterns (see Kamien et al., 1993 & Perkins et al., 1996). Animal studies, for instance, show that the benzodiazepine diazepam generalises completely to lorazepam – another benzodiazepine – and to the barbiturate pentobarbital, but not to drugs belonging to other pharmacological classes such as stimulants, opiates, and antihistamines (Johanson, 1991); and studies in rodents trained to discriminate the effects of the methylxantine caffeine from placebo also show that they generalise from low doses of caffeine to other stimulant drugs, even though they have different mechanisms of action, such as cocaine, amphetamine, methylphenidate and ephedrine (as well as other methylxanthines with a similar mechanism of action) (Stolerman et al., 2011).

In some generalisation studies, as was mentioned at the start of this section, instead of different compounds being compared, different doses of the agents used in training are tested. This is in order to identify the smallest dose of each of these drugs that can still act as DSs. As one example, Lile et al. (2011) found that in humans first trained to discriminate 30 mg oral delta-9-THC from placebo, and later administered with several doses of this cannabinoid during a generalisation phase, 15 mg of delta-9-THC was the smallest dose that could act as DS. In nonhuman animals this technique is also used. For instance, Harris and colleagues (1989) found that rats first trained to discriminate placebo from 20 mg/kg of the anxiogenic drug pentylenetetrazol (PTZ; a drug we will review in detail in a later section), and then trained with sequentially lower doses of this drug, were capable of reliably discriminating a PTZ dose of 10 mg/kg, but no lower.

As for any drug discrimination procedure, a generalisation test can be implemented as a three- or even a four-choice procedure. An example comes from rats trained to discriminate between the atypical antipsychotic clozapine, the typical antipsychotic chlorpromazine, and placebo (Porter et al., 2005), results from the generalisation phase showed that clozapine produced chlorpromazine appropriate responding at lower doses, and clozapine appropriate responding at higher doses. And in pigeons trained to discriminate the psychomotor stimulant amphetamine, the barbiturate pentobarbital, the opioid morphine and saline, the researchers found that pentobarbital generalised to chlordiazepoxide – a drug also with amnestic[33] properties – morphine generalised to methadone, and amphetamine generalised to cocaine and

[33] Amnestic drug: a drug that causes amnesia

methamphetamine. Low doses of phencyclidine generalised to saline, while higher doses partially generalised to pentobarbital and amphetamine (Li & McMillan, 2001).

In the sections that follow, I will pursue generalization effects further, but focussing specifically on animal examples where the emotional or affective properties of the stimuli used seemed to have been the most salient.

3.c. Affective states as discriminative stimuli in animal research

As we have just seen in Section 3.b., drugs that induce subjective effects – as reported by humans – can also act as DSs. Sometimes these effects are likely to include changes in affective states that might be used as DSs, as we will see in the following sections. However, in other cases these shifts in affective states are not necessarily induced by drugs. An interesting example is that of discrimination studies on animals' metacognitive abilities. Although these studies are typically designed to assess animals' capabilities related to monitoring their cognitive knowledge, some researchers posit the hypothesis that animals in these experiments are not monitoring their knowledge (i.e., knowing that they know something) but rather changes in affective states related to uncertainty regarding solving the discrimination tasks they undergo (see Carruthers & Ritchie, 2012).

In relation to drug discrimination experiments and affective states as DSs, we have already reviewed some studies which are consistent with animals using changes in affective states as discriminative cues. We saw, for instance, that mice can discriminate between MDMA – a drug that among different sensory and perceptual effects that could be used as DSs, induces powerfully positive affective consequences – and saline (Fantegrossi et al., 2009). Perhaps more revealing, we also saw that the concurrent administration of anxiety-eliciting compounds can render pigeons less able to discriminate barbiturates (which typically reduce anxiety: Herling and Winger, 1981) from saline (see Section 3.b.i.). Several researchers have built on and complemented this type of work by specifically investigating whether affective states – a particular category of interoceptive stimulus – (see Chapter 1) can be used as DSs in discrimination paradigms. These studies use both drug and non-drug induced affective states, but typically they have been carried out in nonhuman animals, and use only negative affective states as DSs. To the best of our knowledge, studies

investigating the use of affective states as DSs have not been carried out on humans, nor conducted using positive states.

One of the first animal studies potentially using negative affective states as DSs was carried out in the mid 1970s by Weissman (1976). In this study, 12 of 24 food deprived rats were made arthritic by a single intradermal injection of *Mycobacterium butyricum* into a paw, and both groups of rats were then trained to discriminate aspirin suspended in saline (56 mg/kg at a volume of 5 ml/kg) from saline alone. Both groups were capable of discriminating between aspirin and saline significantly above chance, but the arthritic rats did this far better than healthy rats (for example, the correct choice percentage for arthritic rats was significantly higher than the one for nonarthritic rats on the last four out of eight blocks of sessions). Furthermore, Weissman believed that these results probably underestimated the true differences between the arthritic and nonarthritic rats, since two very good discriminators in the arthritic group died before completing the testing phase. A single trial test phase followed in which all surviving rats were administered aspirin *per os* (orally). Results from this phase suggested that arthritic rats may be better than healthy rats at detecting the effects of injected aspirin to the effects induced by oral aspirin (8/10 subjects in the first group versus 6/10 subjects in the latter one). Weissman suggested that the arthritic rats were using their states of pain and its alleviation as DSs. Colpaert (1999) argued that this study represented the first empirical demonstration of chronic pain in animals. To the best of our knowledge this groundbreaking experiment has not been replicated in any other species, nor have other discrimination studies in animals using pain as a DS been carried out. It is relevant to underline, however, that other than an affective (negatively valenced) component, pain also induces other illness symptoms that are not affective states and could still be used as DSs.

Below we will see that other potential states of discomfort, disease relief, or drug induced states akin to states of anxiety in humans, have been used as DSs in animal discrimination studies (see Table 2 for different generalisation criteria used in these studies).

Table 2 – Generalisation studies using affective states as DSs in rodents

Degree of generalisation	**Study**	**Species**	**Drugs/treatment generalised to the training stimulus**
Strong generalization (≥80% choice of the training stimulus lever) in most subjects	Gellert & Holtzman (1979)	Rat	Morphine withdrawal to naltrexone withdrawal
	Gauvin & Holloway (1991)	Rat	Predator exposure (cat) to pentylenetetrazol (PTZ)
	Gauvin et al. (1992)	Rat	Ethanol 'hangover' to PTZ
	Gauvin et al. (1993)	Rat	PTZ to ethanol 'hangover' to
	Vivian & Miczek (1994)	Rat	Social stress (defeat by rival male) to PTZ
	Gauvin et al. (1996)	Rat	Alarm substance pheromone to PTZ
	Gauvin et al. (1997)	Rat	8 hr. photoperiod phase-advance to ethanol 'hangover'
Strong generalization (≥80% choice of the training stimulus lever) not in most subjects	Velucci et al. (1988)	Rat	Social stress (defeat by rival male) to PTZ
	Miczek et al. (1999)	Rat	Social stress (defeat by rival male) to d-amphetamine & Social stress (defeat by rival male) to cocaine
Moderate partial generalization (≥50% choice of the training stimulus lever) in most subjects	Shearman & Lal (1979)	Rat	Bemegride to PTZ & Cocaine to PTZ
	Shearman & Lal (1981)	Rat	Cocaine to PTZ
	Emmet-Oglesby et al. (1983)	Rat	Diazepam withdrawal to PTZ
	Harris et al. (1986)	Rat	Nicotine withdrawal to PTZ

	Wood & Lal (1987)	Rat	Cocaine withdrawal PTZ
	Lal et al. (1988)	Rat	PTZ to ethanol withdrawal
	Wood et al. (1989)	Rat	Cocaine withdrawal to PTZ
Weak partial generalisation (choice of the training stimulus lever on novel drug/treatment significantly different from when on saline)	Mantsch & Goeders (1998)	Rat	15 min. of restraint to cocaine

3.c.i. Drug withdrawal and "hangovers" as DSs

In an early study of drug withdrawal, Gellert and Hozman (1979) found that in rats made dependent upon morphine and trained to discriminate the opioid antagonist naltrexone from placebo, the amount and time course of naltrexone-appropriate responding following abrupt withdrawal of morphine seemed to be directly related to the degree of physical dependence (as indicated by loss of body weight, perhaps the most reliable single index of the morphine withdrawal syndrome in the rat). This effect was only seen for the group of rats administered morphine *per os* and supplemented by the implantation of morphine pellets – an administration method known to induce severe withdrawal symptoms. In contrast, the increase in naltrexone-appropriate responding was relatively weak for rats administered a morphine drinking solution, known to induce a lesser morphine dependence and therefore milder withdrawal symptoms. Unfortunately, the loss of body weight was not measured for this group, so similar comparisons were not made. It was concluded that these animals were probably using as a DS the narcotic withdrawal syndrome, which in humans is described as a distinctly unpleasant experience including weakness, lack of motivation, irritability, nausea, depression and dysphoria and anxiety (see Section 4). Delving deeper into the nature of the withdrawal experience, other studies highlight a role for anxiety-like states. These have all done so via the use of pentylenetetrazol ("PTZ"; also known as Cardiazol or Metrazol): a well-

characterized anxiogenic drug, which in humans was initially described as a convulsant, but later found to produce anxiety at subconvulsive doses and panic attacks at convulsant doses (Jung et al., 2002). During early studies on its pre-convulsant effects, subjects administered with non-convulsant doses of PTZ reported intense anxiety (Rodin, 1958; Rodin & Calhoun, 1970). In fact, because of this effect most subjects refused to participate further after their initial experience with this drug. This prompted a physician volunteer to experience PTZ-induced anxiety, who described his experience thus: "within a matter of seconds, I experienced catastrophic anxiety and said to myself now I know what these patients are going through, I can't possibly take that. It was a sense of utter distress and impending catastrophe. There is no doubt that this was one of the most anxiety producing events of my life" (Lal & Emmet-Oglesby, 1983). PTZ was first used as DS in a drug discrimination paradigm in a series of experiments in rats carried out by Shearman and Lal (Shearman & Lal, 1979). In these experiments, subjects were trained to discriminate subconvulsant doses of PTZ from placebo (saline). The rats successfully acquired the discrimination. Once trained, the anxiolytic drugs chlordiazepoxide, diazepam, flurazepam, clobazam, and meprobamate were found to all be effective antagonists of PTZ-responding in these rats in a dose-dependent manner. Subjects generalised the PTZ DS to bemegride (an anxiogenic drug) and cocaine in a dose-dependent manner. Similarly, a group of rats trained to discriminate bemegride versus saline generalised the bemegride stimulus to PTZ. Since this early work, a number of other studies have convincingly shown the anxiety-like nature of the state induced by this drug in animals (evidence detailed in Section 3.c.ii). The experiments reviewed below also show that drug withdrawal and PTZ have similar aversive effects.

One early example came from Emmett-Oglesby and colleagues (1983a & 1983b), who found that rats terminating chronic treatment with large doses of diazepam generalised the internal stimuli produced by withdrawal to the interoceptive DS produced by PTZ. In this case, 6 of 10 subjects selected the PTZ appropriate lever following injection of saline after 1 day of diazepam withdrawal. Interestingly, the 6 rats selecting the PTZ chose the placebo lever after being injected with 20 mg/kg diazepam. This result is consistent with the finding that in humans, the early stages of withdrawal from dependence on benzodiazepines are characterized by intense anxiety; indeed, for some drugs such as diazepam, anxiety may be the chief complaint of the patient withdrawing from dependence (Petursson and Lader, 1981).

Similarly, Wood and colleagues (Wood & Lal 1987; Wood et al., 1989) found that withdrawal from cocaine also produced a PTZ-like stimulus, as demonstrated by the selection of the PTZ lever by these subjects. This increased in intensity proportionally to the prior duration of chronic cocaine exposure. Furthermore, this effect was reversed by administration of the anxiolytic drug, diazepam. States of ethanol-withdrawal have also been found to produce a PTZ-like interoceptive stimulus. The use of PTZ discrimination to characterize anxiety induced by ethanol-withdrawal was first described by Lal and colleagues (1988) with rats selecting the PTZ lever during ethanol-withdrawal. In this experiment blood levels of ethanol had dropped to near zero levels at the time of withdrawal, ruling out any possible direct effects of ethanol. This study initiated a line of research into the mechanisms of ethanol-withdrawal and hangover associated anxiety. Gauvin and colleagues, for instance, found that rats trained to discriminate PTZ from saline did not generalise the PTZ stimulus to ethanol, but to high acute doses of ethanol administered at various time points prior to discrimination test session. Interestingly, the researchers reported that subjects responded on the PTZ-appropriate level in a quantitative fashion, which was dose- and time-dependent, and suggested that this behavioural discrimination paradigm may be a valid method for assessing animal analogues of human "hangover" states (Gauvin et al., 1992).

Similar signs of hangover from a large acute dose of ethanol were seen in rats trained to discriminate chlordiazepoxide – a benzodiazepine with sedative, hypnotic, anxiolytic, anticonvulsant, and muscle relaxant properties – and PTZ. Initially, rats responded on the chlordiazepoxide-appropriate lever after ethanol administration. At 12-14 hour after ethanol, a shift in responding to the PTZ-appropriate lever was taken to be a sign of acute withdrawal (i.e., hangover) from ethanol. In agreement with this interpretation, the shift to PTZ responding was blocked by administration of ethanol or chlordiazepoxide (Gauvin et al., 1989). In an additional study by this group, subjects were trained in a two choice discrimination task using as DSs ethanol delayed effects (EDE) versus "normal" basal homeostasis. Rats were thus injected with either 4 g/kg of ethanol or the equivalent volume of saline (SAL) 18 hour before the sessions. Additionally, each rat was injected with one 1 ml/kg injection of saline 15 min. before the discrimination sessions. Subjects succeed in acquiring the discrimination, and it was reported that during the generalisation phase rats showed a time-dependent, cyclic, return from the experimental "hangover" state to the "normal"

state, by 48 hour. Moreover, pre-treatment with low doses of ethanol and chlordiazepoxide did not generalise to the "hangover" state, and blocked the stimulus attributes of hangover. Interestingly, since all subjects responded on the EDE-appropriate lever when administered with 5.6 mg/kg PTZ, the researchers concluded that these results demonstrated the saliency and anxiogenic dimensionality of experimentally induced hangover (Gauvin et al., 1993). Finally, another experiment carried out by these researchers studied the similarities between the DS effects of morphine withdrawal and alcohol "hangover". This focussed on the negative effects that follow over-consumption of ethanol, and that are not dependent on prior addiction. In this fascinating experiment, adult male rats were again successfully trained to discriminate ethanol's delayed negative effects from baseline, non-hangover states. The trained rats were then shown to generalize significantly from the former state to morphine withdrawal, suggesting subjective similarities between the two states. When subjected to a sudden time advance-shift of 8 hours (mimicking "jetlag") the trained rats also generalised to this from hangover (Gauvin et al., 1997).

Not every state of hangover or drug withdrawal known to induce anxiety in humans fully generalise to PTZ in rats. At least this is what a study conducted by Harris and colleagues (1986) suggests. In this experiment rats were also trained to discriminate PTZ from saline, and then injected with nicotine intraperitoneally (0.64 mg/kg on the first day, and 1.25 mg/kg on subsequent days, in two groups for 21 and 15 days respectively). The researchers found that after acute nicotine administration, 35% of the rats pressed the PTZ lever. At 48 hour after termination of nicotine treatment 50% of rats selected the PTZ lever, a number of rats greater than that in a control group tested after an equivalent period without training. These results – which were similar for both nicotine-administered groups, and data were therefore pooled – were interpreted as partial generalisation of the state induced by withdrawal from chronic nicotine to the one induced by PTZ. Furthermore, since the anxiolytic diazepam antagonized the generalisation effect, it was concluded that withdrawal from chronic nicotine produces a weak PTZ-like stimulus that might be related to the state of anxiety observed in humans after nicotine withdrawal. As a whole, these experiments further showed the usefulness of operant drug discrimination procedures for the characterization of states of hangover and withdrawal through the use of these states as DSs.

3.c.ii. Studies using pentylenetetrazol (PTZ) as DSs: Animal anxiety?

A variety of further pharmacological studies of PTZ and its DS effects have shown that: 1) anxiogenic drugs, but not non-anxiogenic stimulant nor convulsant drugs generalize to the PTZ discriminative stimulus; and 2) the only drugs effective in antagonizing the PTZ DS are anxiolytic compounds to which tolerance does not develop. This led Lal and Shearman (1982) to conclude that the DS induced by PTZ is probably directly related to the anxiogenic properties of this drug, and therefore that the PTZ-saline paradigm represents an animal model of anxiety that can be used to detect potential anxiogenic and anxiolytic activity in new compounds. In a related paper, titled "Behavioural analogues of anxiety: animal models", Lal and Emmett-Oglesby (1983) made a similar case, not only reviewing previous findings on the PTZ drug discrimination paradigm, but also citing newer findings in rats suggesting that anxiety-provoking situations in humans, such as withdrawal from drug dependence (see previous section), also generalize to PTZ by these animals. Subsequent findings strengthened this case yet further, but showing that applying anxiety-provoking stimuli to animals, such as exposing rats to predator cues, also induces PTZ-appropriate responding in animals trained to distinguish PTZ from saline. This evidence will be reviewed later in this section.

Some evidence that anxiolytic drugs interfere with the acquisition or performance of PTZ-responding was presented in Section 3.c.i. Additional examples include cases from Lal and colleagues (Lal et al, 1980; Lal & Shearman 1980), who found that in rats trained to discriminate PTZ from saline, pretreatment with the anxiolytic drugs diazepam and valproic acid antagonized the DS produced by PTZ in a dose dependent manner. Shearman and Lal (1981) found that diazepam but not haloperidol significantly antagonized the DS produced by PTZ in rats trained to discriminate PTZ from saline, further supporting the hypothesis that both drugs have anxiogenic effects. Altogether, these studies show that the DS properties of PTZ can be attenuated by anxiolytics, thus supporting the hypothesis initially posed by Shearman and Lal that PTZ DS properties may specifically be related to an "anxiety-inducing" action of this drug. In agreement with this finding, many studies also found that the DS produced by PTZ generalises to other anxiogenic drugs, but not to stimulants or convulsants. For instance, Shearman and Lal (1978; 1979) found that

the discriminative stimuli induced by bemegride and cocaine – both drugs with anxiogenic properties (at high doses in the case of cocaine; the acute forced administration of cocaine at higher doses is frequently reported to produce anxiogenic effects in rats [e.g. Covington & Miczek, 2003]) – generalised to PTZ in rats previously trained to discriminate PTZ from saline. In a subsequent study, Shearman and Lal (1981) also found that high doses of cocaine (20mg/kg) but not diazepam nor haloperidol significantly generalised to the PTZ DS in rats trained to discriminate PTZ from saline, a result that replicated and supported their initial studies. In a similar study, Wood and Lal (1987) showed that male rats trained to discriminate cocaine from saline in a two-lever operant procedure generalised high doses of cocaine (20 mg/kg partially, 40 mg/kg fully) to a PTZ stimulus after chronic cocaine administration.[34]

Since the hypothesis that PTZ at subconvulsant doses induce states of anxiety was supported, this model of anxiety has been used in other animal species. Their findings provide us with a good picture of nonhuman animals' abilities to use the effects of PTZ as DSs, and reveal other potential states of anxiety – both drug and non-drug induced – that animal subjects appear to find similar to the states induced by PTZ. Rats can generalise from drug-induced anxiety states to non-drug-induced anxiety states: some studies investigated whether the interoceptive cue generated by this anxiogenic drug resembles that which might occur naturally during the lives of these animals. For instance, some experiments have evaluated whether aggressive defeat of the test male by another male produced an interoceptive cue which resembled (that is, strongly generalised to) the effects of PTZ. Were this the case, it would indicate that the internal state induced by the drug mimics some aspect of that occurring during behavioural interactions resulting in fear or anxiety-like responses (Vellucci et al., 1988; Vivian et al., 1994). When PTZ trained rats were injected with saline intraperitoneally, placed in the home cage of an aggressive male for 10 minutes and then placed into the operant chamber to perform the discrimination task, it was

[34] Wood and Lal (1987) argued that failure to replicate Shearman and Lal's previous results were unknown. However, they suggested that chronic administration of cocaine produced tolerance to the disruptive effects of this drug, and higher doses of cocaine then tested produced full substitution for the PTZ stimulus, thus supporting the hypothesis that high doses of cocaine are dysphoric and anxiogenic.

found that over a third of the animals did indeed complete the task on the PTZ lever, with a latency to lever-pressing no different from that which normally followed injection of PTZ. Furthermore, those animals continuing to choose the saline lever did so with a latency markedly longer than that normally following injection with saline and, interestingly, with larger numbers of interpolated responses on the PTZ lever (Vellucci et al., 1988). Altogether, these findings suggest that the state induced by PTZ in rats might share some elements with the state generated by an agonistic encounter with a rival male rat (i.e., social defeat stress). In a similar study, Vivian and colleagues (1994) found full generalisation to PTZ (i.e., PTZ lever responses ≥80% of the total responses) in 15 out of 25 defeated rats, while only 3 rats fully generalised aggressive defeat to saline. In an attempt to further explore the minimal social defeat conditions that were required for PTZ lever selection, PTZ-trained rats with previous defeat experience were placed into the home cage of the aggressive resident but within a protective wire mesh cage. Researchers reported that during these encounters, intruder rats oriented toward the resident and often remained in a crouch posture while they emitted ultrasonic vocalizations. Saline was administered after the first ultrasonic vocalization, and it was found that 7 out of 25 rats fully generalised to PTZ, while 9 rats did it to saline. These findings are thus consistent with the hypothesis that some male rats find the state generated by actual aggressive defeat more anxiogenic (i.e., more similar to the effects of PTZ) than the state induced by exposure to an aggressive conspecific while protected by a wire mesh. In both studies such post-defeat generalization effects were attenuated through pre-treatment with anxiolytics such as midazolam and clordizepoxide. Furthermore, defeat by an aggressive conspecific had specific effects on operant responding when midazolam was used as an additional DS, only engendering PTZ- appropriate, and not midazolam appropriate responding. Interestingly, Velucci and colleagues pointed out that failure to produce generalization to PTZ in a larger number of animals probably resulted in large part from inadequacies of these procedures. The 'magnitude' of the aggressive encounter and its precise form and outcome cannot be controlled by the experimenter and is, therefore, highly variable. Furthermore, removing the test male from the resident male's home cage and placing him in the operant chamber in order to begin the trial is itself fear reducing, since it removes the animal from the source of fear.

Predator-cues, another ethologically-relevant aversive stimulus, have similar effects: training to use PTZ as a DS followed by exposure to a domestic cat

engendered 92% PTZ-appropriate responding in 10 out of 12 rats when tested for stimulus generalization in the discrimination task. The other two subjects displayed freezing, defecation and urination behaviours that lasted for the entire two-min test session (Gauvin & Holloway, 1991).[35]

In a similar vein, but working with a different species, a series of studies by Carey and colleagues yielded interesting findings regarding pigs' capabilities for using states potentially analogous to human anxiety as DSs. The first of these studies showed that in an operant chamber consisting of a modified Skinner box set up with two levers – one for the PTZ condition and one for the saline one – pigs, like rats, can use the state induced by an injection of PTZ to make a particular lever-pressing response. Interestingly, here pigs were trained under a novel procedure in which they had to only press the PTZ lever when injected with PTZ, and to select both levers alternately following an intravenous injection of the same volume of the saline vehicle. A positive correlation between response time on the PTZ lever and the dose of PTZ was also observed for each pig (Carey et al., 1992). After showing pigs' capabilities to discriminate PTZ from saline in a drug discrimination paradigm, a second study followed in which pigs were found to generalise from the state generated by a PTZ injection to a conditioned emotional state (CES) consisting of expecting a non-injurious electric shock. When pigs were presented with the neutral tone stimulus during a saline session, subjects selected both levers alternately (the operant reinforced after being administered with saline). However, after two pairings of the tone with the application of an electric shock, all subjects started to show a significant bias towards the PTZ-designated lever during the subsequent application of the CES alone. Furthermore, this generalisation of the CES to the PTZ cue was antagonised by pre-treatment with the anxiolytic diazepam. Carey and Fry concluded

[35] To increase the level of arousal and stimulate interaction with the sequestered rats, the cat had been pretreated for 10 min. with an ecologically relevant stimulant (catnip) prior to predator/prey exposure, which significantly increased the amount of time the cat spent investigating, prowling, and pouncing in the presence of the rats. Gauvin & Holloway (1991) reported that the rats were behaviourally excited, and the attacking cat elicited defensive burying into the wood chips. During the 20-min exposure, rats thus continually attempted to bury themselves under the other rats, producing a frenzied "push and shove" between rats.

that this study provides further confirmation of the anxiogenic nature of PTZ in these pigs (Carey & Fry, 1993). In an interesting final study, published in 1995, Carey and colleagues used this PTZ discrimination paradigm to evaluate the responses of pigs when presented with different environmental putative anxiogenic stimuli. Thus, replacing the usual wire mesh floor of the Skinner box with a smooth wooden board induced a significant preference for the PTZ lever in 3 of the 4 subjects. When presented with a novel object – a small rubber ball – in the food bowl of the box at the start of test sessions, 2 of the 3 pigs tested selected the PTZ lever; interestingly, during this session these subjects choosing the PTZ lever did not eat the food reinforcements delivered into the food bowl. A reduction of the normal temperature in the operant chamber from 20-22ºC to 5ºC elicited a PTZ response in 2 out of the 4 subjects. Transporting subjects individually in a trailer for 20 min. elicited a PTZ response in only one of three animals tested, whereas an antagonistic encounter with an unknown male pig resulted in 2 out of the 4 pigs selecting the PTZ lever. Exposure to the smell of sow urine and to the smell of a pig carcass resulted in PTZ response in 50% of the subjects respectively, whereas the preputial gland secretion induced a period of anxiety in one of the three pigs tested. Finally, the smell of melted thymol did not elicit PTZ responses in any of the 4 animals tested (Carey and Fry, 1995). Although this study was carried out with only four subjects, and to the best of our knowledge has not been replicated, these data nevertheless suggest that pigs might be capable of generalizing the effects induced by the anxiogenic drug PTZ to potentially anxiogenic non-drug induced states.

3.c.iii. Studies using cocaine and d-amphetamine as DSs: other forms of animal anxiety?

Some anxiety-related studies assessed whether states generated by drugs such as cocaine and d-amphetamine share similarities with non-drug induced states of potential anxiety in rats. Miczek and colleagues, for instance, studied whether rats generalise the states induced by high doses of d-amphetamine and cocaine to "social defeat stress". In this case nine out of 35 rats in the d-amphetamine group, and six of 18 in the cocaine group fully substituted the rival male defeat experience for d-amphetamine or cocaine respectively. Since a substantial number of rats generalised

this social stress experience to the d-amphetamine or cocaine cue only partially or inconsistently, the researchers interpreted these results as suggesting strong individual differences regarding whether they find these two states similar (Miczek at al., 1999). However, the same concern pointed out by Velucci and colleagues regarding their study on social defeat and PTZ-effects in rats apply to this case: "removal of the test male from the resident male's home cage and placing him in the operant chamber in order to begin the trial is itself fear reducing, since it removes the animal from the source of fear". Another possible interpretation, however, is that the cocaine dose used in this experiment (10 mg/kg) might not have been high enough to induce anxiety in these rats. This is consistent with previous studies showing that this dose was not high enough to engender PTZ responding in rats trained to discriminate PTZ from saline (see Sections 3.c.i and 3.c.ii).

Finally, in another anxiety related study in rats two groups of subjects were trained to discriminate injected cocaine from saline. The first group (n=10) was given 10 mg/kg of cocaine, whereas the second one (n= 6) was given 20 mg/kg of this same drug. During this discrimination phase rats first received intraperitoneal saline pre-treatment prior to the start of the test session while placed into a commercially available restraining apparatus for 15 min. and then were promptly transferred to the operant boxes for generalisation testing. Interestingly, subjects from both groups generalised the effects induced by the stressful state induced by movement restriction to cocaine. Yet, only in the first group the number of rats meeting the 85% generalisation criterion (6/10) when restraint was administered subsequent to an injection of saline was significantly greater than when saline was administered alone. The researchers pointed out that one possible interpretation of these results is that a component of the interoceptive state produced by cocaine is associated with a state of anxiety. However, they also suggested the possibility that it was not interoceptive cues associated with stress, but rather those produced by stress removal that generalised to cocaine. In these experiments, rats were restrained immediately prior to generalisation testing, but the stressor was not actually present during the test sessions. Thus, removal of the rats from the restraining devices and placement into a familiar environment (the discrimination box) may have engendered a positive interoceptive state associated with safety from an aversive stimulus (Mantsch & Goeders, 1998).

As a whole, all those studies reviewed from Section 3.c to 3.c.iii provide us with a good picture of how the use of affective states as DSs by nonhuman animals can be a useful technique in order to investigate these animals' emotional worlds.

4. Conscious awareness of the stimuli used as DSs?

In Section 3 we saw evidence suggesting that the subjects of DD experiments need to detect the effects of drugs in order to succeed in the discrimination tasks. But, do they need to *consciously* access these internal stimuli? Unfortunately we cannot assess this directly, but next I review some evidence from human and animal drug discrimination and generalisation experiments relevant to this question. I also review whether or not researchers working with animals have interpreted their own data as revealing animals' subjective experiences.

4.a. Correlational self-report evidence from human drug discrimination studies

When thinking about discrimination in humans, most people would probably intuitively agree that subjects only succeed in discriminating different stimuli because they are aware of them – think, for instance, of a task in which subjects are rewarded for telling apart a red versus a green light. However, examples such as blindsight and blindsmell (see Chapter 2) suggest that this sort of discrimination might be possible without conscious processing. Therefore, it is conceivable that reported awareness of the discriminated stimuli is actually a consequence of nonconscious discrimination. As we have seen already, human DD studies most often consist of an operant discrimination task akin to that used in animals, and also a self-report questionnaire-based task which tries to capture the interoceptive effects induced by drugs (see Section 3b). As we will see next, by evaluating the relationships between these two tasks we can assess whether conscious processing of these interoceptive drug cues is necessary for telling apart drugs in an operant discrimination paradigm.

To the best of my knowledge only a few studies have empirically assessed the relationship between human drug discrimination performance and self-reported drug effects. In a study of cocaine dependents trained to discriminate oral cocaine from placebo, Oliveto and colleagues (1998) found that during a generalisation phase in

which subjects were administered smaller cocaine doses, the visual analogue scale (VAS) self-report items "Similar to cocaine" (R = +0.84), "Similar to placebo" (R = -0.77), and "High" (R = +0.70) correlated best with drug discrimination performance at these various lower cocaine doses. A study by Duka and colleagues (1998a) using oral ethanol as the active drug also found significant correlations between the results of self-report and of drug discrimination tasks. In this case, subjects – moderate drinkers – were trained to discriminate placebo from ethanol. Once discrimination was acquired, they were provided with smaller doses of ethanol in a generalisation phase. Subjects were able to discriminate the three higher doses of ethanol from placebo, and their abilities to do so significantly correlated with feelings of light-headedness at the two lower of the four doses tested (R = +0.78, p = 0.00 at the 0.025 g/kg dose and R = +0.57, p = 0.02 for the 0.05 dose), with similar tendency for the third lower dose (R = +0.47, p = 0.06); while at the highest dose, discrimination ability was significantly related to changes in taste (R = + 0.57, p = 0.02). According to the researchers, the lack of correlation between light-headedness and discrimination performance at this highest dose may have been a ceiling effect, because the high level of light-headedness in most subjects at this dose did not allow a statistical relationship to be revealed. Alternatively, this finding may reflect that more salient stimuli detract from the importance of the other cues (i.e., the salience of taste cues at this high dose prevented subjects from using and reporting the light-headedness cue). Another study by this group reported that the generalisation performance of tobacco smokers previously trained to discriminate placebo from nicotine (administered in chewing gum) correlated best with the scores for the subjective VAS self-report "Sensations in the mouth" (R = +0.66, p < 0.001), and secondly with self-reported effects on "Heart rate" (R = +0.41, p < 0.02) (Duka et al., 1998b). Since subjects had to chew the gum for 10 minutes, the researchers suggested that the strength of the stimulus "Sensations in the mouth" might have prevented subjects from identifying and reporting subtler DSs (Dora Duka, personal communication).

The one similar study to focus on unambiguously negative effects involved experimentally-induced drug withdrawal. Oliveto and colleagues (2002) trained opioid dependent subjects to distinguish between an intramuscular injection of the opioid antagonist naloxone ('Drug A') from placebo ('Drug B') under a three choice drug discrimination procedure, in which subjects identified the drug condition as ''A,'' ''B,'' or ''N'' (neither A nor B – 'novel'). Once the discrimination was

acquired, smaller doses of naloxone were tested during the generalisation phase, and it was found that the VAS items "Similar to naloxone" ($R = +0.87$, $p < 0.05$), "Similar to placebo" ($R = -0.82$, $p < 0.05$), and "Bad" (effects) ($R = +0.41$, $p < 0.05$), correlated significantly with naloxone-appropriate. Note that an R of 0.87, the highest I came across in any study, means an R^2 of 0.757; in other words, here 76% of the variance in DD was being explained by variance in SR

Finally, a recent study assessed the relationship between drug discrimination performance and ratings of subjective effects across several experiments pooled (Reynolds et al., 2013): six different experiments from the same lab, using identical procedures, analysed together. Subjects were all trained to discriminate d-amphetamine from placebo, and presented with smaller d-amphetamine doses once they had acquired the discrimination. For all doses of d-amphetamine in the generalisation phase, there were significant correlations between drug appropriate responding and the subjective ratings on 15 items from the Drug-Effect Questionnaire (DEQ) (see Appendix Table 1 in Section 3.b.). The items "Active, alert, energetic" ($R = +0.51$), "Any effect" ($R = +0.50$, $p < 0.0001$), "Stimulated" ($R = +0.45$), and "Talkative, Friendly" ($R = +0.41$, $p < 0.0001$) correlated best with drug discrimination, while the correlations found for other DEQ, Adjective Rating Scale, and ARCI items were weaker. The researchers concluded that "the current findings demonstrate that diverse subjective effects contribute to the discriminative effects of D-amphetamine" and "on the basis of the degree of correspondence between responses in drug discrimination and ratings of subjective effects, it seems reasonable to suggest that subjective effects contributed to the accurate discrimination of D-amphetamine".

Altogether, these results from human studies using different drugs as DSs empirically demonstrate that operant discrimination performance and self-reported states are closely related. Nevertheless, we still cannot address whether the subjective effects induced by psychoactive drugs *guide* subjects' discrimination performance, since no causal relationship is revealed. In fact, these findings do not even rule out the possibility that subjects can discriminate psychoactive drugs without consciously accessing any of the effects induced by these drugs. In other words, and taking drug generalisation as an example, these correlational results do not rule out the possibility that subjects could discriminate drugs at doses smaller than the ones at which they can self report any drug induced subjective effect, and thus that SR follows DD.

4.b. Threshold effects in human DD studies

As seen in section 3., DD human studies typically assess significant changes in performance in both the DD and SR tasks during the generalisation phase. Additionally, some of these studies also assess the lowest drug dose that induces operant choices significantly different from those in the placebo condition (*DD threshold dose*); and the lowest drug dose that yields scores significantly different from placebo in the subjective self-report questionnaires filled after the DD task (*SR threshold dose*). Since the ability to emit verbal reports implies conscious processing of information (see Chapter 2), by comparing these thresholds we can gain more insight into whether DD can happen without any related subjective effects.

As one example illustrative of these studies, Lile and colleagues (2012) found that in humans previously trained to discriminate 30 mg of delta-9-THC from placebo, 15 mg of delta-9-THC was the smallest dose that could act as DS at the group level, and that their results from self-report questionnaires (see Section 3.b and Appendix Table 1) corresponded with this: 15 mg of delta-9-THC was also the smallest dose whose effects subjects were able to describe (in this case yielding high scores for the drug–effect questionnaire items "Any Effect", "High", "Like Drug" & "sedated"). The table below shows the studies reporting these thresholds:

Table 3 – Studies reporting DD and SR threshold doses

Type 1 studies: DD and SR thresholds are nsd (same dose)
Type 2 studies: DD threshold < SR threshold
Type 3 studies: SR threshold < DD threshold
All results correspond to performances at the group level

Type	Study	Training drug vs. placebo	DD Threshold	SR Threshold
1	Lile et al., 2009	25 mg delta-9-THC po	7.5 mg delta-9-THC	7.5 mg delta-9-THC
1	Lile et al.,	25 mg delta-	10 mg	10 mg

	2010	9-THC po	delta-9-THC	delta-9-THC
1	Sevak et al., 2011	10 mg methamphetamine po	5 mg methamphetamine	5 mg methamphetamine
1	Lile et al., 2011a	150 mg cocaine po	150 mg cocaine	150 mg cocaine
1	Lile et al., 2011b	30 mg delta-9-THC po	5 mg delta-9-THC	5 mg delta-9-THC
1	Lile et al., 2012a	30 mg delta-9-THC po	15 mg delta-9-THC	15 mg delta-9-THC
1	Lile et al., 2012b	30 mg delta-9-THC po	15 mg delta-9-THC	15 mg delta-9-THC
1	Lile et al., 2005	15 mg d-amphetamine po	2.5 mg d-amphetamine	2.5 mg d-amphetamine
2	Stoops et al., 2005	30 mg methylphenidate po	20 mg methylphenidate	30 mg methylphenidate
2	Kelly et al., 1997	0.45 g/liter of body water (lbw) ethanol po	0.30 g/lbw	0.45 g/lbw
2	Rush et al., 2003	15 mg d-amphetamine po	5 mg d-amphetamine	10 mg d-amphetamine
3	Rush et al., 2004	15 mg d-amphetamine po	10 mg d-amphetamine	2.5 mg d-amphetamine
3	Lile et al., 2006	30 mg methylphenidate po	20 mg methylphenidate	10 mg methylphenidate
3	Rush et al., 1998	20 mg d-amphetamine po	5 mg d-amphetamine	2.5 mg d-amphetamine
3	Lile et al., 2014	30 mg delta-9-THC po	15 mg delta-9-THC	5 mg delta-9-THC

As we can see, in many of the studies above (*Type 1*) DD and SR thresholds are not significantly different. This means that the smallest drug dose phase that induces changes in behaviour is the smallest dose that induces reportable changes in feelings assessed by the SR questionnaires. This finding is thus consistent with the hypothesis that DD only happens when accompanied by reportable changes in feelings. In the studies labelled as *Type 3*, the SR threshold is smaller than the DD. This finding suggests that, sometimes, drugs may be able to induce reportable changes in states at doses that cannot be discriminated from placebo in the operant DD task (perhaps suggesting that these stimuli cannot gain control of behaviour). This evidence also backs up the hypothesis that DD cannot happen without any associate SR. However, some of the studies above suggest that this may not always be the case: the smaller DD threshold than the SR one in the *Type 2* studies suggest that under certain conditions, humans may be able to behaviourally discriminate drugs from placebo at doses that do not induce reportable feelings (at least as assessed by the drug sensitive SR questionnaires used in these studies, though other explanations are possible as will be reviewed in the next chapter).

Overall, we have just seen that these pieces of evidence are consistent with the results of the correlational studies reviewed in Section 4.b.: they largely add evidence to the hypothesis that operant discrimination performance and self-reported states are closely related in verbal humans, but they also back up the possibility that humans could potentially discriminate drugs in DD paradigms – at least at week levels of performance – without consciously processing any of the effects induced by these drugs.

4.c. Acquisition studies: interpreting their results in terms of subjective experience in humans and other animals

Throughout Section 3.b we saw that animals can only discriminate psychoactive drugs that humans report being able to tell apart due to the subjective effects (i.e., feelings) they induce. For example, Kamien and colleagues (1993) found that results from amphetamine studies show similar patterns of generalisation to other psychomotor stimulants in humans, pigeons, rats and rhesus monkeys. These findings are consistent with Gauvin and colleagues (1992) consideration that "novel

test drugs from pharmacological classes outside the training drug stimulus class have generally engendered saline appropriate responding" (in drug versus saline two choice paradigms). Some other interesting cases are also in agreement with the hypothesis that, perhaps, animals use the subjective effects of drugs as DSs. For instance, Lal and colleagues (1978) discovered that rats can rapidly learn to discriminate small doses of haloperidol – an antipsychotic dopamine antagonist that does not normally induce subjective effects – from saline, but only when pre-treated with amphetamine, a stimulant dopamine agonist. This result suggests that these rats used as DSs the effects of haloperidol drug on the state induced by amphetamine. Similarly, naloxone is a pure opioid antagonist that unmanipulated subjects find hard to discriminate. However, rats can discriminate this drug from saline if they are pretreated with morphine 24 hours earlier. If the morphine treatment is not given, in contrast, naloxone treated rats choose the saline lever. This suggests that it was only the state induced by naloxone on the effects previously induced by morphine that these animals could use as a DS (Lal et al., 1978). Similarly, we saw that unlike healthy rats, only hypertensive subjects seem capable of using the state generated by reduced blood pressure as DS (Lal & Yaden, 1985), and that the use of PTZ as a DS by rats is abolished by anti-anxiety drugs (e.g., Lal et al., 1980; Lal & Shearman, 1980).

In light of these results, Albert Weissman's early finding that arthritic rats can discriminate the effect of aspirin better than healthy rats (1976) led Colpaert (1999) to write the following bold conclusion: "arguably for the first time ever has been made an actual observation of chronic pain in animals"; even more strongly, he argued that "while at some point in time it was felt that subjective experiences were uniquely human and utterly inaccessible, the drug discrimination paradigm carries the exciting promise of rendering amenable to experimental analysis in animals experiences, which, by their nature, have been considered as out of the range of scientific inquiry". This view is in agreement with Lal (1979)'s statement that if there is a drug known to provide relief from a symptom which is specific of a disease or state, the presence or absence of this disease or state can be determined by the use of drug discrimination.

4.d. Generalisation experiments in drug discrimination studies: interpreting their results in terms of subjective experience in humans and other animals

In Section 3.b.ii we saw that in humans, DD procedures are characterized as having *specificity* within the different pharmacological classes according to the effects they induce: subjects usually generalise to drugs that induce similar subjective effects to the training drug, and not to drugs of different classes. Such agents need not necessarily share an identical mechanism of action: just be able to produce similar pharmacological effects in the subjects. We also argued that animal data also show striking similarities with these human patterns: as Chait and colleagues put it (1984), drugs with similar DS properties in laboratory animals, i.e. drugs that animals generalise between (e.g., between different stimulants, or between different opioids) also tend to produce similar subjective effects in humans (e.g. Chait et al., 1984). In the next lines I will summarize what the researchers conducting these studies concluded in terms of results in terms of subjective experience.

To start with, we saw several studies showing that rats generalise between drugs with similar subjective effects, and also about antagonistic effects of anxiolytic compounds. The finding that rats generalise states generated by PTZ to the ones induced by bemegride and high doses of cocaine respectively, both drugs known to induce states of anxiety in humans, was interpreted by Shearman and Lal (1978; 1979) as suggesting that the DSs produced by these drugs are "related to an 'anxiety-inducing' action". In another work, Harris and colleagues (1987) concluded that "all anxiogenic drugs substitute for PTZ, and all anxiolytic drugs block the PTZ stimulus". In agreement with this view, Gauvin and colleagues (1989) concluded that findings on PTZ drug discrimination "support the view that the interoceptive stimulus produced by PTZ is anxiogenic in nature". These researchers argued that "evidence from several studies supports the view that the interoceptive discriminative stimulus produced by PTZ in rats is best correlated with the anxiogenic effect of this compound in man, since: 1) clinically efficacious anxiolytics block them; 2) compound such as the ß-carbolines that are anxiogenic in man generalize to them; 3) although PTZ is a convulsant, nonanxiolytic anticonvulsants do not block the PTZ cue; … and 4) drugs and drug-withdrawal states characterized as 'anxiety-producing' in humans generalise to the PTZ cue in rats", concluding that "this impressive data base strongly supports the view that PTZ discrimination in rats parallels the subjective reports of anxiety in humans".

These types of evidence are consistent with the hypothesis that not just humans but also other animals need to consciously process the effects induced by

psychoactive drugs in order to discriminate them in DD paradigms. An even more compelling case can be made by taking into account studies showing that animals can generalise drug-induced states to potentially similar states not induced by drugs (see Section 3.c.i for a more detailed description of these studies). In all these studies it seemed highly plausible that subjective states explained the animals' behaviour, but what do experts in the field believe? Here I review the interpretations and conclusions of the researchers who conducted these experiments in order to gain more insight on whether animals need to consciously process these drug effects in order to discriminate them in DD paradigms. I focus on cases where withdrawal or hangover states generalized to anxiogenic drugs, and where non-drug induced states generalized to anxiogenic drugs.

In section 3.c. we saw that rats trained to discriminate saline from PTZ, a compound that is anxiogenic to humans, generalise from it to other drug-induced states of potential anxiety. For example, partial generalisation from a 48-hour nicotine withdrawal to PTZ was interpreted by Harris and colleagues (1987) as "a withdrawal stimulus related to human anxiety". In the same vein, finding that rats generalise from PTZ to states of diazepam withdrawal led Emmet-Oglesby (1983) to write "the withdrawal state detected by this procedure is related to the long-lasting and pervasive anxiety during withdrawal reported in the clinical literature", and to state that the drug discrimination paradigm "provides the most sensitive and accurate behavioural analog in animals to what humans verbally report about subjective drug experiences". Discussing their work on alcohol withdrawal, Lal and colleagues (1988) likewise argued that "the discrimination of PTZ by animals can be used to investigate aspects of ETOH withdrawal that are related to the occurrence of anxiety in humans", so "providing information about subjective symptoms of clinical significance". They concluded boldly that "questions concerning subjective events may…be approached through animal models of subjective drug effects. Drug discrimination, in which animals discriminate internal stimuli arising from drugs, is one such model". Similarly, generalisation from both high doses of cocaine and withdrawal from them to PTZ was interpreted by Wood and Lal (1987) as suggesting "that high doses of cocaine produce anxiogenic stimuli", and "that cocaine withdrawal is anxiogenic". The researchers concluded that the DD paradigm is "a bioassay for detecting subjective effects, including anxiety". Wood and colleagues (1989) further argued for "the utility of drug discrimination as an *in vivo* assay for

detecting subjective effects". A final example focuses on commonalities between two forms of withdrawal. Finding that rats generalise between naltrexone withdrawal and morphine withdrawal, Gellert and Holtzman (1979) argued that "several lines of evidence suggest that the discriminative stimuli engendered by naltrexone are related to withdrawal phenomena", concluding that "drug discrimination procedures may afford an animal model for studying the subjective effects of narcotic analgesics".

Other ingenious generalisation experiments, on pigs as well as rats, investigated generalisations between anxiogenic drugs and other, non-drug-related but arguably anxiogenic experiences. They are not always clear-cut. For example, when Mantsch and Goeders (1998) showed that rats generalise the state induced by high doses of cocaine to the stressful state induced by movement restriction, they were unsure how to interpret it. However, both their suggestions centred on subjective experiences. One possibility, inspired by the way that high doses of cocaine can induce states of anxiety in humans was that "a component of the subjective effects of cocaine may be associated with 'anxiety'". Alternatively, they suggested that the "removal of the rats from the restraining devices and placement into a familiar environment (the discrimination chambers) may have engendered a "relaxation" interoceptive state associated with safety", such that "similarities between this state and positive subjective effects of cocaine such as those reported in humans (e.g., euphoria and a sense of well-being) may have accounted for the generalization observed". In other studies, the valence of the affective states in question have been less ambiguous. Rats generalising PTZ states to aggressive defeat by another male were interpreted by Vellucci and colleagues (1988) as "adding some credence to the claims that the PTZ discrimination is in some sense a viable model of anxiety". Discussing their own very similar study, Vivian and colleagues (1994) concluded that rats generalise a PTZ state to an agonistic encounter because "exposure to a threatening and attacking conspecific resulting in defeat produces an anxiety like state". Likewise, finding that rats generalise from PTZ to predator cues led authors Gauvin and Holloway (1991) to suggest that "the PTZ drug discrimination task provides a behavioral assay for a hypothesized psychological 'affective' metric space best categorized as a single continuum bounded on one end by 'anxiety,' and 'anxiolysis' on the opponent end". Gauvin and colleagues (1996) put forward similar reasons for finding that rats generalise from exposure to an alarm pheromone to PTZ. They argued that "an ethologically relevant, strain dependent pheromone can induce

subjective states in rats similar to a 15 mg/kg PTZ training cue … based on the affective dimension". They concluded "the present study implicates a significant role of olfactory stimuli in inducing changes in the interoceptive subjective effects in rats". In another inventive study by this group, Gauvin and colleagues (1997) interpreted the finding that rats can generalise states of morphine withdrawal to states of alcohol 'hangover', and even to a sudden time shift of 8 hours (mimicking "jetlag") as providing evidence "for the subjective similarity between ETOH hangover, opiate withdrawal states, and the physiological disruption induced by circadian phase advances".

As for the similar studies conducted on pigs, when these animals generalised from PTZ to various potential environmental causes of anxiety (see Section 3.c.ii), Carey and Fry (1992, 1995) argued that "an anxiogenic drug can be used to test for the presence or absence of a subsequent anxiety state in response to a variety of environmental stimuli", and furthermore, that "our approach…provid[es] the animal with a direct means of expressing its own psychological state. The PTZ discrimination paradigm therefore provides a valuable scientific tool for future studies of animal welfare". Equally bold was their claim that "ambiguities in the assessment of animal welfare can be overcome by employing the anxiogenic drug PTZ in a pharmacological conditioning procedure which enables self-expression of the psychological state of anxiety" (Carey & Fry, 1995).

Thus researchers often present the most parsimonious explanation for generalisations between drug-induced and non drug-induced states as being that they are affectively similar. Gauvin and colleagues (1989) argued that "the internal contextual events that determine stimulus generalization profiles may well be the affective state of the organism", concluding that "the drug discrimination paradigm is a viable model to assess interoceptive affective components associated with drug administration". Returning to Colpaert, as he eloquently put it: "the drug discrimination paradigm thus appears to offer an experimental access to the subjective experience, or perception, of stimuli that are produced by drugs or that arise from other, physiologically defined, but invariably internal conditions. In this capacity, drug discrimination studies have begun to provide insights into the psychophysiology of subjective perception" (Colpaert, 1999). More recently, Díaz & Velázquez (2000) also argued that "by strong analogy we can infer that animals perceive or are aware of the subjective effects of drugs".

Were these animals really conscious of the internal states they used as DSs? As we have just seen, some of the authors of the studies above cautiously placed those terms representing human affective states in inverted commas, stressing the point that these nonhuman animals' affective states probably differ from those typically experienced by humans. Even if the nonhuman mammals in these experiments were experiencing states homologous to human affective states, this *per se* does not necessarily imply that these states were processed consciously: in Chapter 2 we saw that affective states in humans are not always consciously processed. However, a relevant point here is that these animals were using states as DSs in an operant discrimination paradigm. As seen in Chapter 2, making non-routine bindings that require creation of unforeseen combinations – such as learning to discriminate drug conditions from placebo, and linking this knowledge to an operant in order to be rewarded – is considered one of the roles of conscious processing. Plausible as this is, unfortunately pigs and rats cannot use language to self-report it. In the next section I will try to shed more light on this difficult question by taking into account evidence from the human drug discrimination studies, which crucially have a verbal self-report component.

4.e. Conclusions: is the use of interoceptive cues as DSs in humans and animals guided by awareness of subjective feelings?

In Section 4.a. we have seen experimental studies showing that in humans, the operant discrimination task and self-report data are closely correlated in DD studies. This is consistent with the hypothesis that conscious awareness of drug-induced states might play a crucial role in operant discrimination performance in verbal humans. However, we saw that these correlational results do not prove any causal relationship between the conscious awareness of drug-induced states and the ability to tell apart different drugs: they just suggest that these two events are closely related.

We also saw that evidence from drug generalisation studies in humans reporting DD and SR threshold doses is consistent with this view. However, even though DD thresholds in these studies are often equal to or larger than the SR thresholds, in some of this work the DD thresholds are smaller than the SR ones, thereby suggesting that humans may be able to use states induced by drugs – and perhaps other types of interoceptive stimuli – as DS without consciously processing

these discriminative cues. Thus although it has been argued that "drug discrimination in humans is due to their subjective effects" (Díaz & Velázquez, 2000), the discrimination of interoceptive DSs could still be a non-conscious process, with awareness of the discriminated stimuli being a consequence (not a cause) of this process.

With regard to nonverbal beings, studies reviewed in Section 3.a. showed that rhesus monkeys can use gut fill (pulsations in the jejunum) or bladder fill as DSs; that rats can use states generated by electric stimulation of their gastrointestinal mucosa as DSs, and that hunger can also be used as DSs. Perhaps even more impressively, among other examples we saw that only hypertensive rats trained to discriminate the antihypertensive drug clonidine seem capable to generalise its effects to other antihypertensive drugs when these drugs have different mechanisms of action. Not being clear whether conscious processing of these states is necessary for them to act as DSs, results from drug discrimination and generalisation experiments reviewed in Section 4.b provide additional evidence consistent with the hypothesis that at least some mammals might be conscious of the interoceptive DSs they discriminate. These studies show that 1) states that in humans would either cancel out certain drug effects or conversely make them more detectable, similarly affect rats' abilities to use the drugs as DSs. For example, rats trained to discriminate PTZ from placebo stop picking the PTZ lever even when on this drug, if at the same time they have been dosed with anxiolytics; 2) Rats generalise between drugs (or drug withdrawal effects) that cause similar subjective experiences in humans, even when the drugs' chemical natures and modes of action are quite different. For example, rats trained to discriminate PTZ from placebo pick the PTZ lever when exposed to diverse other anxiogenic drugs. Likewise, rats trained to discriminate caffeine from placebo pick the caffeine lever when exposed to low doses of others stimulants, such as cocaine, amphetamine, methylphenidate and ephedrine; and rats trained to discriminate alcohol "hangover" from normal baseline states pick the hangover lever if they are then subjected to morphine withdrawal; 3) Rats also generalise between drug and non-drug states that seem likely to have similar subjective effects (from what we know from humans, and/or from how they affect classic tests of rat anxiety), which as we saw are usually interpreted by the researchers as relevant to subjective experience.

In light of this body of evidence, some have also drawn parallels with humans to argue that "drug discrimination in animals closely matches human drug

discrimination" (Díaz & Velázquez, 2000; see also Oliveto et al., 2002). However, we have seen that some authors have been more cautious than others when comparing the states animals may use as DSs to potentially homologous states in humans. Perhaps more importantly, some evidence from human DD studies suggests that the use of internal states as DSs might not always require conscious processing of these stimuli. Thus overall, and regardless of species, the main question then perhaps remains unanswered: does conscious awareness of the effects induced by drugs guide discrimination performance, or at least very strong DD? Or, to put it more broadly, is the use of interoceptive cues as DSs caused by the awareness of subjective feelings? In order to achieve a better understanding of this type of relationship, Chapter 4 presents results from a meta-analysis of 41 drug discrimination experiments in humans using different psychoactive drugs as DSs.

5. Summary

In this chapter we have seen that animals of different species are constantly processing information in order to use it as DSs. In Section 2 we saw these stimuli are often of an exteroceptive nature, and that animals in their natural environments spontaneously learn to identify those stimuli that indicate when performing a certain action will probably yield a reward. We have also seen that nonhuman animals can be trained by humans to use different stimuli as DSs, and that this possibility provides us with a window into their perceptual and cognitive worlds.

In Section 3 we saw that animals also use interoceptive stimuli, i.e., stimuli that originate inside the body, such as hunger or pain, as DSs. Section 3.b showed that there is a large body of work – both in human and nonhuman animals – on the use of drugs as DSs. There I introduced the concepts of drug discrimination and drug generalisation, which are key to the operant DD paradigm typically used in these studies. Crucially, we saw that both human and nonhuman subjects in these experiments only discriminate drugs that humans report to induce subjective effects, i.e. psychoactive drugs; likewise, we saw that only drugs that induce similar subjective feelings in humans are normally generalised in these experiments. In Section 3.c we saw that many of these drugs induce changes in affective states, and I reviewed evidence showing that some nonhuman mammals (namely rats and pigs) are

capable of generalising between drug and non-drug induced states related to negative affect (in most cases seemingly like states of anxiety in humans).

Since the body of evidence presented in Section 3 was consistent with the hypothesis that drugs can be used as DSs only if subjects are conscious of at least some of the feelings these typically induce in humans, Section 4 tried to shed some light on this complex question. In Section 4.a. I reviewed several relevant DD studies in humans. By assessing the relationship between subjects' performance in the operant DD task – analogous to the one used in animal DD studies – and the self-reports induced by the drug discriminated, these studies showed that the operant DD performance is closely related to the conscious processing of the drug effects. In Section 4.b., however, we saw that even though most studies reporting threshold doses for the DD and SR tasks support the view that humans cannot discriminate drugs in an operant paradigm unless they consciously process any of the effects induced by these substances, some other studies support the opposite view. In Section 4.c. we saw that those researchers conducting the experiments reviewed in Section 3.c. all more or less cautiously agreed in concluding that those generalisations were probably due to similar subjective states. Moreover, we saw that some of them regarded the DD paradigm as a window into the subjective states of subjects in absence of self-reports. In Section 4.c., however, I discussed that despite this impressive body of work, the absence of evidence from humans clearly showing that effects of psychoactive drugs in particular (and of any other sort of interoceptive information) must be consciously processed to be used as DSs must prevent us from inferring that those animals were definitely consciously processing the information used as DSs. Next I present a strategy to assess whether humans cannot use drugs as DSs unless they are conscious of any of the effects they induce.

Chapter 4

Meta-analysis: Testing the relationship between conscious processing of a stimulus and its ability to act as a DS

1. Introduction and overview

In this chapter I further evaluate whether subjects using interoceptive states as discriminative stimuli (DSs) in an operant discrimination paradigm need to be consciously aware of these DSs in order to successfully discriminate between them and a placebo (see Section 4 of the previous chapter). To do so, I present results from a meta-analysis of operant discrimination experiments in humans using the effects of different psychoactive drugs as DSs. The background to this meta-analysis is reviewed in Section 2, while Section 3 describes the methods used, Section 4 presents the results, and these findings are discussed in Section 5.

2. Meta-analysis of drug discrimination experiments: theoretical aspects

In Section 4 of the Chapter 3 we went through evidence from both human and animal operant drug discrimination studies consistent with the hypothesis that conscious processing of the internal states induced by different drugs used as DSs co-occurs with, and is perhaps even necessary for, success in these discrimination and generalisation paradigms. Among these pieces of evidence, Section 4.b. reviewed the frequent strong relationship between the results of the operant drug discrimination tasks and the subjective self-report questionnaires that human subjects typically fill out during these experiments: all studies investigating this found good correlations between performance in the operant drug discrimination task and several items of the subjective self-report questionnaires. However, these correlational results of course cannot address whether conscious awareness of the states induced by the psychoactive drugs actually *causes* the operant discrimination: they just show that these two

phenomena are closely related. Nor do they even demonstrate that operant drug discrimination cannot happen at doses for which there are no associated self-reports (SR) (and therefore conscious processing). Furthermore, some of the correlations between the operant drug discrimantion (DD) performance and SR performance were weak (see, for instance, Reynolds et al., 2013), and some SR thresholds were higher than DD ones (see Table 3 – Chapter 3), therefore suggesting that DD may sometimes happen without any related conscious processing.

To investigate in more detail how discrimination behaviour and subjective state inter-relate, we therefore carried out a meta-analysis on all published human experiments in which subjects trained to discriminate a dose of an active drug from placebo using a two alternative forced choice paradigm[36] (2AFC), are required to discriminate a number of smaller doses of the active drug dose (interspersed with placebo), while at the same time reporting their subjective state. From each of these experiments we extracted two threshold doses: a) a *discriminative stimulus threshold dose (DS threshold dose)*: the lowest drug dose that induced operant choices significantly different from those in the placebo condition; and b) *a self-report threshold dose (SR threshold dose)*: the lowest drug dose that yielded scores in the subjective self-report questionnaires that were significantly different from placebo. By comparing these two groups of threshold doses we hoped to achieve a clearer image of the relationship between these two metrics and thence evaluate whether the conscious awareness of the drug effects is always present when drugs act as DSs. The hypothesis, that success in drug discrimination cannot occur without conscious awareness of the drug's self-reportable effects, predicts that the two thresholds will significantly covary, and that the SR threshold doses will be smaller than or equal to the equivalent DS threshold doses. In contrast, if the two thresholds do not covary, and if drugs can act as DSs at lower doses than those generating subjective effects, we would reject the hypothesis by showing that the effects of drugs can be discriminated without conscious processing. We also pulled out data on DD – i.e, percentage of drug choice – performance at the SR thresholds to assess whether it is strong or weak

[36] A Two Alternative Forced Choice Paradigm (2AFC) typically sets a decision making condition in which a choice must be made between two responses based on limited information about which is correct – that is, which will be rewarded (see Bogacz et al., 2006).

at that point: it identifies levels of performance above which clear self-reported subjective effects of the drug should occur.

We used drug data in this meta-analysis because, to the best of our knowledge, studies using states of emotion as DSs (e.g., anxiety) have not been carried out in humans. However, we hoped that these analyses would be relevant to other sorts of interoceptive states, i.e. non-drug ones, when they are used as DSs in similar operant paradigms. Furthermore, these results could potentially shed light on the vast literature on DD in nonhuman animals also using 2AFCs paradigms.

3. Methods

3.1. Article search

Searches of the database Web of Knowledge© (Thomson Reuters, Toronto, Canada) were conducted on the 30th of July 2012, and updated on the 8th of August of 2015. Ten individual searches were run each time, using the following combinations of key topics in English language plus "NOT mice + NOT rats + NOT pigeons + NOT primates + NOT monkeys + NOT animals". They yielded a total of 211 papers as follows:

i. "Discriminative stimul*+ drug + self-report – 33 papers
ii. "Discriminative stimul* + drug + subjective – 66 papers
iii. "Discriminative stimul*+ interoceptive + self-report – 1 paper
iv. "Discriminative stimul*+ interoceptive + subjective – 2 papers
v. "Drug discrimination" + self-report – 34 papers
vi. "Drug discrimination" + subjective – 63 papers
vii. "Stimulus control"+ drug+ self-report – 3 papers
viii. "Stimulus control"+ drug+ subjective – 8 papers
ix. "Stimulus control"+ interoceptive + self-report – 0 papers
x. "Stimulus control"+ interoceptive+ subjective – 1 papers

3.2. Experiment selection

From these, all experiments that met the following two conditions were extracted:

1) Subjects underwent a double blind discrimination acquisition phase to be trained to discriminate a drug dose from placebo in a 2AFC.

2) Subjects who passed this acquisition phase then underwent a generalisation phase (see Section 3.b for an in depth explanation of the generalisation phase in DD experiments) consisting of both operant two alternative forced choice discrimination tasks, and self-report questionnaires. In these double blind generalisation tasks, subjects had to classify or compare different doses of the active drug used during the acquisition phase – interspersed with placebo trials – as either active drug dose or placebo. Again, this was done by using concurrent operant 2AFC and SR paradigms.

The resulting experiments numbered 79 in total, with 31.71% of these experiments using subjects addicted to the active drug. Twenty-four of these either directly reported the DD and SR threshold doses needed for our analysis, or provided data (e.g. figures) from which it was possible to calculate these thresholds. Each group of researchers involved in these 24 experiments was then contacted via e-mail to ensure we used all their studies that were relevant to our meta-analysis. The researchers contacted were: W. K. Bickel (Virginia Tech Carilion Research Institute, US), T. Duka (University of Sussex, US), J. A. Lile (University of Kentucky; US), K. A. Perkins (University of Pittsburgh; US), A. Oliveto (University of Arkansas for Medical Sciences, US), K. L. Preston (US National Institute on Drug Abuse – NIDA), and C. R. Rush University of Kentucky, US). This approach yielded 17 additional published experiments for inclusion in the meta-analysis (Oliveto et al., 1993; Kelly et al., 1997; Rush et al., 1997; Rush et al., 1998; Perkins et al., 1999; Rush et al., 2000; Rush & Baker, 2001; Rush et al., 2003; Lile et al., 2004; Lile et al., 2005; Stoops et al., 2005; Stoops et al., 2006; Vansickel et al., 2006; Sevak et al., 2009; Sevak et al., 2011; Lile et al., 2012; Lile et al., 2014).

Thus, our final dataset consisted of 41 pairs of thresholds, from experiments that used a wide range of psychoactive drugs as DSs, with diverse subjective effects. These included stimulants (caffeine, cocaine, nicotine, d-amphetamine, methamphetamine, and methylphenidate); a cannabinoid (delta-9-tetrahydrocannabinol); opioids (naloxone, hydromorphone); a GABA-A antagonist (flumanezil); an alcohol (ethanol); a benzodiazepine (triazolam); an imidazopyridine (zolpidem); and a barbiturate (pentobarbital).

3.2.1. Extracting the relevant values

When the DS and SR threshold doses were not reported in the papers but required calculating, we used the following procedure: values were extracted from graphs showing the scores for the relevant tasks, and the smallest doses significantly different from placebo in both the generalisation phase of the operant discrimination task and the self-report tasks were identified by statistically comparing subjects' scores for the different drug conditions to the scores for the placebo condition using either unpaired t-tests (when means and SEMs were reported), or paired t-tests when papers provided values for individual subjects (GraphPad Software Inc., CA, USA).

Across studies, these threshold doses showed a very great variation in magnitude because of the diversity of drugs used: the maximum dose (training dose) used ranged from 0.15 mg/70 kg (naloxone, Oliveto et al., 2002), up to 18900 mg/70kg (ethanol, Kelly et al., 1997 – reported as 0.45g/litre of body water (lbw), for an estimated 42 lwb in a 70kg subject). These proved unsuitable for parametric analyses even after log transformations, and were therefore standardised by dividing each dose tested by the maximum dose used in each experiment, while leaving the placebo condition as a value of 0. All values then fell on a 0-1 scale. (see https://www.biomedware.com/files/documentation/Preparing_data/Methods_for_data_standardization.htm).

The following information was also extracted from each paper, to help us understand both their limitations and their relevance to animal studies. The percentage of drug option choices – i.e., the percentage of drug option choice at the group level – at the DD threshold and at the SR threshold were calculated for each experiment. After checking for normality, their means and 95% confidence intervals

were then calculated. Finally, we noted the number of SR items reported in the methods of each of these experiments, versus the number of items shown in the results that enabled us to calculate the thresholds above.

3.3. Statistical Analyses

3.3.1. Regression: can DD discrimination performance predict SR discrimination?

The SR thresholds were regressed on their respective DS thresholds. SR was chosen as the dependent variable, since we were interested in investigating whether SR can be predicted from DD choice. This analysis was conducted in JMP 12.0.1 (SAS Institute Inc., NC, USA, 2013), using general linear models (GLMs). Addiction status (i.e., whether subjects in each experiment were addicted to the drug they were trained to discriminate from placebo) was included as a blocking factor, to assess whether there are significant differences in DD and SR discrimination performances according to whether subjects were dependent on the drugs they were trained to discriminate. This decision was based on findings showing that drug addicts have been shown to perform simple operant actions (pressing a lever) for self-administering drug doses they report not to feel (Lamb et al.,1991; Comer et al., 2008; see also Chapter 2, Section 3), therefore suggesting that these group's DD capabilities may differ from those of non-addicts. Log_{10} transformations were applied to both groups of thresholds to achieve normality of the residuals.

Additionally, we assessed whether the slope of the regression was significantly different from 1, in order to assess whether the DD thresholds were not significantly smaller than the SR thresholds (a finding that would reject out hypothesis). For this, as for all statistical analyses, tests were two-tailed and the conventional probability value of $P<0.05$ was chosen to determine significance.

3.3.2. Tests of Differences

a) Are DD and SR thresholds significantly different?

Neither Arcsine Square Root, Logit, Log, nor Box-Cox transformations were effective for achieving normality of the residuals in a GLM with addiction status was included as a blocking factor**,** and so the nonparametric Wilcoxon signed rank test was used to compare the thresholds (GraphPad Software Inc., CA, USA).

b) Are the percentages of drug option choices at the DD and SR thresholds significantly different?

A paired t-test was conducted in JMP 12.0.1 (SAS Institute Inc., NC, USA, 2013) to assess whether the percentages of drug option choices (i.e., selection of the drug option in the operant task) at the DD and SR threshold doses were significantly different.

c) Are the percentages of drug option choices at the SR thresholds significantly different from 50% (i.e., above chance or guessing levels)?

A one sample t-test was conducted in JMP 12.0.1 (SAS Institute Inc., NC, USA, 2013) to assess whether the percentages of drug option choices at the SR thresholds were significantly different than 50%.

3.3.3. Bland & Altman Agreement Method

Additionally, we used a test suggested by Bland and Altman (1986) (see also Kwiecien et al., 2011) to further assess agreement between two variables. This method consists of plotting the differences between each pair of variables (DD threshold minus SR threshold) against the average of each pair. When 95% of the plotted values fall within the limits of agreement – defined as the mean difference plus and minus two standard deviations of the differences – it is considered that there is good agreement between the two variables. This result would support the

prediction that the two groups of thresholds are not different. Each difference falling above agreement limit represented by the mean difference plus two standard deviations would show that the SR is smaller than the DD one; on the other hand, each difference falling below the limit represented by the mean difference minus two standard deviations would show that the DD is smaller than the SR one.

4. Results

Across the 41 studies, results were available for an average of approximately 22% of the items that had been assessed via SR questionnaires (see Table 4). Table 4 shows the percentages of drug option choices at the DD and SR thresholds for 40 of the 41 experiments (for one study [Kelly et al., 1997] it was not possible to calculate these figures since data were not provided and no relevant graph was shown). The mean at the DD threshold was 65.59% (with a lower 95% CI of 59.61% and an upper 95% CI of 71.56%), and the mean at the SR threshold was 65.45% (with a lower 95% CI of 58.19% and an upper 95% CI of 72.71%).

Table 4 – Drug option choices at DD and SR thresholds / Number of SR items per experiment, ordered by percentage of drug choice at the SR threshold.

Nº	**Study**	**% of drug choice at DD threshold**	**% of drug choice at SR threshold**	**SR items reported in a way we could use/all SR items measured per study**
8	Preston et al. (1992)	100	100	4/15
22	Oliveto et al. (2002)	100	100	6/16
26	Oliveto et al. (1993)	100	100	3/15
21	Smith & Bickel (1999)	97.38	97.38	6/14

3	Smith & Bickel (2001)	90	90	1/16
37	Stoops et al. (2006)	58.97	89.48	6/27
38	Stoops et al. (2005)	66.38	87.82	8/27
18	Rush et al. (2002)	84.62	84.62	4/71
35	Rush et al. (1997)	84.39	84.39	3/39
30	Lile et al. (2004)	83.86	83.86	2/48
25	Perkins et al. (1999)	83.35	83.35	4/8

5	Lile et al. (2005)	80.98	80.98	4/48
28	Vansickel et al. (2006)	59.55	76.92	4/38
20	Preston & Bigelow (1998)	73.56	73.56	7/25
32	Rush et al. (2001)	76.39	76.39	3/39
24	Duka et al. (1998a)	48.83	76.38	3/6
31	Rush et al. (2003)	71.08	71.08	4/37
1	Perkins et al. (1996)	70.79	70.79	5/12
39	Sevak et al. (2011)	70.64	70.64	7/27
17	Rush et al. (2003)	49.84	68.77	4/48
23	Duka et al. (1998b)	37.24	68.46	2/13
29	Lile et al. (2005)	68.42	68.42	6/48
11	Lile et al. (2011)	65.50	65.50	6/48
14	Lile et al. (2009)	61.56	61.56	4/28
36	Lile et al. (2012)	60.95	60.95	3/20
9	Perkins et al. (2005)	60.23	60.23	8/13
2	Perkins et al. (1996)	45	60	5/12
16	Rush et al. (2004)	79.21	55.98	4/48
12	Lile et al. (2011)	51.85	51.85	4/20
13	Lile et al. (2010)	50.16	50.16	4/20
10	Lile et al. (2012)	50.06	50.06	4/20
7	Preston & Bigelow	50	50	6/15

	(2000)			
27	Sevak et al. (2009)	49.82	49.82	4/48
40	Lile et al. (2014)	64.28	40	4/20
6	Perkins et al. (1997)	36.6	36.6	2/11
19	Perkins et al. (1999)	25.14	32.43	6/11
34	Rush et al. (1998)	53.24	32.06	8/47
4	Oliveto et al (1992)	49.71	28. 46	3/14
15	Lile et al. (2006)	69.06	21.33	4/48
33	Rush et al. (2000)	44.8	7.92	4/39
41	Kelly et al. (1997)	–	–	2/16

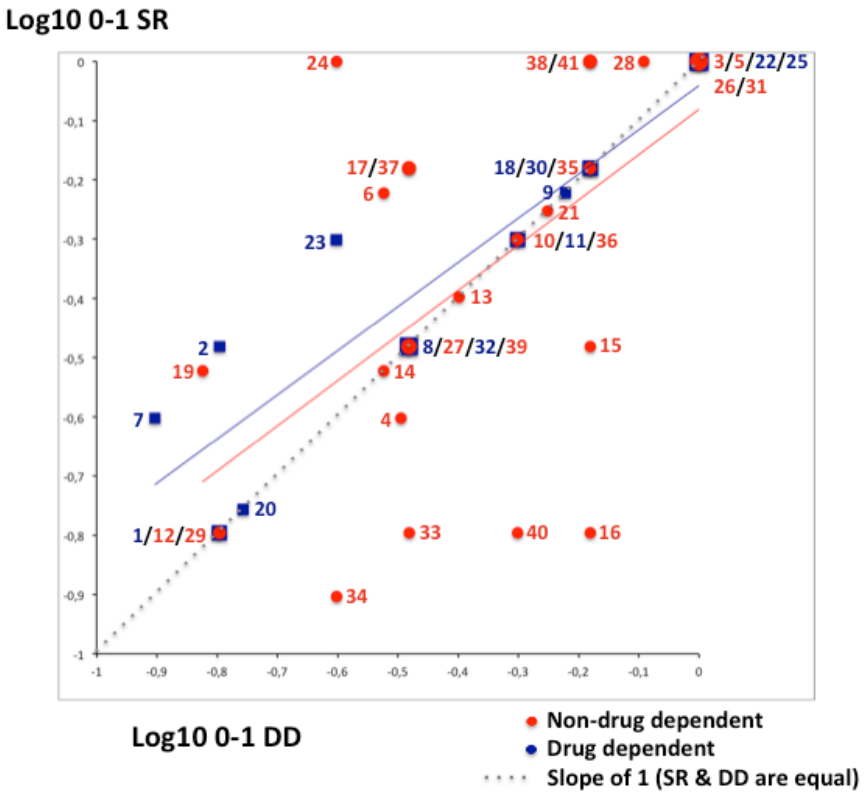
In all 40 studies, subjects at the group level showed significantly altered SR from the placebo condition when the percentage of drug choices made in the DD task was at 100%. In 36 of the 40 – thus 90% of the studies in this meta-analysis – SR was significantly altered above placebo levels if the percentage of drug choices made in the DD task was above 90%. In 28/40 studies (70%), SR was altered if the drug choice was above 80%, whereas in 52.5% of the studies (21/70), SR was altered if the drug choice was at least 70%.

4.1. Regression of SR thresholds against DD thresholds

Regressing the SR thresholds on the DD thresholds revealed significant covariation ($F_{1,37} = 32.9$, $p < 0.0001$; see Fig. 1). The R^2 of the model was 0.48. There was no significant main effect of addiction status ($F_{1, 37} = 0.40$, $p = 0.53$), nor was there an interaction between addiction status and DD threshold ($F_{1, 37} = 0.004$, $p = 0.95$), showing that the nature and slope of the relationship between the two thresholds was unaffected by this variable (next page).

Figure 2 – Log_{10} 0-1 SR (threshold doses for subjective self report/maximum dose in each study) regressed on Log_{10} 0-1 DD (group of threshold doses for the drug discrimination task/maximum dose in each study)

Each number represents one experiment as labelled in Table 2. The red dots are experiments where subjects were non-dependents. The red line represents the regression line for the non-dependents.



The slope of the regression line (addicts and non-addicts combined) assessed using a JMP custom test was not significantly different from 1 ($F_{1,38} = 1.76$, $p = 0.1853$).

4.2. Tests of Differences

a) DD and SR dose thresholds

The Wilcoxon Test found no significant difference between the pairs of thresholds ($W = 50$, $n = 41$, $p = 0.2113$).

b) Drug option choices in the operant task at DD and SR thresholds

The paired t-test found no significant difference between the percentages of drug option choices task at the DD and SR thresholds ($t(39) = 0.0525$, $p = 0.9584$).

c) Drug option choices at the SR thresholds compared to chance.

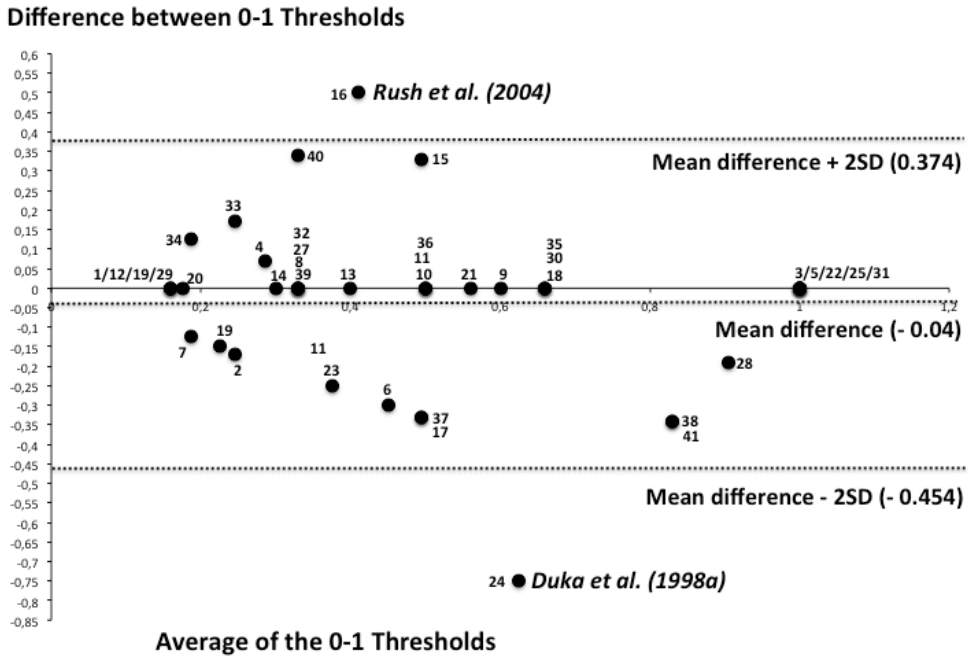
The one sample t-test found a very significant difference ($\mu > 50$) between the percentages of drug option choices at the SR threshold and a score of 50% ($t(39) = 4.3066$, $p < 0.001$).

4.3. Bland & Altman Agreement Method

The Bland & Altman plot showed a good agreement between the two kinds of thresholds. Only 2 of the 41 differences (Duka et al., 1998a, Rush et al., 2004) fell outside the agreement zone, and therefore 95.12% of our values fell within these limits (see Fig. 3). Furthermore, for only one of these two differences (Duka et al., 1998a) was the DD threshold smaller than the SR one (suggesting that DD could happen at doses at which there is no SR effect). The result for Rush and colleagues (2004) suggests the contrary, i.e., that subjects can self-report the drug effects at doses at which they cannot discriminate this drug from placebo in the operant DD task. It is important to underline, however, that a pre-requisite for a valid use of this test is that the differences between the two types of variables are normally distributed. The differences between our DD and SR thresholds did not show normality of the residuals, and, importantly, were not possible to normalize despite using several transformations (namely, Arcsine Square Root, Logit, Log, and Box-Cox).

Figure 3 – Bland & Altman test of agreement

Each number represents an experiment, as labelled in Table 1. The differences were calculated by subtracting the SR threshold value from the DD one.



5. Discussion

Our aim was to investigate whether the successful use of interoceptive states as discriminative stimuli is paralleled by the conscious awareness of those states. We focussed on drug discrimination at threshold levels of deviation from placebo, and used two thresholds as our metrics: the lowest dose at which significant drug discrimination behaviour occurs (the DD threshold), and the lowest dose at which significant changes in self-reported states occur (the SR threshold). We had predicted that the two thresholds would significantly covary, and that the SR threshold doses would be smaller than or equal to the equivalent DS threshold doses. These predictions were met. Here we discuss the potential implications of these results for humans and animals, as well as the limitations of our data.

When the SR thresholds were regressed on the DD thresholds, we found that they significantly covaried, showing that the two thresholds are closely related. Furthermore, the two thresholds were so similar as to be statistically indistinguishable. Thus the regression line's slope was not significantly different from 1, revealing a 1:1 relationship between the two thresholds. The SR and DD thresholds were not significantly different from each other in a test of differences; and the percentages of drug option choices at the DD and SR threshold were not significantly different. The Bland and Altman method to assess agreement was also consistent with this pattern (though the fact that the differences between the two thresholds were not normally distributed suggests prudence when interpreting this piece of evidence). Overall, 24 of the 41 differences had a value of 0: i.e. in 24 studies the two thresholds occurred at the same dose, and 39 values fell within the limits of agreement. Moreover, in only one of the differences falling outside these limits of agreement was the SR threshold substantially greater than the DD threshold. Finally, we saw that at high levels of drug option choice, the DS are higly likely accompanied by reportable feelings induced by these drugs.

Overall, these results cautiously suggest that, in general, groups of human subjects are typically unable to use drugs at DSs if they are below doses that can be subjectively detected, and, conversely, that when groups of subjects successfully use drugs as DSs, these are likely paralleled by significantly detectable subjective effects. This hypothesis is particularly supported for those groups of subjects performing above threshold levels, at high levels of drug choice in the 2AFC task. Groups of subjects with drug option choices above 90% were highly likely to be experiencing subjective effects from the drug, and in every single study, groups of subjects with drug option choices at 100% *were* experiencing subjective effects induced by the drug

This result joins previous findings that across individual subjects within such drug DS studies (see Chapter 3, Section 4.b), subjective self-reported states and the outcomes of the discrimination tasks are closely related. It is also cautiously consistent with the causal hypothesis advanced by Díaz and Velázquez (2000) that the conscious processing of a drug's subjective effects is *necessary* for humans to discriminate psychoactive drugs in a DS operant task. Importantly, the group of experiments in our meta-analysis account for a very wide range of psychoactive drugs, from caffeine to opioids, and were carried out by different research groups using operant drug discrimination techniques that varied slightly from study to study

(typically by the number of points or items available to be distributed between the two choices, and by the time available for the task). Moreover, the addiction status of subjects (i.e., whether dependent on the drug they had to discriminate from placebo) did not affect this discrimination pattern.[37] This indicates that our findings have wide generality across diverse subjects and psychoactive drugs. It also suggests that the use of other interoceptive stimuli as DSs may perhaps be similarly tightly related to – if not dependent on – the conscious processing of such stimuli.

These findings can be interpreted in the context of one of the hypothetical roles of conscious processing: making non-routine bindings, those that require creation of unforeseen combinations (see Chapter 2 – Section 2). Learning to discriminate drug conditions from placebo, and linking this knowledge to an operant action in order to be rewarded, may thus be an example of this, depending on conscious processing. Perhaps, then, success in operant 2AFC drug discrimination paradigms could be used as a biomarker of conscious processing in the absence of verbal reports (see Chapter 2 – Section 4). This hypothesis has obvious implications for the animals used in drug discrimination studies (and perhaps even all studies where interoceptive cues are used as DSs). If in every other way, *"drug discrimination in animals closely matches human drug discrimination"* (Diaz & Velazquez, 2000; see also Chait et al., 1985 & Oliveto et al., 2002), this suggests that similar changes in drug-induced subjective state are manifest within groups of rats picking the drug option in similar tasks. This seems especially likely given that animals are typically not deemed to have generalised to the drug until their choices for the drug option are at 80-85% or greater. Our analyses indicate that in human studies, even doses causing lower levels of drug option choices (between just 58.90% and 73.06%) are often high enough to cause significant changes in self-reported feelings at the group level. They further show that subjects choosing the drug option at 90% or more – as has been seen in some animal studies (see Chapter 3) – are highly likely to be experiencing reportable subjective drug effects.

[37] As we have seen in Section 3 of Chapter 2, some studies have shown that drug dependents can perform operants – e.g., press a lever to be rewarded with drug doses that they report not to feel. Our results suggest that this is not the case when subjects must discriminate drugs by performing a 2AFC task.

5.1. Discussion points and limitations of the data

We must acknowledge, however, the limitations of our study. First, the results are correlational, and so they cannot show any causal relationship between DD and SR: they just show that across the groups of subjects used in the experiments in our meta-analysis, significant DD typically does not seem to happen at doses not eliciting drug-related SRs. Secondly, the analyses only use the overall behaviour of groups of subjects, and so cannot reveal what proportion of individual subjects follow this pattern. Thirdly, it is important to note the very moderate R^2 value: less than half the variation in the SR thresholds was explained by variation in DD. In fact, the R^2 value for this cross-study work was lower than in some within study analyses. For instance, Oliveto and colleagues (1988) found that during a generalisation phase of a cocaine DD study, the SR item "similar to cocaine" correlated best with subjects' DD performance ($R = +0.84/R^2 = 0.706$) (see Chapter 3, Section 4.a.). This suggests either that the data in this meta-analysis are noisy, and/or that the relationship between SR and DD varies with drug and/or the SR items measured in each study. Relatedly we found many undeniable counter-examples to our predicted pattern, and indeed in 11 studies (Perkins et al., 1996; 1997 & 1999 – nicotine nasal spray; Kelly et al., 1997 – ethanol p.o.; Duka et al., 1998a – ethanol p.o. – &1998b – nicotine chewing gum; Preston & Bigelow, 2000 – hydromorphone i.m.; Rush et al., 2003 – d-amphetamine p.o.; Stoops et al., 2005 – methylphenidate p.o. – & 2006 – d-amphetamine p.o.; Vansickel et al., 2006 – triazolam p.o.) the DD threshold was lower than the SR threshold – such that subjects were apparently discriminating the drug without showing detectable changes in subjective effects. This was most marked in one study (Duka et al., 1998a), where the DD threshold was so much lower than the SR threshold that it fell outside the Bland-Altman limits of agreement. Fourth and finally, the mean percentage of drug option choices in the operant task at SR threshold doses (58.90-73.06%) was significantly above 50%: thus as drug doses increased, subjects' behaviour changed from primarily choosing the placebo option to above chance drug choices without any detectable changes in SR feelings.

What we cannot know about both these last two patterns is whether they represent measurement error, or instead are genuine cases of blindsight-like responding (see Chapter 2, Section 2). This meta-analysis certainly cannot rule out

the possibility that at least some individuals are capable of discriminating some types of drugs in a 2AFC paradigm without feeling their effects. However, there are also potentially methodological reasons for these puzzling patterns, and for why we only found a broad association between SR and DD. First, although the scores and results for the most sensitive SR items were typically reported in the papers we used, we could not present a *full* picture of SR because more than 75% of the full range of SR items measured were not reported, and so were unavailable for our analyses. Thus, we cannot rule out that this missing SR data could therefore perhaps explain why so much DD behaviour apparently occurred without accompanying changes in feelings. Second, another methodological limitation was the gaps between the drug doses tested in the generalisation phase. That doses went up in steps instead of incrementally means that the exact thresholds for DD and SR may not always be captured, adding noise to the data.

Chapter 5

Conclusions and future directions

1. Introduction

In this final chapter I will summarize the main points presented throughout this thesis, discuss the conclusions we can draw from them, and consider some future potential research directions for those interested in identifying potential markers of conscious emotion (i.e., sentience). In Section 1 I will discuss the use of the drug discrimination two alternative forced choice paradigm ('DD 2AFC') as a potential behavioural marker of conscious processing of interoceptive states – including emotions – in verbal humans. Section 2 will focus on what can we conclude regarding the performance of non-verbal subjects in this DD paradigm. Finally, Section 3 will deal with a broader question: the general merits of using verbal humans as models for investigating consciousness in non-verbal beings, and in particular in non-human animals.

2. Drug discrimination as a potential marker of the conscious access of emotion and other interoceptive states in verbal humans

In Chapter 2 it was argued the use of emotions or other interoceptive states as discriminative stimuli (DSs) in an operant paradigm might require consciously accessing these states. This is the type of arbitrary learned behaviour involving a "novel binding" (Dehaene, 2014), which is likely to depend on conscious accessing the relevant information for performing the discrimination, and therefore could be a potential tool for identifying states of conscious emotion in the absence of verbal report. In order to assess whether this is the case, throughout this thesis we took advantage of verbal humans' ability to self-report to investigate whether the use of interoceptive stimuli as DSs requires conscious processing of this information. This

was based on the assumption that one can extrapolate from drug-induced changes in affect, to 'pure' or true affective states (something revisited in Section 1.5.2, and in 2.1 for the case of non-human animals). As stressed in Chapter 2, self-report is the only tool we have for validating any non-verbal process (e.g., behaviour or pattern of brain activation) as dependent on or correlating with conscious processing of information.

2.1. Verbal humans are unlikely to use interoceptive stimuli as DSs if these are not consciously accessed

In Chapter 3 I argued that in the absence of human studies using affective states as DSs, the DD paradigm offers the closest possibility of investigating this question. There we saw that it has been suggested that humans discriminate and generalise drugs (in practice typically recreational ones) according to their interoceptive effects, as assessed by the self-report (SR) questionnaires that subjects usually fill right after undergoing the DD operant task (see Chapter 3 for an in depth review of the DD paradigm). I reviewed those papers empirically assessing the relationship between the two tasks' outcomes (DD and SR), which found that they are typically closely related. However, these studies 1) did not investigate whether DD can ever happen without any associated SR (for example at very low doses); and 2) as a consequence of the correlational nature of these analyses, they could not test the hypothesis that conscious processing of the drug effects guides the behavioural discrimination (DD).

Chapter 4 further explored this question by conducting a meta-analysis of those DD studies in humans conducting a two alternative forced choice (2AFC) operant discrimination paradigm. This assessed the relationship between the DD threshold doses (the lowest drug doses that induced operant choices significantly different from those in the placebo condition from each experiment) and SR threshold doses (the lowest drug doses that yielded scores in the subjective self-report questionnaires that were significantly different from placebo in each experiment). The results confirmed that these two tasks yield very similar discrimination outcomes: the two types of thresholds (SR and DD) co-varied and were not significantly different, thus cautiously suggesting that drug doses that subjects behaviourally

discriminate as significantly different from placebo also elicit reportable effects (which are therefore consciously accessed). Moreover, it was found that the percentage of drug option selection at the DD and SR threshold doses were not significantly different: the mean percentage of selecting the drug choice option in the task at the DD threshold was 65.59% (with a lower 95% confidence interval [CI] of 59.61% and an upper 95% CI of 71.56%), and the mean at the SR threshold was 65.98% (lower 95% CI: 58.90 - 73.06%). Thus, 1) according to these results, and 2) considering that in all 40 of the studies in the meta-analysis, 100% DD performance was accompanied by significant self-report and in 90% of these studies 90% of DD performance was also accompanied by significant self-report, any subject in DD task selecting the drug option above 90% of DD performance is highly likely to be consciously accessing subjective information relative to the DS. Altogether, these results suggest that verbal humans need to consciously process the interoceptive effects of drugs – which are often reported as of emotional nature – in order to use them as DSs in this paradigm, especially when making very strong use of the DD option. Further below, however, I discuss some limitations of these findings.

2.2. Using the DD 2AFC for identifying conscious processing of interoceptive effects in the absence of verbal reports: limitations of the meta-analysis

Several limitations of Chapter 4's meta-analysis are relevant regarding the use of the DD 2AFC paradigm as a behavioural marker of the conscious processing of interoceptive states. The first three are methodological, while the fourth and fifth emerged from the results themselves, as follows:

Limitation 1): The correlational nature of its results meant that they could not say anything about the causality of the relationship between DD and SR. Thus, from them we cannot know whether consciously accessing the interoceptive effects elicited by these drugs was actually causing the discrimination – our hypothesis – or just correlating with it. This limitation, however, does not prevent us from using this behavioural task as a tool or marker for identifying conscious processing of these interoceptive effects in absence of verbal reports; *Limitation 2)*: These analyses focused solely on the threshold doses and the related DD performance, preventing us from seeing how SR and DD values relate to each other at weaker or stronger doses or

degrees of generalisation. In order to investigate the relationship between DD and SR across all the doses used in a typical DD 2AFC experiments, we therefore plan to analyse data from individual subjects from four DD studies on amphetamine-related drugs, as will be discussed in Section 2.4.1.

Limitation 3): Although the SR questionnaires are designed to specifically capture the effects of the different groups of psychoactive drugs, it is possible that some subjects could consciously process the effects of drugs but either misunderstood the questions in the SR task, or not have their experiences captured by the small subset of SR items (around 25% of all those measured) that were actually available to us for analysis. Both of these effects would cause 'false negative' inferences, as discussed below in 1.3. The problems with relying on language in these self-report tasks is also something I consider in Section 3, since it is a general weakness of our reliance on verbal self-report.

Limitation 4): The analyses revealed that the regression relationship between the two thresholds, although significant, had a rather low R^2 value: thus there was considerable 'scatter' around the regression line, and only a small fraction of the variance in SR could be inferred from DD behaviour: not something very compatible with SR playing a causal role in the discriminative behaviour.

Limitation 5): Relatedly, some of this scatter arose because in some studies the threshold for DD was lower than the threshold for SR. In fact, in 11 studies the DD threshold occurred at a lower dose than the SR threshold, such that at this dose subjects were apparently discriminating the drug without experiencing any detectable changes in subjective effects (perhaps in a blindsight-like manner, as reviewed in Chapter 2). If one wished to use DD behaviour to infer changes in subjective state, and the lack of self-report did reflect a genuine lack of self-awareness (rather than mere measurement error: a possibility suggested in *Limitation 3)* and also returned in Section 3), then these 11 results would therefore be examples of false positives (or Type I errors).

This type of deviation between the two thresholds was most marked in one study on ethanol discrimination (Duka et al., 1998a), where the DD threshold was substantially lower than the SR threshold. To explore this puzzle further, I therefore plan to investigate whether this result does actually show that verbal humans can discriminate this drug at the threshold level without being able to report any drug effect. I will do this by conducting a DD study using ethanol with Professor Dora

Duka at the Sackler Centre for Consciousness Science (University of Sussex, UK) (see Section 2.4.2. for more details);

2.3. Potential false negatives (Type II errors) when inferring the conscious access of drug effects from behaviour: Subjects not acquiring the discrimination despite being aware

As is commonly reported in the DD studies presented in Chapters 3 and 4, not all the subjects undergoing 2AFC DD tasks acquire the discrimination between the two conditions in the acquisition phase. Obviously, this does not mean that these subjects agreeing to take part in a lab test are non-sentient, and so it may represent a false negative or type II error. Thus, although this discrimination method can reveal whether a subject is capable of consciously processing the effects of drugs (and potentially other non-drug induced interoceptive stimuli used as DSs), and therefore sentient, the obverse is not true: a subject could fail this task and yet still be sentient.

In the case of any verbal subject *not* acquiring the discrimination, the underlying reasons should therefore be further investigated by taking advantage of the SR drug questionnaires typically used in these studies, in order to quantify whether or not the subject can detect the training dose. In an early DD study specifically addressing this question, for instance, Chait and colleagues (1985) found that in those subjects not acquiring the amphetamine versus placebo DD, the dose of amphetamine produced significant effects relative to when on placebo on nearly every SR item of the questionnaires used. Interestingly, although those subjects acquiring the discrimination were more sensitive to the subjective effects of amphetamine than the non-discriminators, this difference in sensitivity reached statistical significance only for ratings of the SR item "hungry" of the drug effect questionnaire (see Chapter 3 – Appendix table 1 for the SR questionnaires most often used in DD studies). According to the researchers, these "subjects did not differ significantly in their response to amphetamine on other subjective scales [used in this study], probably because of the relatively small number of subjects in each group", and pointed out that additional "findings (unpublished) from our laboratory indicate that discriminators are generally more sensitive than non-discriminators to the subjective effects of amphetamine" (Chait et al., 1985). Thus, in this case, the non-

discriminators' failure to learn the DD task despite feeling the drug effects is clearly a type of false negative with regard to the training dose used. One perhaps reflecting that the effects of the drug dose were not strong enough to control the discrimination behaviour (see 'stimulus control' in Chapter 3). In other words, "it could be that some doses do not produce effects that are discriminable in the DD task" (William W. Stoops, personal communication). In Section 2, however, we will see that in non-human animals studies this could be also attributed to cognitive deficits).

However, if subjects *cannot* detect any subjective effects of the training dose (such that experimentally this was a true negative, not a false one), then it might be fruitful to test higher dose/s of the active drug in the acquisition paradigm, since the previous test conditions (drug doses) evidently did not induce effects strong enough for use as DSs by these subjects. Even though these considerations may perhaps seem trivial in the case of verbal humans, they become particularly relevant when investigating conscious processing of these cues in non-verbal subjects (see Section 3).

2.4. Next steps: overcoming the limitations of our meta-analysis

2.4.1. The use of existing data from individual subjects across all doses and SR items used

In order to solve some of the problems set out in Section 2.2, we expect to further explore how DD and SR responses relate to each other during the generalisation phase of 2AFC drug discrimination studies by analysing data from four studies on drugs related to amphetamine from 50 individual subjects (studies # 27, 31, 38 & 39 in Chapter 4's Table 4). These data were collected at the laboratory of Professor Craig R. Rush (University of Kentucky, USA). Unlike the data used for Chapter 4's meta-analysis, they reveal the raw DD and SR scores for each individual subject, across all the doses tested in the generalisation phase, and for all the SR items used in these studies (not just a small subset as we were limited to in Ch. 4). Although the planned analyses are still essentially correlational and so will still not directly address the question of causality discussed above, we aim to achieve a better understanding of the closeness of relationship between the DD and SR by assessing

the extent to which variation in the SR item scores predicts the variation in DD behaviour, both within and between subjects. We will investigate this for the whole set of doses together, and for different levels of DD performance (i.e., percentages of drug option choices) – since perhaps, for example, the relationship is tighter than that suggested by the threshold dose study at stronger drug doses that people can identify more very confidently, and/or at DD drug option choices of 80% or more.

Moreover, we will look at the predictive effect of SR items not individually but also combined. This is a refinement on how such analyses are usually run, and is important because the different subjects may rely on a different SR item, or even combinations of SR items, for discriminating the drug from placebo at each of the different doses. Thus, by combining those SR items found to best predict DD, we expect to find the model that best predicts the variation in DD behaviour across all doses (compared to the results for these SR items alone).

One complete, if these analyses find weak relationships (e.g. R^2 values of under 25% – Rebecca Meagher, personal communication) explaining only some of the variation in behaviour, this would be inconsistent with subjective states (i.e., consciously accessing these states) playing a causal role in discrimination responses. In contrast, a strong relationship in which most or even all of the variation in behaviour was explained by variation in self-reported state (e.g. R^2 values of over 50% – Rebecca Meagher, personal communication), this *would* be consistent with a causal role.

It is important to underline, however, that these data – the only raw data I could gather after contacting the seven key authors conducting the studies identified in the article search presented in the Chapter 4 – are on amphetamine-related drugs only. Ideally I would have liked to have access to raw data from a wider range of psychoactive drugs (as in the Chapter 4's meta-analysis) so the results would have generality across diverse psychoactive drug classes. This is another advantage of my second planned study, based on ethanol, which I describe next.

2.4.2. Tackling the question of causation: working with positive emotion

In order to tackle the question of causality (as well as to generate a new dataset rich in individual-level data across a range of drug doses), we plan to run our own experiment with humans at the University of Sussex School of Psychology and Sackler Centre for Consciousness Science with professors Dora Duka and Zoltan Dienes. This study has two important elements. First it will specifically focus on emotional feelings, and investigate whether subjects can relate changes in affect induced by drugs, to non-drug induced affective states. Via the experimental manipulation of emotional feelings, it will also investigate whether such feelings play a causal role (thus an obligatory role) in an operant 2AFC task. Ethanol was chosen as the drug because it typically induces positive affect (Dora Duka, personal communication); it is very different from the amphetamine-like drugs we already have good data on; it is much easier to acquire and work with than illegal recreational drugs; and finally, we would like to resolve an issue mentioned in Section 2.2. By replicating the Duka et al. (1998a) study which found the DD threshold to be much smaller than the SR one, we hope to investigate the reasons why this occurred, and determine whether subjects can really discriminate the ethanol threshold dose from placebo without reporting any of the subjective effects typically induced by this drug (thus in a blindsight-like manner).

In our planned study, in combination with the various ethanol doses – which will range from zero to the training dose – subjects will be exposed to two pictures of facial expressions (IAPS well-validated tools to influence mood) (Goeleven et al., 2008; Ebner et al., 2010): one previously shown to boost happiness, one shown to induce sadness. At each dose, each subject will be run through the DS operant task, plus asked to self-report her feelings. The hypothesis that emotional states can only be used as DSs in an operant discrimination task if subjects are consciously accessed, playing a causal role in the subject's operant behaviour, then makes the following testable predictions: 1) That subjects pre-trained to discriminate ethanol from placebo in a DD 2AFC operant task will choose the ethanol option key more often when also reporting positive emotional states; 2) that modifying these states with the validated emotional pictures of high positive and negative valence will respectively increase and decrease subjects' choices of the ethanol operant; 3) the SR

increases in positive emotion will statistically predict the increases in ethanol response selection.

By working with human subjects in this way, collecting data on how they feel as they perform these types of tasks (SR), and also experimentally manipulating how they feel, we will thus rigorously test the hypothesis that consciously accessing the internal states actually guides the discrimination of these states. If this hypothesis is confirmed, this would a) show that similar past studies using animals provide strong evidence that these species have conscious emotions; and b) validate this task as a tool for identifying whether other non-verbal subjects have conscious emotions and are thus sentient (see also Section 2 below).

3. The case of non-verbal beings

3.1 What can we conclude about non-verbal beings' performance in the DD 2AFC paradigm?

In Section 3 of Chapter 3 we saw that different species of non-human mammals and birds, can succeed in using drugs (among other interoceptive cues) as DSs in operant discrimination paradigms. Importantly, there it was argued that their performances are similar to that of verbal humans undergoing similar DD tasks since: 1) they only discriminate psychoactive drugs, i.e., drugs that induce subjective effects (as self-reported by verbal humans); and 2) they typically generalise between drugs that induce similar subjective feelings in humans (and thus suggesting these states are likely to be consciously processed and have a causal role in the discrimination behaviour.

Furthermore, we saw that many of these drugs elicit changes in affective states in humans (e.g. positive states, as engendered by recreational drugs tested in human DD studies, but also negative states, like the anxiety caused by PTZ). Because of the strongly emotional nature of these experiments, we emphasised the many examples of rats discriminating between this last drug, as well as other drugs that humans typically report to induce anxiety. We also emphasized how well rats generalise from e.g. PTZ to non-drug induced states which are potentially anxiogenic for this species, such as

the presence of a predator or an agonistic defeat by a rival rat. Although less conclusive, we also reviewed some similar examples in pigs.

In the light of the results from verbal humans discussed above, what more can we infer from these animals' performance, regarding the potential conscious processing of the discriminated states? As argued in Section 3 of Chapter 3, the acquisition and generalisation criteria in these animal studies are often more stringent than in those on verbal humans (often up to 90% selection of the relevant choice in the 2AFC task). We found in our meta-analysis that humans discriminating drugs at about 90% drug option choice levels are very likely to be consciously processing some of the effects induced by these drugs. We do not have strong reasons for thinking that the case of these non-human animals is different, especially as they are often showing far stronger DD behaviour. Thus, it is likely that not only the rats and pigs, but also the mice, gerbils, rhesus monkeys, squirrel monkeys and pigeons (Kamien et al., 1993) known to succeed in DD/interoceptive DS tasks, are thus "self-reporting" interoceptive states of which they are consciously aware.

3.2. Potential false positives and false negatives when inferring the conscious processing of drug induced interoceptive stimuli in non-verbal beings

3.2.1. False positives (Type I errors)

In the case of non-verbal beings, we obviously lack the possibility of comparing operant discrimination results with self-reports as we can in verbal humans. It is therefore possible that non-verbal beings discriminate drugs according to exteroceptive effects induced by the actions of these drugs – e.g., subtle limb tremors instead of the feeling of anxiety – thus representing a false positive. It is also possible that some individuals can show DD responses, even at high drug choice levels, with no significant change in their underlying subjective state (like the Duka alcohol example seemed to illustrate for humans at threshold DD performance; see above). This too would be a false positive.

However, I suspect this type of false positive is unlikely (even in humans), especially given the similarity in the manner in which animals discriminate drug-

induced stimuli that verbal humans report to discriminate due to their subjective effects.

A crucial point with regard to the interpretation of these results has to do with the phylogenetic closeness to verbal humans (our gold standard) (see Chapter 2, Section 4). The closer in phylogeny to humans is the non-verbal subject, the better arguments we have for thinking that he or she is likely to experience something homologous to what verbal humans report to feel when performing the DD task. In practice, these subjects (especially the mammals) all share with humans similar neurophysiological, physiological, and behavioural traits derived from a shared common ancestor (see Chapter 2, Section 4). The risk of subjects relying on different discrimination mechanisms increases as we consider the case of animals of species more distant in phylogeny (i.e., invertebrates). However, in the absence of evidence suggesting that this may be the case, any subject performing similarly to verbal humans in the DD task (i.e., showing above 75% of drug choice) should in principle be considered likely to process the drug effects at a conscious level, regardless of the species.

3.2.2. False negatives

What can we conclude when non-verbal beings do *not* acquire the discrimination, or do not perform at DD levels that in verbal humans are likely to be paralleled by – if not caused by – conscious processing of the effects induced by the drugs? The obvious problem regarding that lack of self-report applies here too. Before inferring from their lack of DD behaviour that they cannot consciously process the effects induced by this drug dose, or, even more strongly, that they are not sentient, it is important to note that, just as in the human cases considered in Section 1.3, the absence of performance could be explained by several reasons aside from a lack of conscious awareness:

3.2.2.1. Reasons related to individual/species sensitivity to the specific drug effects

As in verbal humans, studies in non-human animals often find that not all subjects are capable of learning the discrimination, or that perform under the threshold of drug choice selection found in our meta-analysis. However, again, as in verbal humans, this does not necessarily imply that these subjects are unable to consciously process *any* drug effects, or, more strongly, that they are not sentient. It could instead be attributed to the individual capability of each subject to use the effects induced by a specific dose as DSs, or even to their ability to use that specific drug as a DS. In this case, this subject's failure to use the drug as a DS is a *true* negative in the context of the experiment, but a potential *false* negative that requires further examination. This can potentially be investigated by assessing whether these individuals can discriminate higher doses of the active drug from placebo in the acquisition phase, and/or whether they can discriminate between other types of subjective states (e.g. non-drug induced states).

Similar reasons could be argued at the cross-species level: some species may not be sensitive to certain drugs that do induce subjective effects in humans.

3.2.2.2. Reasons related to cognition and the training paradigm

An unsuccessful acquisition of the DD can also be related to cognitive issues. In this case, the subject is consciously aware of the drug, but she is unable to learn to use it as a DS in the task. Inferring that the subject cannot consciously access the effects of this specific drug dose would then be a false negative (as of course would be inferring that it cannot feel anything at all). This problem could not be tackled by increasing the dose or changing the drug. It may be caused by:

1) Using an inefficient training technique (e.g., inconsistent reinforcement);

2) At a within species level, variation in subjects – and/or stages of development – regarding their abilities to learn the discrimination operant task (e.g. to learn that only in the presence of a DS – the effects of a certain drug – does pressing one lever yields a reward) – something that could be avoided by not including the non-learners in the generalisation phase; and

3) At the cross species level, the different average cognitive characteristics of different species: it is possible that some species may be capable of consciously processing drug effects – and other interoceptive stimuli – but lack the cognitive skills for learning a 2AFC paradigm. The general difficulty of training animals to use discriminative stimuli, even exteroceptive ones, is quite well known in researchers studying cognitive bias for example (Carole Fureix, personal communication).

3.3. Conclusions: investigating conscious emotion with the 2AFC paradigm

Altogether, we have seen that the use of emotion (and other interoceptive states) as DS in a 2AFC paradigm is highly likely to require consciously accessing these states, especially at strong levels of DD performance. If the planned studies to fill the limitations of the meta-analysis presented in Chapter 4 confirm this result, we could then confidently use this behavioural test for investigating whether subjects of different phyla and/or stages of development have conscious emotions. Furthermore, by testing whether these subjects generalise between different states of emotion,[38] we could 'ask' these non-verbal subjects to 'self-report' the emotions they find alike or dissimilar, which, crucially, are highly likely to be consciously processed. A relevant point regarding the interpretation of these results has to do with the question of homology and analogy. In non-verbal mammals and other vertebrates with homologous CNSs, physiology, and behaviour, we can make stronger inferences regarding the phenomenal characteristics of the states of emotion they discriminate and generalise, according to whether these results are similar or dissimilar to those from verbal humans. The more different the nervous systems, physiology and behaviour of the subjects undergoing the test, the more caution when concluding 1) whether they are actually consciously processing these DSs; and especially 2) to what extent these states of emotion potentially used as DSs are similar to the ones felt by verbal humans in a similar paradigm (e.g., anxiety). Although these results are thus

[38] When designing and conducting these discrimination experiments – e.g., choosing different states of emotion for the discrimination task –, we must always take into account that previous "observation of behavioral similarities [i.e., display of similar behavioral repertoire to humans] combined with a principle of erring on the side of caution gives us strong reasons to treat animals as moral subjects" (Shriver, 2016b).

more likely to represent false positives, as argued above, in the absence of better explanations for this behaviour we should conclude that these beings are likely to process these stimuli at a conscious level. Furthermore, we could also evaluate what kinds of emotions they find alike, and according to the biology of each species, come up with, at least, reasonable hypotheses regarding the valence of these states (i.e., whether they find them similarly positive or aversive).

As argued above, there are several potential false negatives for this test. Whereas some of them can be easily tackled (such as increasing the drug dose when the effects of the acquisition dose are too weak to be reliably used as a DS), other potential false negatives are more problematic. Given the many reasons for not succeeding in this test, other than the lack of conscious accessing the DSs, any subject not acquiring the discrimination should not be regarded as incapable of conscious emotion, and we should conclude that this test is not a useful one for investigating this question in these subjects.

4. The use of verbal humans as models for understanding consciousness in non-human animals: future directions

In this final section I will argue that several pieces of evidence used for inferring conscious access of emotion in non-human animals are not informative, and that by taking advantage from self-report (SR) in verbal humans we can potentially overcome these limitations. I will underline that this latter approach is not exempt from limitations, but that, crucially, these do not undermine its usefulness for making progress with inferring states of conscious emotion in absence of verbal reports.

4.1. In the absence of self-report validation we cannot make strong inferences on whether emotions are consciously accessed or not

As argued in Chapter 2, verbal reports – complemented by other types of evidence also related to emotion – are a crucial tool for making solid inferences about whether or not states of emotion are consciously processed. When considering the case of non-verbal animals, we obviously lack this self-report tool, and have to rely on other types of evidence, i.e. behavioural, physiological, and neurophysiological data. As

stressed in Section 3 of Chapter 2, however, many of these pieces of evidence are likely to also occur in the absence of conscious processing. Examples include changes in facial expression or different types of simple associative learning related to the detection of noxious stimuli such as avoidance learning and protective motor reactions, and many physiological responses related to states of emotion, such as changes in heart rate.

This means that many indicators currently interpreted as signs potentially painful states in animals – such as the facial expressions in mice (Langford et al. 2010); the grooming and rubbing of the antenna in crustaceans after applying noxious stimuli (Elwood, 2011; avoidance learning of electric shocks, Elwood, 2012); heart rate (see Sneddon et al., 2014 for examples in mammals and birds), and changes in cortisol output (again see Sneddon et al., 2014 for examples in mammals and birds); or the crustaceans' hypoglycemic hormone, which may be analogous (Elwood, 2011) likewise probably do not depend on the conscious processing of emotion, and therefore *per se* are not actually informative of whether the states of emotion related to this behaviour are consciously accessed or not: they could merely be nociceptive responses not involving awareness (see Mason, 2011).

4.2. Taking advantage of self-report in verbal humans

The fundamental idea underlying this thesis is that we need to take advantage of verbal humans' self-reports for validating non-verbal processes related to emotion that are dependent on, or correlate with, consciously access. By doing this, we can potentially come up with far more solid inferences when investigating conscious emotion in non-verbal beings.

In Chapters 3 and 4, and above, we saw that we can use this approach for validating the use of emotion as DSs in the 2AFC paradigm as a marker of conscious emotion. Similarly, in Section 4 of Chapter 2 I proposed that other behavioural tasks related to emotion, such as the cognitive bias paradigm, or patterns of neurophysiological activation, may – once validated as such – come to provide us with valid tools for investigating conscious emotion in the absence of verbal reports. The best indicators are likely to be those related to the hypothetical functions of conscious processing (as reviewed in Chapter 2), which are, namely: 1) generating

representations of the world; 2) predicting and imagining potential scenarios, and eliminating hopeless ones before even trying them; 3) and learning novel contingencies – i.e. creating 'novel bindings'. Thus, there are still many potential indicators of conscious emotion of this type that no one has yet tried to validate using human SR. Crucially, as underlined in Section 4 of Chapter 2, the inferences we can come up with are strongly determined by the degree of phylogenetic proximity between verbal humans and these non-verbal subjects. While in animals equipped with homologous CNSs, physiology, and behaviour we can extrapolate to from what humans self-report with more confidence, when considering the cases of animals distant in phylogeny (e.g., invertebrates) our conclusions are necessarily less informative: these subjects may rely on different processes not inherited from a common ancestor, and eventually not dependent on conscious processing. However, as argued in Section 2.2.1., in absence of evidence suggesting the contrary, any non-verbal subject displaying these SR validated processes should, in principle, be considered as potentially capable of consciously accesing her emotions.

4.3. Limitations of the SR approach

4.3.1. Finding equivalents in verbal humans

In some instances, however, this approach might be challenging or limited. For example, it is hard to apply in cases of behaviour whose human equivalent is unclear, like ultrasonic vocalizations related to distress in rats (Knutson et al., 2002); rubbing of the lip in the gravel of a tank in trout injected with acetic acid in this structure (Sneddon et al 2003 & Braithwaite, 2010); or species-typical trade-offs between stimulus avoidance and other behaviour (e.g., the finding that food-deprived fish, are less likely to respond to an electric shock in a feeding area than those non-food deprived [Millsopp and Laming 2008]; and that hermit crabs challenged with electric shocks tend to evacuate their shells at lower intensities if they are in a less preferred shell species [Appel & Elwood, 2009]). Perhaps such responses *are* related to the hypothetical functions of conscious processing, and thus do reveal conscious emotion. However, in the absence of SR validation – of the behaviour itself, or the neurophysiological machinery underlying it – the inferences we can reach are far less

conclusive. Provided any of this behaviour is consistent with the functions of consciousness, then still, the best we can conclude is that these types of evidence alone just *suggest* that the conscious processing of emotion may take place. Again, as argued in the previous section and in Chapter 4, it is important to underline that these conclusions are always conditioned to the degree of phylogenetic proximity between verbal humans and these non-verbal subjects. For animals close in phylogeny, we can imagine how these hypothetical states of conscious emotion may "feel like" if consciously processed, and thus, despite we cannot make inferences as solid as when having self-report-validated tools, we can still come up with reasonable hypothesis. As we consider the cases of animals more distant in phylogeny our conclusions in this regard are necessarily more tentative.

4.3.2. Using SR for validating conscious emotion

An additional problem intrinsic to the use of human verbal report as a marker of conscious emotion is that it may be prone to false negatives, for the following reasons.

1) Subjects may feel the emotion but be unable to label it (i.e., use symbols to express it verbally): something also touched on in the 'false negatives' section of this chapter's Section 2. As reviewed in Chapter 2, the conscious accessing of information does not seem to depend on the use of symbolic language, which is a process involving higher order consciousness (see Edelman and Seth, 2009; and Dehaene, 2014). Consequently, SR validated types of evidence are not solely related to information integrated through access consciousness, but also to symbolic higher order processing. Subjects may thus have access consciousness – our interest here – but be unable to use the symbolic higher order processing intrinsic to SR.

2) Verbal reports may not always be accurate (e.g., a subject may misunderstand the SR questionnaires in the DD experiments).

These two problems thus mean that SR is a conservative as a validator of the conscious access of emotional information in humans: probably little prone to false positives, but much more prone to false negatives.

4.4. Conclusions

Even though human verbal self-report is not a perfect marker of conscious access of emotion, it is the best validator that we have for now. In the absence of this 'self-report validation', the hypotheses we can come up with on whether a process – e.g., a certain behaviour or pattern of brain activation – depends on consciously accessing information related to emotion are much weaker.

In conclusion, by using human verbal self-report for validating processes related to the functions of conscious emotion (e.g., behaviour that involve 'non-routine bindings', such as discriminating and generalising states of emotion in an operant paradigm, as presented in this thesis), as well as pieces of neurophysiological evidence – which, in turn, may work as non-verbal validators of behaviour –, we can investigate conscious emotion in non-verbal subjects. As in the case of verbal humans, self-report cannot reveal what *it is like* to be another being *from her own perspective*. Furthermore, any reachable conclusion regarding whether a nonverbal being experiences conscious emotion is strongly shaped by her phylogenetic relationship to verbal humans, and therefore by the questions of homology and analogy. Despite these limitations, however, this approach finally offers a potential window into non-verbal beings' worlds of conscious emotion, and therefore the possibility of making stronger inferences on what it *feels like* to be a bat.

Bibliography

Allen, C. (2004). Animal Pain. *Nous*, *38*(4), 617–643. http://doi.org/10.1111/j.0029-4624.2004.00486.x

Allen, C. (2005). Deciphering animal pain. In M. Aydede (Ed.), *Pain: New Essays on Its Nature and the Methodology of Its Study*. Cambridge MA: Bradford Book/MIT Press.

Allen, C. (2013). Fish Cognition and Consciousness. *Journal of Agricultural and Environmental Ethics*, *26*(1), 25–39. http://doi.org/10.1007/s10806-011-9364-9

Allen, C., & Trestman, M. (2015) "Animal Consciousness", The Stanford Encyclopedia of Philosophy (Summer 2015 Edition), Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/sum2015/entries/consciousness-animal/>.

Altman, J. L., Albert, J.-M., Milstein, S. L., & Greenberg, I. (1977). Drugs as Discriminable Events in Humans. In H. Lal (Ed.), *Discriminative Stimulus Properties of Drugs* (Vol. 22, pp. 187–206). Boston, MA: Springer US. Retrieved from http://link.springer.com/10.1007/978-1-4684-3090-5_11

Anderson, D. J., & Adolphs, R. (2014). A Framework for Studying Emotions across Species. *Cell*, *157*(1), 187–200. http://doi.org/10.1016/j.cell.2014.03.003

Andrews, K. (2015). *The animal mind: an introduction to the philosophy of animal cognition*. Abingdon, Oxon: Routledge.

Appel, J. B., West, W. B., Rolandi, W. G., Alici, T., & Pechersky, K. (1999). Increasing the selectivity of drug discrimination procedures. *Pharmacology, Biochemistry, and Behavior*, *64*(2), 353–358.

Appel, M., & Elwood, R. W. (2009). Motivational trade-offs and potential pain experience in hermit crabs. *Applied Animal Behaviour Science*, *119*(1-2), 120–124. http://doi.org/10.1016/j.applanim.2009.03.013

Baars, B. J. (2001). *In the theater of consciousness: the workspace of the mind*. New York: Oxford Univ. Press.

Bárdos, G., & Adám, G. (1978). Visceroceptive control of operant behavior in rats. *Physiology & Behavior*, *20*(4), 369–375.

Basile, B. M., & Hampton, R. R. (2014). Metacognition as discrimination: Commentary on Smith et al. (2014). *Journal of Comparative Psychology*, *128*(2), 135–137. http://doi.org/10.1037/a0034412

Basile, B. M., Schroeder, G. R., Brown, E. K., Templer, V. L., & Hampton, R. R. (2015). Evaluation of seven hypotheses for metamemory performance in rhesus monkeys. *Journal of Experimental Psychology. General*, *144*(1), 85–102. http://doi.org/10.1037/xge0000031

Bateson, M., Desire, S., Gartside, S. E., & Wright, G. A. (2011). Agitated Honeybees Exhibit Pessimistic Cognitive Biases. *Current Biology*, *21*(12), 1070–1073. http://doi.org/10.1016/j.cub.2011.05.017

Bateson, P. (1991). Assessment of pain in animals. *Animal Behaviour*, *42*(5), 827–839. http://doi.org/10.1016/S0003-3472(05)80127-7

Bayne, T., Cleeremans, A., & Wilken, P. (Eds.). (2009). *The Oxford companion to consciousness*. Oxford ; New York: Oxford University Press.

Bekinschtein, T. A., Peeters, M., Shalom, D., & Sigman, M. (2011). Sea slugs, subliminal pictures, and vegetative state patients: boundaries of consciousness in classical conditioning. *Frontiers in Psychology*, *2*, 337. http://doi.org/10.3389/fpsyg.2011.00337

Bennett, D. A., Geyer, H., Dutta, P., Brugger, S., Fielding, S., & Lal, H. (1982). Comparison of the actions of trimethadione and chlordiazepoxide in animal models of

anxiety and benzodiazepine receptor binding. *Neuropharmacology*, *21*(11), 1175–1179.

Bennett, D. A., & Lal, H. (1982). Discriminative stimuli produced by clonidine: an investigation of the possible relationship to adrenoceptor stimulation and hypotension. *The Journal of Pharmacology and Experimental Therapeutics*, *223*(3), 642–648.

Benoit, S. C., & Davidson, T. L. (1996). Interoceptive sensory signals produced by 24-hr food deprivation, pharmacological glucoprivation, and lipoprivation. *Behavioral Neuroscience*, *110*(1), 168–180.

Berridge, K. C. (2003). Comparing the emotional brains of humans and other animals. In R. J. Davidson, K. R. Scherer, & H. H. Goldsmith (Eds.), *Handbook of affective sciences* (pp. 25–51). Oxford ; New York: Oxford University Press.

Berridge, K., & Winkielman, P. (2003). What is an unconscious emotion?(The case for unconscious "liking"). *Cognition & Emotion*, *17*(2), 181–211. http://doi.org/10.1080/02699930302289

Bland, J. M., & Altman, D. G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet (London, England)*, *1*(8476), 307–310.

Block, N. (1995). On a confusion about a function of consciousness. *Behavioral and Brain Sciences*, *18*(02), 227. http://doi.org/10.1017/S0140525X00038188

Block, N. (2005). Two neural correlates of consciousness. *Trends in Cognitive Sciences*, *9*(2), 46–52. http://doi.org/10.1016/j.tics.2004.12.006

Block, N., Carmel, D., Fleming, S. M., Kentridge, R. W., Koch, C., Lamme, V. A. F., … Rosenthal, D. (2014). Consciousness science: real progress and lingering misconceptions. *Trends in Cognitive Sciences*, *18*(11), 556–557. http://doi.org/10.1016/j.tics.2014.09.004

Bogacz, R., Brown, E., Moehlis, J., Holmes, P., & Cohen, J. D. (2006). The physics of optimal decision making: a formal analysis of models of performance in two-alternative forced-choice tasks. *Psychological Review*, *113*(4), 700–765. http://doi.org/10.1037/0033-295X.113.4.700

Braithwaite, V. (2010). *Do fish feel pain?* Oxford ; New York: Oxford University Press.

Brilot, B. O., Asher, L., & Bateson, M. (2010). Stereotyping starlings are more "pessimistic." *Animal Cognition*, *13*(5), 721–731. http://doi.org/10.1007/s10071-010-0323-z

Brydges, N. M., Leach, M., Nicol, K., Wright, R., & Bateson, M. (2011a). Environmental enrichment induces optimistic cognitive bias in rats. *Animal Behaviour*, *81*(1), 169–175. http://doi.org/10.1016/j.anbehav.2010.09.030

Burešová, O., & Bureš, J. (1979). Extinction of a newly acquired conditioned taste aversion: Effect of gustatory CS administered under anaesthesia. *Behavioural Processes*, *4*(4), 329–339. http://doi.org/10.1016/0376-6357(79)90017-2

Butler, A. B., & Cotterill, R. M. J. (2006). Mammalian and avian neuroanatomy and the question of consciousness in birds. *The Biological Bulletin*, *211*(2), 106–127.

Butler, J., Jayson, G. G., & Swallow, A. J. (1975). The reaction between the superoxide anion radical and cytochrome c. *Biochimica Et Biophysica Acta*, *408*(3), 215–222.

Carey, M. P., & Fry, J. P. (1993). A behavioural and pharmacological evaluation of the discriminative stimulus induced by pentylenetetrazole in the pig. *Psychopharmacology*, *111*(2), 244–250.

Carey, M. P., & Fry, J. P. (1995). Evaluation of animal welfare by the self-expression of an anxiety state. *Laboratory Animals*, *29*(4), 370–379.

Carey, M. P., Fry, J. P., & White, D. G. (1992). The detection of changes in psychological state using a novel pharmacological conditioning procedure. *Journal of Neuroscience Methods*, *43*(1), 69–76.

Carruthers, P. (2003). *Phenomenal consciousness: a naturalistic theory* (1. paperback ed). Cambridge: Cambridge Univ. Press.

Carruthers, P., & Ritchie, J. B. (2012). The emergence of metacognition: affect and uncertainty in animals. In M. J. Beran, J. Brandl, J. Perner, & J. Proust (Eds.), *Foundations of Metacognition* (pp. 76–93). Oxford University Press. Retrieved from http://www.oxfordscholarship.com/view/10.1093/acprof:oso/9780199646739.001.0001/acprof-9780199646739-chapter-005

Chait, L. D., Uhlenhuth, E. H., & Johanson, C. E. (1984). An experimental paradigm for studying the discriminative stimulus properties of drugs in humans. *Psychopharmacology*, *82*(3), 272–274.

Chait, L. D., Uhlenhuth, E. H., & Johanson, C. E. (1985). The discriminative stimulus and subjective effects of d-amphetamine in humans. *Psychopharmacology*, *86*(3), 307–312.

Chalmers, D. J. (1995). Facing up to the problem of consciousness. *Journal of Consciousness Studies*, *2*(3), 200–219.

Clark, R. E. (2011). Eyeblink conditioning and systems consolidation: an ironic yet powerful pairing. *Neurobiology of Learning and Memory*, *95*(2), 118–124. http://doi.org/10.1016/j.nlm.2010.12.003

Clark, R. E., Manns, J. R., & Squire, L. R. (2001). Trace and delay eyeblink conditioning: contrasting phenomena of declarative and nondeclarative memory. *Psychological Science*, *12*(4), 304–308.

Clore, G. L. (1994). Why emotions are never unconscious. In P. Ekman & R. J. Davidson (Eds.), *The nature of emotion: fundamental questions* (pp. 285–290). New York: Oxford University Press.

Colpaert, F. C. (1999). Drug discrimination in neurobiology. *Pharmacology, Biochemistry, and Behavior*, *64*(2), 337–345.

Colpaert, F. C., Niemegeers, C. J., & Janssen, P. A. (1978). Discriminative stimulus properties of cocaine and d-amphetamine, and antagonism by haloperidol: a comparative study. *Neuropharmacology*, *17*(11), 937–942.

Comer, S. D. (2002). Self-Administration of Intravenous Buprenorphine and the Buprenorphine/Naloxone Combination by Recently Detoxified Heroin Abusers. *Journal of Pharmacology and Experimental Therapeutics*, *303*(2), 695–703. http://doi.org/10.1124/jpet.102.038141

Cornu, J.-N., Cancel-Tassin, G., Ondet, V., Girardet, C., & Cussenot, O. (2011). Olfactory Detection of Prostate Cancer by Dogs Sniffing Urine: A Step Forward in Early Diagnosis. *European Urology*, *59*(2), 197–201. http://doi.org/10.1016/j.eururo.2010.10.006

Corwin, R. L., Woolverton, W. L., & Schuster, C. R. (1990). Effects of cholecystokinin, d-amphetamine and fenfluramine in rats trained to discriminate 3 from 22 hr of food deprivation. *The Journal of Pharmacology and Experimental Therapeutics*, *253*(2), 720–728.

Covington, H. E., & Miczek, K. A. (2003). Vocalizations during withdrawal from opiates and cocaine: possible expressions of affective distress. *European Journal of Pharmacology*, *467*(1-3), 1–13.

Crick, F. C., & Koch, C. (2005). What is the function of the claustrum? *Philosophical Transactions of the Royal Society B: Biological Sciences*, *360*(1458), 1271–1279. http://doi.org/10.1098/rstb.2005.1661

Crick, F., & Koch, C. (2003). A framework for consciousness. *Nature Neuroscience*, *6*(2), 119–126. http://doi.org/10.1038/nn0203-119

Custers, R., & Aarts, H. (2010). The Unconscious Will: How the Pursuit of Goals Operates Outside of Conscious Awareness. *Science*, *329*(5987), 47–50. http://doi.org/10.1126/science.1188595

Darwin, Charles. (1872). *The Expression of the Emotions in Man and Animals*. London: Murray.

Dawkins, M. (2015). Animal Welfare and the Paradox of Animal Consciousness. In *Advances in the Study of Behavior* (Vol. 47, pp. 5–38). Elsevier. Retrieved from http://linkinghub.elsevier.com/retrieve/pii/S0065345414000023

Dawkins, R. (1976). *The selfish gene*. Oxford: Oxford University Press.

de Gelder, B., Vroomen, J., Pourtois, G., & Weiskrantz, L. (1999). Non-conscious recognition of affect in the absence of striate cortex: *NeuroReport*, *10*(18), 3759–3763. http://doi.org/10.1097/00001756-199912160-00007

de Gelder, B., & Tamietto, M. (2007) Scholarpedia, 2(10): 3555

DeGrazia, D. (1996). *Taking animals seriously: mental life and moral status*. Cambridge ; New York: Cambridge University Press.

DeGrazia, D. (2014). What Is Suffering and What Sorts of Beings Can Suffer? In R. M. Green & N. J. Palpant (Eds.), *Suffering and Bioethics* (pp. 134–154). Oxford University Press. Retrieved from http://www.oxfordscholarship.com/view/10.1093/acprof:oso/9780199926176.001.0001/acprof-9780199926176-chapter-7

Dehaene, S. (2014). *Consciousness and the brain: deciphering how the brain codes our thoughts*. New York, New York: Viking Adult.

Desborough, J. P. (2000). The stress response to trauma and surgery. *British Journal of Anaesthesia*, *85*(1), 109–117.

De Vos, M., Gandras, K., & Debener, S. (2014). Towards a truly mobile auditory brain–computer interface: Exploring the P300 to take away. *International Journal of Psychophysiology*, *91*(1), 46–53. http://doi.org/10.1016/j.ijpsycho.2013.08.010

Dimberg, U., & Petterson, M. (2000). Facial reactions to happy and angry facial expressions: evidence for right hemisphere dominance. *Psychophysiology*, *37*(5), 693–696.

Duka, T., Stephens, D. N., Russell, C., & Tasker, R. (1998). Discriminative stimulus properties of low doses of ethanol in humans. *Psychopharmacology*, *136*(4), 379–389.

Duka, T., Tasker, R., Russell, K., & Stephens, D. N. (1998). Discriminative stimulus properties of nicotine at low doses: the effects of caffeine preload. *Behavioural Pharmacology*, *9*(3), 219–229.

Ebner, N. C., Riediger, M., & Lindenberger, U. (2010). FACES—A database of facial expressions in young, middle-aged, and older women and men: Development and validation. *Behavior Research Methods*, *42*(1), 351–362. http://doi.org/10.3758/BRM.42.1.351

Edelman, D. B., & Seth, A. K. (2009). Animal consciousness: a synthetic approach. *Trends in Neurosciences*, *32*(9), 476–484. http://doi.org/10.1016/j.tins.2009.05.008

Edelman, G. M. (1989). *The remembered present: a biological theory of consciousness*. New York: Basic Books.

Edelman, G. M. (2003). Naturalizing consciousness: A theoretical framework. *Proceedings of the National Academy of Sciences*, *100*(9), 5520–5524. http://doi.org/10.1073/pnas.0931349100

Ehmann, R., Boedeker, E., Friedrich, U., Sagert, J., Dippon, J., Friedel, G., & Walles, T. (2012). Canine scent detection in the diagnosis of lung cancer: revisiting a puzzling phenomenon. *European Respiratory Journal*, *39*(3), 669–676. http://doi.org/10.1183/09031936.00051711

Ellsworth, P.C. (1994). Levels of thought and levels of emotion. In P. Ekman & R. J. Davidson (Eds.), *The nature of emotion: fundamental questions* (pp. 192–196). New York: Oxford University Press.

Elwood, R. (2012). Evidence for pain in decapod crustaceans. *Animal Welfare*, *21*(1), 23–27. http://doi.org/10.7120/096272812X13353700593365

Elwood, R. W. (2011). Pain and suffering in invertebrates? *ILAR Journal / National Research Council, Institute of Laboratory Animal Resources*, *52*(2), 175–184.

Emmett-Oglesby, M., Spencer, D., Lewis, M., Elmesallamy, F., & Lal, H. (1983). Anxiogenic aspects of diazepam withdrawal can be detected in animals. *European Journal of Pharmacology*, *92*(1-2), 127–130.

Emmett-Oglesby, M. W., Spencer, D. G., Elmesallamy, F., & Lal, H. (1983). The pentylenetetrazol model of anxiety detects withdrawal from diazepam in rats. *Life Sciences*, *33*(2), 161–168.

Epstein, D. H.; Schmittner, J. P.; Walsh, S. L., P. K. L. (2007). Abuse of Marketed Medications. In S. B. Karch (Ed.), *Drug abuse handbook* (2nd ed, pp. 281–311). Boca Raton: CRC Press/Taylor & Francis.

Fantegrossi, W. E., Murai, N., Mathúna, B. O., Pizarro, N., & de la Torre, R. (2009). Discriminative stimulus effects of 3,4-methylenedioxymethamphetamine and its

enantiomers in mice: pharmacokinetic considerations. *The Journal of Pharmacology and Experimental Therapeutics*, *329*(3), 1006–1015. http://doi.org/10.1124/jpet.109.150573

Farah, M. J. (2008). Neuroethics and the Problem of Other Minds: Implications of Neuroscience for the Moral Status of Brain-Damaged Patients and Nonhuman Animals. *Neuroethics*, *1*(1), 9–18. http://doi.org/10.1007/s12152-008-9006-8

Flanagan, O. (1995). *Consciousness reconsidered* (3. print). Cambridge, Mass.: MIT Press.

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *American Psychologist*, *34*(10), 906–911. http://doi.org/10.1037/0003-066X.34.10.906

Frijda, N. H. (2003). Emotions and hedonic experience. In D. Kahneman, E. Diener, & N. Schwarz (Eds.), *Well-being: the foundations of hedonic psychology* (1. papercover ed, pp. 190–210). New York, NY: Russell Sage Foundation.

Gallagher, S., & Zahavi, D. (2015) "Phenomenological Approaches to Self-Consciousness", The Stanford Encyclopedia of Philosophy (Spring 2015 Edition), Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/spr2015/entries/self-consciousness-phenomenological/>.

Gardner, R. A., Gardner, B. T., & Van Cantfort, T. E. (Eds.). (1989). *Teaching sign language to chimpanzees*. Albany: State University of New York Press.

Gatch, M. B., & Lal, H. (2001). Animal models of the anxiogenic effects of ethanol withdrawal. *Drug Development Research*, *54*(2), 95–115. http://doi.org/10.1002/ddr.1209

Gauvin, D. V., Briscoe, R. J., Baird, T. J., Vallett, M., Carl, K. L., & Holloway, F. A. (1997). Cross-generalization of an EtOH “hangover” cue to endogenously and

exogenously induced stimuli. *Pharmacology, Biochemistry, and Behavior*, *57*(1-2), 199–206.

Gauvin, D. V., Goulden, K. L., & Holloway, F. A. (1993). State-dependent stimulus control: cueing attributes of ethanol "hangover" in rats. *Alcoholism, Clinical and Experimental Research*, *17*(6), 1210–1214.

Gauvin, D. V., Harland, R. D., Criado, J. R., Michaelis, R. C., & Holloway, F. A. (1989). The discriminative stimulus properties of ethanol and acute ethanol withdrawal states in rats. *Drug and Alcohol Dependence*, *24*(2), 103–113.

Gauvin, D. V., Harland, R. D., & Holloway, F. A. (1989). Drug discrimination procedures: A method to analyze adaptation level of affective states. *Drug Development Research*, *16*(2-4), 183–194. http://doi.org/10.1002/ddr.430160212

Gauvin, D. V., & Holloway, F. A. (1991). Cross-generalization between an ethologically [correction of ecologically] relevant stimulus and a pentylenetetrazole-discriminative cue. *Pharmacology, Biochemistry, and Behavior*, *39*(2), 521–523.

Gauvin, D. V., Youngblood, B. D., & Holloway, F. A. (1992). The discriminative stimulus properties of acute ethanol withdrawal (hangover) in rats. *Alcoholism, Clinical and Experimental Research*, *16*(2), 336–341.

Gazit, I., Goldblatt, A., & Terkel, J. (2005). The role of context specificity in learning: the effects of training context on explosives detection in dogs. *Animal Cognition*, *8*(3), 143–150. http://doi.org/10.1007/s10071-004-0236-9

Giurfa, M. (2013). Cognition with few neurons: higher-order learning in insects. *Trends in Neurosciences*, *36*(5), 285–294. http://doi.org/10.1016/j.tins.2012.12.011

Glennon, R. A., & Young, R. (Eds.). (2011). *Drug discrimination: applications to medicinal chemistry and drug studies*. Hoboken, New Jersey: Wiley.

Goeleven, E., De Raedt, R., Leyman, L., & Verschuere, B. (2008). The Karolinska Directed Emotional Faces: A validation study. *Cognition & Emotion*, *22*(6), 1094–1118. http://doi.org/10.1080/02699930701626582

Grau, J. W., Barstow, D. G., & Joynes, R. L. (1998). Instrumental learning within the spinal cord: I. Behavioral properties. *Behavioral Neuroscience*, *112*(6), 1366–1386.

Griffiths, R. R., Evans, S. M., Heishman, S. J., Preston, K. L., Sannerud, C. A., Wolf, B., & Woodson, P. P. (1990). Low-dose caffeine discrimination in humans. *The Journal of Pharmacology and Experimental Therapeutics*, *252*(3), 970–978.

Hagbarth, K. E. (1960). Spinal withdrawal reflexes in the human lower limbs. *Journal of Neurology, Neurosurgery, and Psychiatry*, *23*, 222–227.

Harding, E. J., Paul, E. S., & Mendl, M. (2004). Animal behaviour: Cognitive bias and affective state. *Nature*, *427*(6972), 312–312. http://doi.org/10.1038/427312a

Harris, C. M., Emmett-Oglesby, M. W., & Lal, H. (1989). Sensitivity of pentylenetetrazol discrimination increased by a stimulus fading technique. *Psychopharmacology*, *98*(4), 460–464.

Harris, C. M., Emmett-Oglesby, M. W., Robinson, N. G., & Lal, H. (1986). Withdrawal from chronic nicotine substitutes partially for the interoceptive stimulus produced by pentylenetetrazol (PTZ). *Psychopharmacology*, *90*(1). http://doi.org/10.1007/BF00172876

Herling, S., & Winger, G. (1981). Selective blockade of the discriminative stimulus effects of pentobarbital in pigeons. *Psychopharmacology*, *75*(3), 321–323.

Hernstein, R. J., Loveland, D. H., & Cable, C. (1976). Natural concepts in pigeons. *Journal of Experimental Psychology. Animal Behavior Processes*, *2*(4), 285–302.

Herrnstein, R. J., & Loveland, D. H. (1964). Complex Visual Concept in the Pigeon. *Science*, *146*(3643), 549–551. http://doi.org/10.1126/science.146.3643.549

Holtzman, S. G. (1985). Discriminative stimulus effects of morphine withdrawal in the dependent rat: suppression by opiate and nonopiate drugs. *The Journal of Pharmacology and Experimental Therapeutics*, *233*(1), 80–86.

Horta, O. (2010). What is Speciesism? *Journal of Agricultural and Environmental Ethics*, *23*(3), 243–266. http://doi.org/10.1007/s10806-009-9205-2

Hyslop, A. "Other Minds", The Stanford Encyclopedia of Philosophy (Fall 2015 Edition), Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/fall2015/entries/other-minds/>.

Huxley, T. (1874). On the hypothesis that animals are automata, and its history. *Fortnightly Review*, *95*, 555–80.

Institute for Laboratory Animal Research (U.S.) (Ed.). (2009). *Recognition and alleviation of pain in laboratory animals*. Washington, D.C: National Academies Press.

Jack, A. I., & Shallice, T. (2001). Introspective physicalism as an approach to the science of consciousness. *Cognition*, *79*(1-2), 161–196.

James, W. (1884). II.—WHAT IS AN EMOTION ? *Mind*, *os-IX*(34), 188–205. http://doi.org/10.1093/mind/os-IX.34.188

Järbe, T. U., & Swedberg, M. D. (1998). Discriminative stimulus functions of CNS sedative drugs assessed by drug versus drug discrimination procedures in gerbils. *Psychopharmacology*, *135*(2), 201–212.

Johanson, C. E. (1991). Discriminative stimulus effects of diazepam in humans. *The Journal of Pharmacology and Experimental Therapeutics*, *257*(2), 634–643.

Jolij, J., & Lamme, V. A. F. (2005). Repression of unconscious information by conscious processing: Evidence from affective blindsight induced by transcranial magnetic stimulation. *Proceedings of the National Academy of Sciences*, *102*(30), 10747–10751. http://doi.org/10.1073/pnas.0500834102

Jung, M. E., Lal, H., & Gatch, M. B. (2002). The discriminative stimulus effects of pentylenetetrazol as a model of anxiety: recent developments. *Neuroscience and Biobehavioral Reviews*, *26*(4), 429–439.

Kamien, J. B., Bickel, W. K., Hughes, J. R., Higgins, S. T., & Smith, B. J. (1993). Drug discrimination by humans compared to nonhumans: current status and future directions. *Psychopharmacology*, *111*(3), 259–270.

Kaminski, J. (2004). Word Learning in a Domestic Dog: Evidence for "Fast Mapping." *Science*, *304*(5677), 1682–1683. http://doi.org/10.1126/science.1097859

Kattner, F., Ellermeier, W., & Tavakoli, P. (2012). Both trace and delay conditioning of evaluative responses depend on contingency awareness. *Learning and Motivation*, *43*(1-2), 35–47. http://doi.org/10.1016/j.lmot.2012.01.004

Kelly, T. H., Emurian, C. S., Baseheart, B. J., & Martin, C. A. (1997). Discriminative stimulus effects of alcohol in humans. *Drug and Alcohol Dependence*, *48*(3), 199–207.

Kendrick, K. M., da Costa, A. P., Leigh, A. E., Hinton, M. R., & Peirce, J. W. (2001). Sheep don't forget a face. *Nature*, *414*(6860), 165–166. http://doi.org/10.1038/35102669

Kiesel, A., Kunde, W., Pohl, C., Berner, M. P., & Hoffmann, J. (2009). Playing chess unconsciously. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(1), 292–298. http://doi.org/10.1037/a0014499

Kihlstrom, J. F. (1999). The psychological unconscious. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality: theory and research* (2nd ed, pp. 424–442). New York: Guilford Press.

Kihlstrom, J. F., Mulvaney, S., Tobias, B. A., & Tobis, I. P. (2000). The emotional unconscious. In E. Eich, J. F. Kihlstrom, G. H. Bower, J. P. Forgas, & P. M. Niedenthal (Eds.), *Cognition and Emotion* (pp. 30–86). Oxford University Press.

Kirk, R. (2015) "Zombies", The Stanford Encyclopedia of Philosophy (Summer 2015 Edition), Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/sum2015/entries/zombies/>.

Kirton, A., Winter, A., Wirrell, E., & Snead, O. C. (2008). Seizure response dogs: Evaluation of a formal training program. *Epilepsy & Behavior*, *13*(3), 499–504. http://doi.org/10.1016/j.yebeh.2008.05.011

Knutson, B., Burgdorf, J., & Panksepp, J. (2002). Ultrasonic vocalizations as indices of affective states in rats. *Psychological Bulletin*, *128*(6), 961–977.

Koch, C. (2014). In the Playing Ground of Consciousness. *Science*, *343*(6170), 487–487. http://doi.org/10.1126/science.1248710

Koros, E., Kostowski, W., Danysz, W., & Bienkowski, P. (1999). Ethanol discrimination in the rat: lack of modulation by restraint stress and memantine. *Naunyn-Schmiedeberg's Archives of Pharmacology*, *359*(2), 117–122.

Lader, M., Petursson, H. (1981). Withdrawal from long-term benzodiazepine treatment. *British Medical Journal (Clinical Research Ed.)*, *283*, 643–645.

Lal, H. (1979). Interocepttve Stimuli as Tools of Drug Development. *Drug Development and Industrial Pharmacy*, *5*(2), 133–149. http://doi.org/10.3109/03639047909055668

Lal, H., & Emmett-Oglesby, M. W. (1983). Behavioral analogues of anxiety. Animal models. *Neuropharmacology*, *22*(12B), 1423–1441.

Lal, H., Harris, C. M., Benjamin, D., Springfield, A. C., Bhadra, S., & Emmett-Oglesby, M. W. (1988). Characterization of a pentylenetetrazol-like interoceptive stimulus

produced by ethanol withdrawal. *The Journal of Pharmacology and Experimental Therapeutics*, *247*(2), 508–518.

Lal, H.; Shearman, G. T. (1978). Discriminative stimulus properties of pentylenetetrazol in the rat. In F. C. Colpaert, J. A. Rosecrans, & Janssen Research Foundation (Eds.), *Stimulus properties of drugs: ten years of progress* (pp. 181–188). Amsterdam ; New York : New York: Elsevier/North-Holland Biomedical Press ; sole distributors for the USA and Canada, Elsevier/North-Holland.

Lal, H., & Shearman, G. T. (1982). Attenuation of chemically induced anxiogenic stimuli as a novel method for evaluating anxiolytic drugs: A comparison of clobazam with other benzodiazepines. *Drug Development Research*, *2*(S1), 127–134. http://doi.org/10.1002/ddr.430010717

Lal, H., Shearman, G. T., Fielding, S., Dunn, R., Kruse, H., & Theurer, K. (1980). Effect of valproic acid on anxiety-related behaviors in the rat. *Brain Research Bulletin*, *5*, 575–577. http://doi.org/10.1016/0361-9230(80)90093-3

Lal, H., Shearman, G. T., Fielding, S., Dunn, R., Kruse, H., & Theurer, K. (1980). Evidence that GABA mechanisms mediate the anxiolytic action of benzodiazepines: a study with valproic acid. *Neuropharmacology*, *19*(8), 785–789.

Lal, H., & Yaden, S. (1985). Discriminative stimuli produced by clonidine in spontaneously hypertensive rats: generalization to antihypertensive drugs with different mechanisms of action. *The Journal of Pharmacology and Experimental Therapeutics*, *232*(1), 33–39.

Langford, D. J., Bailey, A. L., Chanda, M. L., Clarke, S. E., Drummond, T. E., Echols, S., … Mogil, J. S. (2010). Coding of facial expressions of pain in the laboratory mouse. *Nature Methods*, *7*(6), 447–449. http://doi.org/10.1038/nmeth.1455

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1990). Emotion, attention, and the startle reflex. *Psychological Review*, *97*(3), 377–395. http://doi.org/10.1037/0033-295X.97.3.377

Leberer, M. R., & Fowler, S. C. (1977). Drug discrimination and generalization in pigeons. *Pharmacology, Biochemistry, and Behavior*, *7*(5), 483–486.

Lile, J. A., Kelly, T. H., & Hays, L. R. (2010). Substitution Profile of the Cannabinoid Agonist Nabilone in Human Subjects Discriminating Δ9-Tetrahydrocannabinol: *Clinical Neuropharmacology*, *33*(5), 235–242. http://doi.org/10.1097/WNF.0b013e3181e77428

Lile, J. A., Kelly, T. H., & Hays, L. R. (2011). Separate and combined effects of the cannabinoid agonists nabilone and Δ9-THC in humans discriminating Δ9-THC. *Drug and Alcohol Dependence*, *116*(1-3), 86–92. http://doi.org/10.1016/j.drugalcdep.2010.11.019

Lile, J. A., Kelly, T. H., & Hays, L. R. (2012a). Separate and combined effects of the GABAB agonist baclofen and Δ9-THC in humans discriminating Δ9-THC. *Drug and Alcohol Dependence*, *126*(1-2), 216–223. http://doi.org/10.1016/j.drugalcdep.2012.05.023

Lile, J. A., Kelly, T. H., & Hays, L. R. (2012b). Separate and combined effects of the GABA reuptake inhibitor tiagabine and Δ9-THC in humans discriminating Δ9-THC. *Drug and Alcohol Dependence*, *122*(1-2), 61–69. http://doi.org/10.1016/j.drugalcdep.2011.09.010

Lile, J. A., Kelly, T. H., & Hays, L. R. (2014). Separate and combined effects of the GABAA positive allosteric modulator diazepam and Δ9-THC in humans discriminating Δ9-THC. *Drug and Alcohol Dependence*, *143*, 141–148. http://doi.org/10.1016/j.drugalcdep.2014.07.016

Lile, J. A., Kelly, T. H., Pinsky, D. J., & Hays, L. R. (2009). Substitution profile of Δ9-tetrahydrocannabinol, triazolam, hydromorphone, and methylphenidate in humans discriminating Δ9-tetrahydrocannabinol. *Psychopharmacology*, *203*(2), 241–250. http://doi.org/10.1007/s00213-008-1393-3

Lile, J. A., Stoops, W. W., Durell, T. M., Glaser, P. E. A., & Rush, C. R. (2006). Discriminative-stimulus, self-reported, performance, and cardiovascular effects of atomoxetine in methylphenidate-trained humans. *Experimental and Clinical Psychopharmacology*, *14*(2), 136–147. http://doi.org/10.1037/1064-1297.14.2.136

Lile, J. A., Stoops, W. W., Glaser, P. E. A., Hays, L. R., & Rush, C. R. (2004). Acute administration of the GABA reuptake inhibitor tiagabine does not alter the effects of oral cocaine in humans. *Drug and Alcohol Dependence*, *76*(1), 81–91. http://doi.org/10.1016/j.drugalcdep.2004.04.010

Lile, J. A., Stoops, W. W., Glaser, P. E., Hays, L. R., & Rush, C. R. (2011). Discriminative stimulus, subject-rated and cardiovascular effects of cocaine alone and in combination with aripiprazole in humans. *Journal of Psychopharmacology*, *25*(11), 1469–1479. http://doi.org/10.1177/0269881110385597

Lile, J. A., Stoops, W. W., Vansickel, A. R., Glaser, P. E. A., Hays, L. R., & Rush, C. R. (2005). Aripiprazole Attenuates the Discriminative-Stimulus and Subject-Rated Effects of D-Amphetamine in Humans. *Neuropsychopharmacology*, *30*(11), 2103–2114. http://doi.org/10.1038/sj.npp.1300803

Lile, J. A., Stoops, W. W., Wagner, F. P., Glaser, P. E. A., & Rush, C. R. (2005). Oxazepam does not modulate the behavioral effects of d-amphetamine in humans. *Pharmacology Biochemistry and Behavior*, *82*(2), 270–279. http://doi.org/10.1016/j.pbb.2005.08.012

Li, M., & McMillan, D. E. (2001). Four-choice drug discrimination in pigeons. *Behavioural Pharmacology*, *12*(8), 621–628.

Mantsch, J. R., & Goeders, N. E. (1998). Generalization of a restraint-induced discriminative stimulus to cocaine in rats. *Psychopharmacology*, *135*(4), 423–426.

Mason, G. J. (2011). Invertebrate welfare: where is the real evidence for conscious affective states? *Trends in Ecology & Evolution*, *26*(5), 212–213. http://doi.org/10.1016/j.tree.2011.02.009

Matheson, S. M.; Bateson, M. (2007). Performance on a categorisation task suggests that removal of environmental enrichment induces "pessimism" in captive European starlings (Sturnus vulgaris). *Animal Welfare*, *16*, 33–36.

McCulloch, M. (2006). Diagnostic Accuracy of Canine Scent Detection in Early- and Late-Stage Lung and Breast Cancers. *Integrative Cancer Therapies*, *5*(1), 30–39. http://doi.org/10.1177/1534735405285096

McGreevy, P., & Boakes, R. A. (2007). *Carrots and sticks: principles of animal training*. Cambridge ; New York: Cambridge University Press.

Mendl, M., Brooks, J., Basse, C., Burman, O., Paul, E., Blackwell, E., & Casey, R. (2010). Dogs showing separation-related behaviour exhibit a "pessimistic" cognitive bias. *Current Biology*, *20*(19), R839–R840. http://doi.org/10.1016/j.cub.2010.08.030

Miczek, K. A., Mutschler, N. H., van Erp, A. M., Blank, A. D., & McInerney, S. C. (1999). d-amphetamine "cue" generalizes to social defeat stress: behavioral sensitization and attenuated accumbens dopamine. *Psychopharmacology*, *147*(2), 190–199.

Miksic, S.; McCarten, M. A., Lal, H. (1978). A comparison of discriminative stimuli produced by naloxone, cyclazocine and morphine in the rat. In F. C. Colpaert, J. A. Rosecrans, & Janssen Research Foundation (Eds.), *Stimulus properties of drugs: ten years of progress* (pp. 177–180). Amsterdam ; New York : New York:

Elsevier/North-Holland Biomedical Press ; sole distributors for the USA and Canada, Elsevier/North-Holland.

Millsopp, S., & Laming, P. (2008). Trade-offs between feeding and shock avoidance in goldfish (Carassius auratus). *Applied Animal Behaviour Science*, *113*(1-3), 247–254. http://doi.org/10.1016/j.applanim.2007.11.004

Moody, M. F., & Parriss, J. R. (1961). The discrimination of polarized light by Octopus: a behavioural and morphological study. *Zeitschrift Für Vergleichende Physiologie*, *44*(3), 268–291. http://doi.org/10.1007/BF00298356

Mumford, G. K., Evans, S. M., Kaminski, B. J., Preston, K. L., Sannerud, C. A., Silverman, K., & Griffiths, R. R. (1994). Discriminative stimulus and subjective effects of theobromine and caffeine in humans. *Psychopharmacology*, *115*(1-2), 1–8.

Murphy, S. T., & Zajonc, R. B. (1993). Affect, cognition, and awareness: affective priming with optimal and suboptimal stimulus exposures. *Journal of Personality and Social Psychology*, *64*(5), 723–739.

Nagel, T. (1974). What is it like to be a bat? *Philosophical Review*, *83*(October), 435–50.

Nelson, T. O. (1996). Consciousness and metacognition. *American Psychologist*, *51*(2), 102–116. http://doi.org/10.1037/0003-066X.51.2.102

Norton, J. J. S., Lee, D. S., Lee, J. W., Lee, W., Kwon, O., Won, P., … Rogers, J. A. (2015). Soft, curved electrode systems capable of integration on the auricle as a persistent brain–computer interface. *Proceedings of the National Academy of Sciences*, *112*(13), 3920–3925. http://doi.org/10.1073/pnas.1424875112

Ohman, A., & Soares, J. J. (1994). “Unconscious anxiety”: phobic responses to masked stimuli. *Journal of Abnormal Psychology*, *103*(2), 231–240.

Oliveto, A. H., Bickel, W. K., Hughes, J. R., Higgins, S. T., & Badger, G. (1997). Functional antagonism of the caffeine-discriminative stimulus by triazolam in humans. *Behavioural Pharmacology*, *8*(2-3), 124–138.

Oliveto, A. H., Bickel, W. K., Hughes, J. R., Higgins, S. T., & Fenwick, J. W. (1992). Triazolam as a discriminative stimulus in humans. *Drug and Alcohol Dependence*, *30*(2), 133–142. http://doi.org/10.1016/0376-8716(92)90018-8

Oliveto, A. H., Bickel, W. K., Hughes, J. R., Terry, S. Y., Higgins, S. T., & Badger, G. J. (1993). Pharmacological specificity of the caffeine discriminative stimulus in humans: effects of theophylline, methylphenidate and buspirone. *Behavioural Pharmacology*, *4*(3), 237–246.

Oliveto, A. H., McCance-Katz, E., Singha, A., Hameedi, F., & Kosten, T. R. (1998). Effects of d-amphetamine and caffeine in humans under a cocaine discrimination procedure. *Behavioural Pharmacology*, *9*(3), 207–217.

Oliveto, A. H., Rosen, M. I., Woods, S. W., & Kosten, T. R. (1995). Discriminative stimulus, self-reported and cardiovascular effects of orally administered cocaine in humans. *The Journal of Pharmacology and Experimental Therapeutics*, *272*(1), 231–241.

Oliveto, A., Sevarino, K., McCance-Katz, E., & Feingold, A. (2002). Butorphanol and nalbuphine in opioid-dependent humans under a naloxone discrimination procedure. *Pharmacology, Biochemistry, and Behavior*, *71*(1-2), 85–96.

Pain terms: a list with definitions and notes on usage. Recommended by the IASP Subcommittee on Taxonomy. (1979). *Pain*, *6*(3), 249.

Pang, R., Turndorf, H., & Quartermain, D. (1996). Pavlovian fear conditioning in mice anesthetized with halothane. *Physiology & Behavior*, *59*(4-5), 873–875.

Panksepp, J. (2005). *Affective neuroscience: the foundations of human and animal emotions*. Oxford: Oxford Univ. Press.

Pappens, M., Vandenbossche, E., Van den Bergh, O., & Van Diest, I. (2015). Interoceptive fear learning to mild breathlessness as a laboratory model for unexpected panic attacks. *Frontiers in Psychology*, *6*. http://doi.org/10.3389/fpsyg.2015.01150

Perkins, K. A., D'Amico, D., Sanders, M., Grobe, J. E., Wilson, A., & Stiller, R. L. (1996). Influence of training dose on nicotine discrimination in humans. *Psychopharmacology*, *126*(2), 132–139.

Perkins, K. A., Fonte, C., Blakesley-Ball, R., & Wilson, A. S. (2005). The discriminative stimulus, subjective, cardiovascular, and reinforcing effects of nicotine as a function of light physical activity. *Nicotine & Tobacco Research: Official Journal of the Society for Research on Nicotine and Tobacco*, *7*(5), 791–800. http://doi.org/10.1080/14622200500262931

Perkins, K. A., Fonte, C., Sanders, M., White, W., & Wilson, A. (1999). Effects of training dose and two- versus three-choice testing procedure on nicotine discrimination responding in humans. *Psychopharmacology*, *145*(4), 418–425.

Perkins, K. A., Sanders, M., D'Amico, D., & Wilson, A. (1997). Nicotine discrimination and self-administration in humans as a function of smoking status. *Psychopharmacology*, *131*(4), 361–370.

Perkins, K. A., Sanders, M., Fonte, C., Wilson, A. S., White, W., Stiller, R., & McNamara, D. (1999). Effects of central and peripheral nicotinic blockade on human nicotine discrimination. *Psychopharmacology*, *142*(2), 158–164.

Persson, I. (1993). A Basis (Interspecies) Equality. In P. Singer & P. Cavalieri (Eds.), *The Great Ape Project* (pp. 183–193). St. Martin's Griffin.

Pickel, D., Manucy, G. P., Walker, D. B., Hall, S. B., & Walker, J. C. (2004). Evidence for canine olfactory detection of melanoma. *Applied Animal Behaviour Science*, *89*(1-2), 107–116. http://doi.org/10.1016/j.applanim.2004.04.008

Pluhar, E. B. (1995). *Beyond prejudice: the moral significance of human and nonhuman animals*. Durham: Duke University Press.

Popper, K. (1978). Natural Selection and the Emergence of Mind. *Dialectica*, *32*(3-4), 339–355. http://doi.org/10.1111/j.1746-8361.1978.tb01321.x

Porter, J. H., Prus, A. J., Vann, R. E., & Varvel, S. A. (2005). Discriminative stimulus properties of the atypical antipsychotic clozapine and the typical antipsychotic chlorpromazine in a three-choice drug discrimination procedure in rats. *Psychopharmacology*, *178*(1), 67–77. http://doi.org/10.1007/s00213-004-1985-5

Preston, K. L., & Bigelow, G. E. (1998). Opioid discrimination in humans: discriminative and subjective effects of progressively lower training dose. *Behavioural Pharmacology*, *9*(7), 533–543.

Preston, K. L., & Bigelow, G. E. (2000). Effects of agonist-antagonist opioids in humans trained in a hydromorphone/not hydromorphone discrimination. *The Journal of Pharmacology and Experimental Therapeutics*, *295*(1), 114–124.

Preston, K. L., Liebson, I. A., & Bigelow, G. E. (1992). Discrimination of agonist-antagonist opioids in humans trained on a two-choice saline-hydromorphone discrimination. *The Journal of Pharmacology and Experimental Therapeutics*, *261*(1), 62–71.

Provenza, F. D., Lynch, J. J., & Nolan, J. V. (1994). Food aversion conditioned in anesthetized sheep. *Physiology & Behavior*, *55*(3), 429–432. http://doi.org/10.1016/0031-9384(94)90096-5

Prus, A. J., Philibin, S. D., Pehrson, A. L., & Porter, J. H. (2005). Generalization to atypical antipsychotic drugs depends on training dose in rats trained to discriminate 1.25 mg/kg clozapine versus 5.0 mg/kg clozapine versus vehicle in a three-choice drug discrimination task. *Behavioural Pharmacology*, *16*(7), 511–520.

Ramachandran, V. (1998). The perception of phantom limbs. The D. O. Hebb lecture. *Brain*, *121*(9), 1603–1630. http://doi.org/10.1093/brain/121.9.1603

Raybuck, J. D., & Lattal, K. M. (2014). Bridging the interval: theory and neurobiology of trace conditioning. *Behavioural Processes*, *101*, 103–111. http://doi.org/10.1016/j.beproc.2013.08.016

Reagh, Z. M., & Knight, D. C. (2013). Negative, but not positive emotional images modulate the startle response independent of conscious awareness. *Emotion*, *13*(4), 782–791. http://doi.org/10.1037/a0032286

Regan, T. (2004). *The case for animal rights* (Updated with a new preface, [2004 ed.]). Berkeley: University of California Press.

Reynolds, A. R., Bolin, B. L., Stoops, W. W., & Rush, C. R. (2013). Relationship between drug discrimination and ratings of subjective effects: implications for assessing and understanding the abuse potential of D-amphetamine in humans. *Behavioural Pharmacology*, *24*(5 and 6), 523–532. http://doi.org/10.1097/FBP.0b013e328364505f

Rivas, E. (2005). Recent use of signs by chimpanzees (Pan troglodytes) in interactions with humans. *Journal of Comparative Psychology*, *119*(4), 404–417. http://doi.org/10.1037/0735-7036.119.4.404

Rolls, E. T. (2014). *Emotion and decision-making explained* (First edition). Oxford ; New York, NY: Oxford University Press.

Rowlands, M. (1997). Contractarianism and Animal Rights. *Journal of Applied Philosophy*, *14*(3), 235–247. http://doi.org/10.1111/1468-5930.00060

Ruiz-Padial, E., & Vila, J. (2007). Fearful and Sexual Pictures Not Consciously Seen Modulate the Startle Reflex in Human Beings. *Biological Psychiatry*, *61*(8), 996–1001. http://doi.org/10.1016/j.biopsych.2006.08.046

Ruiz-Padial, E., Vila, J., & Thayer, J. F. (2011). The effect of conscious and non-conscious presentation of biologically relevant emotion pictures on emotion modulated startle and phasic heart rate. *International Journal of Psychophysiology*, *79*(3), 341–346. http://doi.org/10.1016/j.ijpsycho.2010.12.001

Rush, C. R. (2003). Risperidone Attenuates the Discriminative-Stimulus Effects of d-Amphetamine in Humans. *Journal of Pharmacology and Experimental Therapeutics*, *306*(1), 195–204. http://doi.org/10.1124/jpet.102.048439

Rush, C. R., & Baker, R. W. (2001). Behavioral pharmacological similarities between methylphenidate and cocaine in cocaine abusers. *Experimental and Clinical Psychopharmacology*, *9*(1), 59–73.

Rush, C. R., Baker, R. W., & Rowlett, J. K. (2000). Discriminative-stimulus effects of zolpidem, triazolam, pentobarbital, and caffeine in zolpidem-trained humans. *Experimental and Clinical Psychopharmacology*, *8*(1), 22–36.

Rush, C. R., Kelly, T. H., Fillmore, M. T., & Hays, L. R. (2003). Discriminative-stimulus effects of triazolam in light and moderate drinkers. *Alcoholism, Clinical and Experimental Research*, *27*(4), 638–646. http://doi.org/10.1097/01.ALC.0000062742.29158.6E

Rush, C. R., Kelly, T. H., Hays, L. R., & Wooten, A. F. (2002). Discriminative-stimulus effects of modafinil in cocaine-trained humans. *Drug and Alcohol Dependence*, *67*(3), 311–322.

Rush, C. R., Kollins, S. H., & Pazzaglia, P. J. (1998). Discriminative-stimulus and participant-rated effects of methylphenidate, bupropion, and triazolam in d-

amphetamine-trained humans. *Experimental and Clinical Psychopharmacology*, *6*(1), 32–44.

Rush, C. R., Madakasira, S., Goldman, N. H., Woolverton, W. L., & Rowlett, J. K. (1997). Discriminative stimulus effects of zolpidem in pentobarbital-trained subjects: II. Comparison with triazolam and caffeine in humans. *The Journal of Pharmacology and Experimental Therapeutics*, *280*(1), 174–188.

Rush, C. R., Stoops, W. W., Wagner, F. P., Hays, L. R., & Glaser, P. E. A. (2004). Alprazolam attenuates the behavioral effects of d-amphetamine in humans. *Journal of Clinical Psychopharmacology*, *24*(4), 410–420.

Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological Review*, *110*(1), 145–172. http://doi.org/10.1037/0033-295X.110.1.145

Sankey, C., Richard-Yris, M.-A., Leroy, H., Henry, S., & Hausberger, M. (2010). Positive interactions lead to lasting positive memories in horses, Equus caballus. *Animal Behaviour*, *79*(4), 869–875. http://doi.org/10.1016/j.anbehav.2009.12.037

Savage-Rumbaugh, E. S., & Lewin, R. (1994). *Kanzi: the ape at the brink of the human mind.* New York: Wiley.

Savage-Rumbaugh, S., & Brakke, K. E. (1996). Animal language: Methodological and interpretative issues. In C. Allen & D. Jamison (Eds.), *Readings in Animal Cognition* (pp. 269–288). MIT Press.

Schechter, M. D. (1983). Discriminative stimulus control with imipramine: transfer to other anti-depressants. *Pharmacology, Biochemistry, and Behavior*, *19*(5), 751–754.

Schwitzgebel, E. (2014) "Introspection", The Stanford Encyclopedia of Philosophy (Summer 2014 Edition), Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/sum2014/entries/introspection/>.

Seth, A. K., Baars, B. J., & Edelman, D. B. (2005). Criteria for consciousness in humans and other mammals. *Consciousness and Cognition*, *14*(1), 119–139. http://doi.org/10.1016/j.concog.2004.08.006

Seth, A. K., Dienes, Z., Cleeremans, A., Overgaard, M., & Pessoa, L. (2008). Measuring consciousness: relating behavioural and neurophysiological approaches. *Trends in Cognitive Sciences*, *12*(8), 314–321. http://doi.org/10.1016/j.tics.2008.04.008

Sevak, R. J., Stoops, W. W., Hays, L. R., & Rush, C. R. (2009). Discriminative Stimulus and Subject-Rated Effects of Methamphetamine, d-Amphetamine, Methylphenidate, and Triazolam in Methamphetamine-Trained Humans. *Journal of Pharmacology and Experimental Therapeutics*, *328*(3), 1007–1018. http://doi.org/10.1124/jpet.108.147124

Sevak, R. J., Vansickel, A. R., Stoops, W. W., Glaser, P. E. A., Hays, L. R., & Rush, C. R. (2011). Discriminative-Stimulus, Subject-Rated, and Physiological Effects of Methamphetamine in Humans Pretreated With Aripiprazole: *Journal of Clinical Psychopharmacology*, *31*(4), 470–480. http://doi.org/10.1097/JCP.0b013e318221b2db

Shearman, G., & Lal, H. (1979). Discriminative stimulus properties of pentylenetetrazol and bemegride: some generalization and antagonism tests. *Psychopharmacology*, *64*(3), 315–319.

Shearman, G. T., & Lal, H. (1981). Discriminative stimulus properties of cocaine related to an anxiogenic action. *Progress in Neuro-Psychopharmacology*, *5*(1), 57–63.

Sherrington, S. (1906). *The Integrative Action of the Nervous System*. New York: Charles Scribner's Sons.

Shettleworth, S. J. (1998). *Cognition, evolution, and behavior*. New York: Oxford University Press.

Shriver, A. (2006). Minding Mammals∗. *Philosophical Psychology*, *19*(4), 433–442. http://doi.org/10.1080/09515080600726385

Shriver, A. (2016a). Commentary on Key on Fish Pain. *Animal Sentience*.

Shriver, A. (2016b). The Unpleasantness of Pain in Humans and Other Animals. In Bain D, Brady M, & Corns J (Ed.), *The Pain Project (official title TBA)*.

Sinelnikova, E. M., Dvoretskova, T. V., & Kagan, Z. S. (1975). [Intermediate plateaux in kinetics of the reaction catalyzed by biodegradative L-threonine dehydratase from Escherichia coli]. *Biokhimiia (Moscow, Russia)*, *40*(3), 645–651.

Singer, P. (2011). *Practical ethics* (3rd ed). New York: Cambridge University Press.

Slucki, H., Adam, G., & Porter, R. W. (1965). Operant discrimination of an interoceptive stimulus in rhesus monkeys1. *Journal of the Experimental Analysis of Behavior*, *8*(6), 405–414. http://doi.org/10.1901/jeab.1965.8-405

Smith, B. J., & Bickel, W. K. (1999). Flumazenil discrimination by humans under a two-response and a novel-response procedure. *The Journal of Pharmacology and Experimental Therapeutics*, *291*(3), 1257–1268.

Smith, B. J., & Bickel, W. K. (2001). Effects of alprazolam, caffeine, and zolpidem in humans trained to discriminate triazolam from placebo. *Drug and Alcohol Dependence*, *61*(3), 249–260.

Sneddon, L. U. (2003). The evidence for pain in fish: the use of morphine as an analgesic. *Applied Animal Behaviour Science*, *83*(2), 153–162. http://doi.org/10.1016/S0168-1591(03)00113-8

Sneddon, L. U., Elwood, R. W., Adamo, S. A., & Leach, M. C. (2014). Defining and assessing animal pain. *Animal Behaviour*, *97*, 201–212. http://doi.org/10.1016/j.anbehav.2014.09.007

Sobel, N., Prabhakaran, V., Hartley, C. A., Desmond, J. E., Glover, G. H., Sullivan, E. V., & Gabrieli, J. D. (1999). Blind smell: brain activation induced by an undetected air-borne chemical. *Brain: A Journal of Neurology*, *122 (Pt 2)*, 209–217.

Soldoff, S., & Slucki, H. (1974). Operant discrimination of interoceptive urinary bladder stimulation in the monkey☆☆☆. *Physiology & Behavior*, *12*(4), 583–587. http://doi.org/10.1016/0031-9384(74)90206-6

Steiner, J. E. (1979). Human facial expressions in response to taste and smell stimulation. *Advances in Child Development and Behavior*, *13*, 257–295.

Steiner, J. E., Glaser, D., Hawilo, M. E., & Berridge, K. C. (2001). Comparative expression of hedonic impact: affective reactions to taste by human infants and other primates. *Neuroscience and Biobehavioral Reviews*, *25*(1), 53–74.

Stolerman, I. P., Childs, E., Ford, M. M., & Grant, K. A. (2011). Role of training dose in drug discrimination: a review. *Behavioural Pharmacology*, *22*(5 and 6), 415–429. http://doi.org/10.1097/FBP.0b013e328349ab37

Stoops, W. W., Lile, J. A., Glaser, P. E. A., & Rush, C. R. (2005). Discriminative Stimulus and Self-Reported Effects of Methylphenidate, d-Amphetamine, and Triazolam in Methylphenidate-Trained Humans. *Experimental and Clinical Psychopharmacology*, *13*(1), 56–64. http://doi.org/10.1037/1064-1297.13.1.56

Stoops, W. W., Lile, J. A., Glaser, P. E. A., & Rush, C. R. (2006). A low dose of aripiprazole attenuates the subject-rated effects of d-amphetamine. *Drug and Alcohol Dependence*, *84*(2), 206–209. http://doi.org/10.1016/j.drugalcdep.2006.02.004

Strong, V., Brown, S. W., & Walker, R. (1999). Seizure-alert dogs — fact or fiction? *Seizure*, *8*(1), 62–65. http://doi.org/10.1053/seiz.1998.0250

Tamietto, M., & de Gelder, B. (2010). Neural bases of the non-conscious perception of emotional signals. *Nature Reviews Neuroscience*, *11*(10), 697–709. http://doi.org/10.1038/nrn2889

Tye, M. (2015) "Qualia", The Stanford Encyclopedia of Philosophy (Fall 2015 Edition), Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/fall2015/entries/qualia/>.

Valentino, R. J., Herling, S., & Woods, J. H. (1983). Discriminative stimulus effects of naltrexone in narcotic-naive and morphine-treated pigeons. *The Journal of Pharmacology and Experimental Therapeutics*, *224*(2), 307–313.

Van Gulick, R. (2014) "Consciousness", The Stanford Encyclopedia of Philosophy (Spring 2014 Edition), Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/spr2014/entries/consciousness/>.

Vansickel, A. R., Hays, L. R., & Rush, C. R. (2006). Discriminative-Stimulus Effects of Triazolam in Women and Men. *The American Journal of Drug and Alcohol Abuse*, *32*(3), 329–349. http://doi.org/10.1080/00952990500479266

Varner, G. E. (2012). *Personhood, ethics, and animal cognition: situating animals in Hare's two level utilitarianism*. New York: Oxford University Press.

Velázquez, D. N.;Díaz, J. L. (2000). La discriminación del efecto de las drogas y la conciencia animal. *Salud Mental*, *23*, 1–7.

Vellucci, S. V., Martin, P. J., & Everitt, B. J. (1988). The discriminative stimulus produced by pentylenetetrazol: effects of systemic anxiolytics and anxiogenics, aggressive defeat and midazolam or muscimol infused into the amygdala. *Journal of Psychopharmacology*, *2*(2), 80–93. http://doi.org/10.1177/026988118800200203

Vivian, J. A., Weerts, E. M., & Miczek, K. A. (1994). Defeat engenders pentylenetetrazole-appropriate responding in rats: antagonism by midazolam. *Psychopharmacology*, *116*(4), 491–498.

Vrana, S. R., Spence, E. L., & Lang, P. J. (1988). The startle probe response: a new measure of emotion? *Journal of Abnormal Psychology*, *97*(4), 487–491.

Weinberger, N. M., Gold, P. E., & Sternberg, D. B. (1984). Epinephrine enables Pavlovian fear conditioning under anesthesia. *Science (New York, N.Y.)*, *223*(4636), 605–607.

Weiskrantz, L. (2002). *Blindsight: a case study and implications* (Reprint). Oxford: Clarendon.

Weissman, A. (1976). The discriminability of apsirin in arthritic and nonarthritic rats. *Pharmacology, Biochemistry, and Behavior*, *5*(5), 583–586.

Werner, P. (1975). [Use of the Cessna 206 for medical evacuations]. *Annales De L'anesthésiologie Française*, *16*(1), 41–44.

Wiens, S., & Ohman, A. (2002). Unawareness is more than a chance event: comment on Lovibond and Shanks (2002). *Journal of Experimental Psychology. Animal Behavior Processes*, *28*(1), 27–31.

Wiens, S. & Öhman, A. (2007). Probing unconscious emotional processes: On becoming a successful masketeer. In J. A. Coan & J. J. B. Allen (Eds.), *Handbook of emotion elicitation and assessment* (pp. 65–90). Oxford ; New York: Oxford University Press.

Williams, L. E., Bargh, J. A., Nocera, C. C., & Gray, J. R. (2009). The unconscious regulation of emotion: Nonconscious reappraisal goals modulate emotional reactivity. *Emotion*, *9*(6), 847–854. http://doi.org/10.1037/a0017745

Winkielman, P.; Berridge, P.; Wilbarger, J. (2000). Unconscious affect for doing without feeling: Subliminal facial expressions alter human consumption. *Unpublished Manuscript.*

Winkielman, P., & C. Berridge, K. (2004). Unconscious Emotion. *Current Directions in Psychological Science*, *13*(3), 120–123. http://doi.org/10.1111/j.0963-7214.2004.00288.x

Winkielman, P., & Schwarz, R. B. Z. & N. (1997). Subliminal affective priming resists attributional interventions. *Cognition and Emotion*, *11*(4), 433–465.

Wood, D. M., & Lal, H. (1987). Anxiogenic properties of cocaine withdrawal. *Life Sciences*, *41*(11), 1431–1436.

Wood, D. M., Laraby, P. R., & Lal, H. (1989). A pentylenetetrazol-like stimulus during cocaine withdrawal: Blockade by diazepam but not haloperidol. *Drug Development Research*, *16*(2-4), 269–276. http://doi.org/10.1002/ddr.430160220

Woody, C., Kriplani, A., O'Connor, P., Pratte, J.-F., Radeka, V., Rescia, S., … Yu, B. (2004). RatCAP: a small, head-mounted PET tomograph for imaging the brain of an awake RAT. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, *527*(1-2), 166–170. http://doi.org/10.1016/j.nima.2004.03.114

Wu, W., Moreno, A. M., Tangen, J. M., & Reinhard, J. (2013). Honeybees can discriminate between Monet and Picasso paintings. *Journal of Comparative Physiology. A, Neuroethology, Sensory, Neural, and Behavioral Physiology*, *199*(1), 45–55. http://doi.org/10.1007/s00359-012-0767-5

Yoshida, M., & Isa, T. (2015). Signal detection analysis of blindsight in monkeys. *Scientific Reports*, *5*, 10755. http://doi.org/10.1038/srep10755

Young, R. (2009). Drug Discrimination. In J. J. Buccafusco & J. J. Buccafusco (Eds.), *Methods of behavioral analysis in neuroscience* (2nd ed, pp. 39–58). Boca Raton: CRC Press.

Zajonc, Robert B. (1998). Emotions. In Susan T. Fiske, Daniel T. Gilbert, Gardner Lindzey (Ed.), *The handbook of social psychology, Vols. 1 and 2* (pp. 591–632). New York: McGraw-Hill.

Zemack-Rugar, Y., Bettman, J. R., & Fitzsimons, G. J. (2007). The effects of nonconsciously priming emotion concepts on behavior. *Journal of Personality and Social Psychology*, *93*(6), 927–939. http://doi.org/10.1037/0022-3514.93.6.927

Appendix Table 1: Versions of the self-report questionnaires most often used in combination with drug discrimination

The Addiction Research Center Inventory (ARCI)	The short form of the ARCI is often used, and consists of 49 true/false questions and contained five major subscales: MBG [an index of euphoria]; PCAG [an index of sedation]; LSD [an index of dysphoric changes]; and BG and Amphetamine scales [empirically derived amphetamine-sensitive scales].
The Profile of Mood States (POMS) **related methods**	This 72-item adjective rating scale yields scores on eight mood clusters (e.g., Fatigue). Subjects rate each item by selecting one of five response options: "Not at all," "A little bit", "Moderately", "Quite a bit" and "Extremely."

The Adjective-Rating Scale	Questionnaire that consists of different items (e.g., active, dizzy, euphoric, fatigued…) that subjects rate by selecting among one of five response options: "not at all", "a little bit", "moderately", "quite a bit", and "extremely" (scored numerically from 0 to 4, respectively). According to their pharmacological nature, these items are also contained in subscales such the *sedative*, and the *stimulant* ones. Often presented as a continuous measure in the form of visual analogue scales (VAS). Here subjects typically rate the different items by placing an arrow along a 100-point line anchored at either end by the terms "none" and "extremely".
The Drug–Effect Questionnaire	Questionnaire that consists of different items (e.g., any effect; bad effects; high; like drug…) that participants have to rate using a 5-point scale similar to the one described above.
The Stimulant-Sensitive Adjective-Rating Scale	Questionnaire that consists of several items (e.g., nervous; drug effect; sleepy; sweaty…), also rated on a 5-point scale, but with the characteristic that responses to individual items are summed to create a composite score, with a maximum total score of 84.

The pharmacological class questionnaire	Subjects must categorize the drug effect as being most similar to one of the different classes of psychoactive drugs (e.g., opiates, barbiturates, antidepressants, hallucinogens…). The questionnaire usually provides descriptive labels and examples for each of these different classes.