This is a preprint version of the paper: Jurjako, M. and Malatesti, L. 2016. Instrumental rationality in psychopathy: implications from learning tasks. *Philosophical Psychology*, 29(5): 717-731.

Instrumental rationality in psychopathy: implications from learning tasks

Abstract

The issue whether psychopathic offenders are practically rational has attracted philosophical attention. The problem is relevant in theoretical discussions on moral psychology and in those concerning the appropriate social response to the crimes of these individuals. We argue that classical and current experiments concerning the instrumental learning in psychopaths cannot directly support the conclusion that they have impaired instrumental rationality, construed as the ability for transferring the motivation by means-ends reasoning. In fact, we defend the different claim that these experiments appear to show that psychopaths in certain circumstances are not aware of the relevant means for their ends. Moreover, we suggest how further empirical research could help to settle the issue.

Keywords: psychopathy, instrumental learning, instrumental rationality, epistemic accessibility

1. Introduction

The issue whether people who are classified as having psychopathy are rational is central in three recent interrelated philosophical debates. First, sentimentalists maintain that psychopaths exemplify the case of the immoral rational agent (Aaltola, 2014; Nichols, 2004; Prinz, 2006). Such a case would undermine the rationalist account of moral judgment and motivation. Rationalists, instead, question the rationality of psychopaths (Maibom, 2005, 2010; Kennett, 2010). Second, some externalists about moral judgment claim that, against internalism, psychopaths are rational, possess moral understanding and are not

motivated to act morally. Internalists, besides questioning the assumption that psychopaths make moral judgments, might question their rationality (see e.g. Sinnott-Armstrong, 2014). Third, the outcomes of these two debates might also be relevant for the discussion of the moral and legal responsibility of psychopathic offenders (Malatesti & McMillan (eds.), 2010; Kiehl & Sinnott-Armstrong, (eds.) 2013).

Investigating the rationality of psychopathic individuals requires focusing on acceptable requirements of rationality and offering relevant evidence to assess whether they satisfy these requirements. Heidi Maibom, in her seminal paper "Moral Unreason: The Case of Psychopathy" (Maibom, 2005, and see also her 2010), considers in her discussion of the rationality of psychopathy a plausible requirement of *instrumental rationality*. According to this requirement, if a person is instrumentally rational, then she intends to use the means at her disposal that are necessary to achieve her ends. Consequently, a person is *instrumentally irrational* if she fails to intend to use such means.

Moreover, Maibom argues that certain empirical studies that show that psychopaths perform worse than controls in certain forms of instrumental learning are relevant in assessing their instrumental rationality (Maibom, 2005, 2010). This type of learning involves establishing associations between a set of stimuli, the subject's response and rewarding or punishing stimulus in a certain context. Specifically, Maibom takes the poor performance of psychopaths on instrumental learning tasks to show that they are instrumentally irrational. According to her line of reasoning, due to their learning impairment, they fail to intend to use the means that are at their disposal to achieve the end of solving the task.

We argue that the empirical evidence adduced by Maibom (e.g. Lykken, 1957; Newman, Patterson, & Kosson, 1987) and newer results on the instrumental learning of

psychopaths (Brazil et al., 2013a) leave the issue whether they are instrumentally irrational undecided. Most significantly, we claim that the bearing of these empirical studies on the problem of the rationality of psychopaths, can only be assessed once several conceptual issues concerning the formulation of the requirements of practical rationality are settled. In particular, we argue that the main source of difficulty in using these results is that they cannot tell us whether certain relevant instrumental reasons are available to psychopathic participants. We will also offer some suggestions on how this issue could be addressed experimentally.

We will proceed as follows. First, we describe the diagnosis of psychopathy and briefly survey the classical experiments that show peculiarities in instrumental learning in psychopaths. Second, we present and criticize Maibom's argument that these peculiarities imply that psychopaths are instrumentally irrational. Third, we argue that even newer experiments on the instrumental learning of psychopaths cannot settle the issue. Finally, we offer some recommendations for framing the experiments on instrumental learning in psychopaths that might help in addressing the issue whether they have impairments in instrumental rationality.

2. Psychopathy and instrumental learning: the classic experiments

The Psychopathy Checklist-Revised (PCL-R), elaborated by Robert Hare and collaborators, is currently widely used for the diagnosis of psychopathy (Hare, 2003; Forth, Kosson, & Hare, 2003; Hare & Neumann, 2010; for a survey of alternative diagnostic tools to the PCL-R, see Fowler & Lilienfeld 2013). The PCL-R is used by trained clinicians to measure psychopathy by means of semi-structured interviews and intensive study of the history of the subject, which should be supported by available file records, on 20 items:

1. Glib/superficial charm. 2. Grandiose sense of self-worth. 3. Need for

> stimulation/proneness to boredom. 4. Pathological lying. 5. Conning/manipulativeness. 6. Lack of remorse or guilt. 7. Shallow affect. 8. Callous/lack of empathy. 9. Parasitic lifestyle. 10. Poor behavioral controls. 11. Promiscuous sexual behavior. 12. Early behavioral problems. 13. Lack of realistic long-terms goals. 14. Impulsivity. 15. Irresponsibility. 16. Failure to accept responsibility for one's own actions. 17. Many short-term marital relationships. 18. Juvenile delinquency 19. Revocation of conditional release, 20. Criminal versatility.

For each element in the list, there is a score ranging from 0 to 2 points; the maximum total score is thus 40 points. When a subject scores 30 or more points he/she is considered psychopathic. This cut-off value is usually adopted in North America; in Europe, a value of 25 is often used (Cooke & Michie, 1999).

The PCL-R has proved to be a unifying diagnostic tool for scientific research. Many functional, neural and even genetic peculiarities related to psychopathy have been investigated by using it (see Blair, Mitchell, & Blair, 2005; Patrick (ed.), 2006; Glenn & Raine, 2014). Maibom and other philosophers interested in the rational capacities of psychopaths have relied on numerous classic empirical studies that focus on psychopaths' poor behavioral inhibition in *instrumental learning* tasks (see Maibom, 2005, 2010; Kennett, 2006, 2010). These tasks involve learning in a certain situation the relation between the stimuli and the appropriate response through rewards or punishments. For example, in a certain situation – a classroom – a child learns to raise her hand – a response – in order to receive the attention of a teacher – a rewarding stimulus.

David Lykken (1957) showed that psychopaths have difficulties in *passive avoidance learning*. This type of learning involves responding to stimuli that are associated to rewards

and not responding to those that involve punishments. Lykken used a "mental maze" in his experiment. This maze consists of 20 decision points with four possible choices that are indicated by four response levers. By pressing the only correct lever, the participants advance to the next decision point. But this does not happen if they press any of the three remaining incorrect levers. In addition, just one of these incorrect levers is associated to a punishment consisting in an electric shock. Passive avoidance learning is measured by the increased avoidance of the punished responses in repeated trials. Lykken observed that psychopaths were more prone than non-psychopathic controls to commit passive avoidance errors.

Other studies show that psychopaths have difficulties in *response extinction* tasks. In these tasks participants learn to respond to a stimulus that elicits a reward and then learn to avoid responding to the same type of stimulus when it elicits punishment. Newman, Patterson, & Kosson (1987) developed the one-pack card-playing task to investigate response inhibition learning. At the beginning of the task, participants, who face a computer screen, are given 10 chips, each having a certain monetary value. At the beginning of every trial, the sentence "do you want to play?" appears on the screen. The participants have two choices: play the card by pressing one button or quit the game by pressing a different one. In addition, they are instructed to play as many cards as they like. If they choose to play, they are shown one card taken from a nonstandard 100 card pack. If the participants press the button to play and a face card (Jack, Queen, King or Ace) follows, they win and receive one chip. If their decision to play is followed by a number card, they lose and one chip is taken from them. The probability of getting number cards, and thus losing, increases by 10% with each 10 played cards. So, if 100 cards are played it is impossible to win. The measured parameter in this task is the number of cards played before quitting. Newman and colleagues and other researchers have established that

compared to non-psychopathic controls, adult psychopaths and children with psychopathic tendencies have difficulties in inhibiting the response that was initially rewarded (Fisher & Blair, 1998; Newman, Patterson, & Kosson, 1987; Newman, Patterson, & Howland, 1990; O'Brien & Frick, 1996). They continue to play the cards even when they are punished for doing so.

Psychopaths appear also to be impaired in *response reversal* learning experimental tasks where participants learn to change strategy when initially rewarded stimuli, after a set number of trials, elicit punishment (Lapierre, Braun, & Hodgins, 1995; Mitchell et al., 2002). Mitchell and others used the intradimensional/extradimensional (ID/ED) multi-component instrumental learning task (Mitchell et al., 2002) where participants have to learn to choose between two stimuli presented to them simultaneously on a computer screen by touching them using a mouse on the basis of feedback provided in the same screen with the words 'correct' or 'incorrect'. The correct stimuli are specified by one of their dimensions, for instance their shape. The ID/ED task involves nine learning stages. The second, the seventh and the ninth stage concern response reversal relative to what the participants have learned in the immediately precedent stage (the acquisition phase). In the second stage, for instance, after the participants have learned in the first stage to respond to stimuli of a certain shape, the reward contingencies are reversed. Thus, the participant must inhibit responding to stimuli with one shape and instead respond to stimuli with a different shape.

Finally, in comparison to non-psychopathic controls, psychopaths manifest shortcomings in the Iowa gambling task (Bechara et al., 1994; Blair, Colledge, & Mitchell, 2001; Mitchell, et al., 2002). This experimental paradigm consists of a card game in which participants make 100 consecutive card selections from four decks (A, B, C, and D) that are presented on a computer screen. Different monetary rewards and punishments are

associated to each of these decks. Choosing decks A and B is disadvantageous and results in a sizeable net loss. For instance, over ten selections from deck A, the participant gains US$ 1000 (facsimile), but there are also five unpredictable losses ranging from US$ 150 to 350 thereby bringing the total loss to US$ 1250. Over 10 selections from deck B, the participant gains US$ 1000, but there is one loss of US$ 1250. On the other hand, Decks C and D have a higher frequency of punishment, but these punishments are of a lower magnitude. So, choosing decks C and D is advantageous, rendering a net gain. Decks C and D enable a gain of US$ 500, but even smaller losses (ranging from US$ 25 to 75 in deck C and one US$ 250 loss in deck D) thereby rendering a net gain of US$ 250. Psychopathic participants, as compared to controls, show non risk-aversive behavior in selecting the decks throughout the task. Accordingly, they sustain major losses. Having reviewed the principal results concerning the peculiarities of psychopaths in instrumental learning, let us move to the philosophical interpretation of them in relation to the issue of practical, and more in particular, instrumental rationality.

3. The facets of practical rationality

Reliance on empirical data is of outmost importance in addressing the issue of the practical rationality of psychopaths. However, inferring conclusions about psychopaths' rationality from empirical data on instrumental learning tasks requires preliminary conceptual or theoretical clarifications. The concept of practical rationality is multifaceted and different requirements might be central in ascribing it. Therefore, we have to establish which facets of practical rationality and which principles of ascription are more relevant to the interpretation of the data that we have described so far.

This is a preprint version of the paper: Jurjako, M. and Malatesti, L. 2016. Instrumental rationality in psychopathy: implications from learning tasks. *Philosophical Psychology*, 29(5): 717-731.

Heidi Maibom (2005, p. 241) has advanced a plausible principle that is of central importance in framing the issue of the relation of the instrumental learning and practical rationality in psychopaths:

> (i) who wills the end, also wills the means that are indispensably necessary to his actions and that lie in his power.

This principle, that can be called a principle of *instrumental rationality*, is implicit in the standard picture, which is usually adopted in the psychological literature on practical rationality. According to accounts of this type, rationality involves reasoning in accordance with the axioms of the classical decision theory (Samuels & Stich, 2004). However, here we follow a more philosophical tradition, where this requirement of instrumental rationality is made explicit (see Kant, 1785/1993, GMM, 417) and is often viewed as a more general functional requirement on the ascription of mental states (Davidson, 2001; Maibom, 2005; Millar, 2004).

Practical rationality does not amount to nor is exhausted by principle (i). Similarly, other principles of rationality might be invoked to test the practical rationality of psychopaths. For example, following O'Neill (2001), Maibom advances principle (i) in a list of other requirements to be tested with psychopaths:

> Practical rationality requires that who wills the end, also wills:
>
> (a) the means that are indispensably necessary to his actions and that lie in his power,
>
> (b) some sufficient means to the end,
>
> (c) to make available necessary and/or sufficient means to the end if such means aren't already available,

> (d) that the various specific intentions that are involved in adopting a maxim are mutually consistent, and
>
> (e) that the foreseeable consequences of acting on the specific intentions are consistent with the underlying intention. (Maibom, 2005, p. 241)

It is obvious that the earlier reviewed instrumental learning tasks are not designed to test the ability of a person to be creative in solving a task or to adopt consistent maxims. Rather they are meant to test the abilities to learn certain responses by learning stimuli/reinforcement contingencies, often by classical and operant conditioning (Blair, 2006). Since they have the latter function they bear more directly on the issue of the instrumental rationality of psychopaths as evaluated in the requirement (a) (and to a certain extent (b)).

According to Maibom, passive avoidance errors and failures in the gambling task indicate that psychopaths do not will the necessary means for their ends because, on average, they are worse than the controls at learning to avoid punished stimuli in the first task, while in the second task they are worse at learning to chose from rewarding decks of cards. However, according to Maibom, learning in these cases presents necessary means for accomplishing their ends (Maibom, 2005, pp. 242-243).

Evaluating Maibom's argument requires clarifying the proviso in (i) which states that means must *lie in the power* of the agent. Clearly, her reasoning is sound only if it can be proved that psychopaths, in the relevant experimental tasks, do not will means that are in their power. However, neither Maibom nor O'Neill, from whom Maibom takes condition (i) and who talks about *available* means (O'Neill, 2001, pp. 311-312), spell out the notion further. So, let us suggest some possible readings of it.

Maibom's reasoning could be based on the assumption that the fulfillment of principle (i), including the proviso about available means, can be spelled out in terms of the successfulness of the action of the agent without any further evidence. In fact, rationality of intentions naturally translates into rationality of actions, and intentions are naturally construed as states that ordinarily produce and legitimize ascription of intentionality to behavior (see Davidson, 2001). So, the specific differences in the actions of psychopaths and non-psychopaths in the experiments at issue would be direct evidence of their not willing the *available* means to their ends.

We respond to this reading of (i) by noting that ascriptions of intentions refer to unobservable and inward aspects of action and their rationality does not necessarily reflect the successfulness of ensuing actions (O'Neill, 2001, p. 306). For example, I might intend to save my friend by shooting his attacker, but I shoot my friend accidentally. Also, I might intend to pursue necessary and/or sufficient means for my ends but due to some physical obstacle I might be disabled from acting on my intentions. Thus, requirements of rationality, such as (a), seem to refer to internal states of the subjects (for the thesis that rationality supervenes on the mind see [Broome, 2013, pp. 34, 151-152]).

Maibom's argument could be based on the assumption that *available* means in requirement (i) are those that the control group or the average normal (healthy) person who is represented by the control group, actually wills in order to successfully complete the tasks at issue. In this case, the internal states and the abilities of the control group would offer an account of availability that extends to psychopaths too.

Although this notion of availability might be plausible in some contexts and have considerable heuristic value, it cannot figure in a satisfactory general formulation of principle (i). The normal practice of establishing the availability of information that is relevant for ascribing rationality often relies, implicitly or explicitly, on a reference class

that includes a normal distribution of the relevant capacities. However, this practice breaks down in cases where, intuitively, lack of certain normal cognitive or affective abilities does not affect our judgments of rationality. For example, it is rational for a blind person to choose apples by their taste and not by their color. This seems to be the case even though the blind person would fail in a task in which the requirement is to choose good tasty apples based on perceiving their color. *Generally,* when rational requirements are at issue, the performance of a specific group on some task cannot prescribe what means or other considerations are available to a subject involved in the same task. Instead, we have to consider carefully the capacities (or incapacities) of that agent in relation to the considerations relevant to the specific task. Therefore, the supporter of the suggested notion of availability should show that performances of psychopaths in the learning tasks, that we have considered so far, are not affected by impairments that undermine the ascription of availability grounded on what control groups are capable of doing. However, we think that this cannot be shown. To see this let us first clarify further the notion of availability that is most relevant for the interpretation of the instrumental learning shortcomings in psychopaths.

We maintain that it is plausible that the psychopaths' performances in certain learning tasks are due to a specific form of *epistemic inaccessibility* of the relevant information that does not undermine ascriptions of instrumental rationality. In the context of ascriptions of rationality, epistemic inaccessibility generally is taken to designate the standard idea that instrumental rationality is relativized to what the agent believes or knows about his/her circumstances (see e.g. Broome, 2013; Kolodny & Brunero, 2013). Bernard Williams famously illustrates this idea with the case of a person who orders gin and tonic but unknowingly receives a glass full of petroleum. The person is rational even if she drinks petroleum because relative to her belief she does not have access to the relevant information

about the petroleum (Williams, 1981). However, we would like to suggest a more specific form of epistemic inaccessibility by spelling out its sources and form.

We maintain that there are impairments in the peripheral parts of the decision-making system that affect the accessibility to some relevant information but do not warrant the withdrawal of ascriptions of instrumental rationality as specified by principle (i). Consider, for example, a color-blind person who is faced with solving a response reversal task. She needs to learn to respond to yellow circled stimuli and avoid responding to red squared stimuli. Since the background color of the screen is green she manages to learn to respond to yellow stimuli and to avoid responding to the red square since she does not perceive it. Now suppose that after this pattern of responses is established, the goal of the task is reversed, what was previously rewarded is now punished and *vice versa*. In the reversed situation it is safe to say that the person will make more mistakes in learning to change the response than an average person with normal color vision. Intuitively however, this person is not irrational, since her disability produced insensitivity to the relevant information in this particular situation.

Maibom's argument for the instrumental irrationality of psychopaths fails because their performance in the instrumental tasks is explainable in terms of epistemic inaccessibility. In fact, two principal explanatory hypotheses advanced by experts in the field support the claim that psychopaths are not sensitive to the connection between the stimulus and punishment and do not detect changes in the reward/punishment contingency. Joseph Newman and colleagues' *response modulation* hypothesis (RMH) and James Blair and colleagues' *integrated emotions system* (IES) hypothesis appear to be the principal competing accounts of the constellations of functional impairments in psychopaths, that relate to those exposed on instrumental learning tasks (Blair & Mitchell 2009; Moul, Killcross, & Dadds, 2012). While the former purports to account for their performance in

the relevant tasks in terms of attentional deficits, the latter explains them in terms of affective/emotional dysfunctions. In this paper we do not try to adjudicate the merits of these two accounts. What is important here is that both of these leading accounts of the impairments of psychopaths place those deficits at the level of information processing (cf. Blair & Mitchell, 2009, p. 547) and not at that of motivation. This latter level concerns the effects of internalized information, for example, on motor execution. In fact, these views try to explain the peculiarities in the behavior of psychopaths by reference to deficits in forming affect representations or in the automatic allocation of attention (cf. Moul, Killcross, & Dadds, 2012, p. 792). We support these claims by focusing, first, on Blair's theory.

James Blair's IES (see Blair et al. 2006, p. 157) articulates the view, that has been famously advanced by Antonio Damasio (1994) as the *somatic marker hypothesis*, that psychopaths have emotional impairments that affect their decision making and learning (for the fundamental differences between Blair's and Damasio's approaches, see Blair, Mitchell & Blair, 2005, pp. 92-94). IES explains the deficits of psychopaths in instrumental learning task in terms of impairments in neural mechanisms, most notably amygdala and ventromedial prefrontal cortex, that are involved in affectively targeting or representing certain stimuli (see also Blair, Mitchell, & Blair, 2005; Blair, 2006, 2008). These mechanisms are taken to be responsible for categorizing or evaluating certain options or prospects and their outcomes as good or bad. By emotionally marking a certain type of stimulus as good or bad, the function of these neural mechanisms is to establish associations between that type of stimuli and responses, thus inducing a person to pursue that course of action or to refrain from it.

According to the IES model the problem on instrumental tasks is reduced to the problem of the inaccessibility of the affective representations that code certain choice options as bad

and others as good. For example, consistently with the present account Blair and colleagues propose that a:

> [d]ysfunction within the amygdala alone may make the individual *less sensitive* to changes in reinforcement contingencies and thus only give rise to impairment on tasks such as the one-pack card playing task and the gambling task where the changes are less easy to *detect*. (Blair, Colledge, & Mitchell 2001, p. 508, our emphasis)

The emphasis here is on the dysfunction in the amygdala that, according to the present hypothesis, makes psychopaths less sensitive in detecting the changes in instrumental tasks that are relevant for successfully solving those tasks. Moreover, the inability to detect the relevant information is due to impairments in *affective coding* of stimulus-reinforcement information (cf. Blair et al. 2006, p. 163). The IES model supports the claim that the task-relevant information is inaccessible to psychopaths because the right coding format, namely the affective representations, necessary for detecting changes and solving different instrumental tasks is defective and therefore unavailable to them (ibid., p. 157). In other words, the information is not coded in the way that enables the formation of stimuli-punishment (or reward) associations that can be used for direct choice of actions (Blair et al. 2006).

Similarly the response modulation hypothesis (RMH), advanced by Joseph Newman (1998) and his associates, predicts that the performance of psychopaths in instrumental learning derives from the actual inaccessibility of some information (Hiatt & Newman, 2006; Koenigs & Newman, 2013). As the supporters of this view state:

> behavior that characterizes criminal psychopaths results from failing to access information that nonpsychopathic individuals do access. (MacCoon & Newman, 2006, p. 803)

According to the RMH model, psychopaths have deficits in shifting attention from the stimuli that are in their primary focus to secondary or contextual cues, which are outside of their primary focus. For example, on this account, the experiments with psychopaths on passive avoidance and response reversal show that they are impaired in directing attention to the peripheral or contextual information that is pertinent to the task. Psychopaths do not automatically shift attention in response reversal or gambling tasks because they do not detect the changes in the reinforcement contingencies (e.g. that the response to previously rewarded stimuli is now being punished) (Blair, Colledge, & Mitchell, 2001; Mitchell et al., 2002). In other words, the reinforcement contingencies are inaccessible to the psychopath because once the dominant set of responses is established then "due to a difficulty in switching attention from their primary focus to less salient contingencies in their environment" they are unable to notice (or they tend to miss) the information that is outside of their primary focus and thus cannot be used by the system to automatically adjust action (Moul, Killcross, & Dadds, 2012, p. 790, see also p. 791; see also Koenigs & Newman, 2013; Newman, 1998).

There are, thus, two important respects in which psychopaths seem to be similar to a colorblind person. First, what explains the performance on instrumental tasks is the deficit in the sensory domain, which, although it can have important effects in the decision-system, pertains to what we could call its "periphery". This is why we take those inaccessibility-inducing impairments not to count as issuing in irrationality. Second, the impairment exposed on instrumental tasks is not global. If we were to change colors in the task, the color-blind person could perform normally. Similarly, when the stimuli are made more salient in certain respects, psychopaths tend to solve many instrumental tasks successfully (Koenigs & Newman, 2013; Moul, Killcross, & Dadds, 2012; these cases will be introduced in section 4 and 5). This point is important because it precludes the argument

that the inaccessibility of the information shows that psychopaths lack general capacities that are necessary for ascribing rationality.[1]

To recapitulate, two claims are so far central for our argument that the experiments on the instrumental learning in psychopaths do not show that they fail to satisfy the principle of practical rationality (i). Our first claim is that the two principal explanations of the performance of psychopaths in the instrumental learning tasks can be reasonably interpreted as showing that the relevant means are not actually available to them. The second claim is that *actual epistemic accessibility* spells out the notion of availability relevant in principle (i). However, both claims can be criticized. We describe these worries and sharpen our argument by responding to them in the following sections. Let us begin with our second claim that concerns actual epistemic accessibility.

4. Epistemic accessibility and relevantly similar scenarios

Maibom's line of reasoning could be defended by maintaining that principle (i) refers to *dispositional* or *counterfactual* epistemic accessibility. In this case, a consideration is accessible to the agent if she is *able* to know or become aware of the relevant information. The ensuing principle of instrumental rationality (i) is more demanding than the one that involves actual accessibility. According to this formulation, for instance, the person who drinks petroleum is not rational because, despite her not believing that the glass contains petroleum, she knows that the bartender is her arch enemy and, therefore, she *could* have known that there is something wrong with the drink. Without questioning the plausibility of this reading of epistemic accessibility, let us see whether it can be used to support the conclusion that psychopathic impairments in instrumental learning justify the conclusion that they do not will the necessary means that are available to them.

A plausible formulation of epistemic accessibility should be psychologically realistic. In fact, the 'ought to' in a rational requirement implies a 'can' concerning the psychological abilities of the agent. We propose, thus, to restrict the scope of counterfactual epistemic accessibility in terms of certain *relevantly similar scenarios*.[2] The idea is that some information is accessible to an agent in a certain situation when there is a scenario, which is relevantly similar to that situation, where the agent is sensitive to that information and uses it. Scenarios are relevantly similar to the given situation when the tasks are similar, relevant capacities (or incapacities) of the agent are held fixed, and there are no external conditions that would hinder those capacities (or compensate the incapacities). Let us consider, for instance, the case of a blind person who in a circumstance cannot use relevant visual information about colors as means for some of her ends. We cannot say that that information is accessible to her and that she is irrational because in a scenario where she is cured she would be able to access that information and use it appropriately in the same task. That scenario is not relevantly similar to the situation at issue. In fact, it does not involve a psychological incapacity that should be held fixed across the relevant scenarios that are used to establish what the person is capable of doing.

To our conclusion that psychopaths are impaired in accessing the relevant information in certain learning tasks, it could be then objected that that there is evidence of scenarios where they do access such information. For example, in 'punishment-only' version of the passive avoidance task, where successful solution depends on learning to avoid punished stimuli and respond to stimuli that are not punished, psychopaths show no impairments (Blair, 2006; Newman, 1998). Similarly, in go/no-go tasks where selecting cards from a deck was either punished or rewarded, psychopaths performed normally when the requirement of the task was to *effortfully* allocate attention to the cumulative net earnings at the point of decision (Koenigs & Newman, 2013; Newman, Patterson, & Kosson, 1987).

Even in response-reversal tasks psychopaths are not completely unsuccessful. Rather, they commit more errors and are slower to learn to adapt their behavior. Under some conditions they even perform equally well to the comparison group (see Brazil et al., 2013a). From these considerations one might again argue that psychopaths are more irrational than the comparison groups on the original tasks.

We respond that the previous argument overlooks the fact that these experiments do not offer *relevantly similar* scenarios to the classical results on instrumental learning. We need the distinction between *personal level* and *sub-personal level* explanations and their mutual relations to articulate this point. Roughly speaking, the personal level explanations account for the behavior and mental life of an individual by referring to mental states, usually spelled out as propositional attitudes (Bermúdez, 2005, pp. 30-31). States that we are conscious of are paradigmatically personal states. This awareness is taken to involve sensitivity to the kind of revisions required of these mental states and the inferential transitions that they afford given the changes in the world or in our mental economy (Broome, 2013; Millar, 2004). So, personal level states are cognitively penetrable because they involve a rational sensitivity to changes in propositional attitudes. For example, perception that there is a duck or a rabbit projected on a screen will be influenced by beliefs about what we see on the screen. In addition, these states normally exhibit an inferential integration (see Bermúdez, 2005, p. 31). For example, having a desire for a certain item and the belief that a certain course of action will satisfy that desire render rational and intelligible forming the intention to perform that action and acting accordingly.

Sub-personal level explanations, on the other hand, concern the operations of biological or psychological systems that do not operate with personal level mental states and their specific relations. Below the level of the person, explanations invoke concepts of cognitive functioning that operate on different time scales and levels of organization. For example, a

sub-personal type of explanation could involve visual classification of stimuli that takes place at the level of an encapsulated perceptual module.[3]

The instrumental learning experiments where psychopaths perform as controls are not relevantly similar scenarios for evaluating whether they can access the relevant information when they show an impaired performance. This is so because the sub-personal automatic mechanisms involved are different. In fact, the response modulation hypothesis (Koenigs and Newman, 2013) and integrated emotions system (see e.g. Blair et al., 2006) support only the deficiency in the sensory domain. That is, psychopathic individuals either have deficits in automatic focusing of attention to the relevant stimuli (RMH) or the error-related information is not affectively represented in the cognitive system (IES). Therefore, in both cases, the information cannot be used at the personal level for successfully solving the task (see e.g. Blair, et al., 2006, p. 157). There is ample evidence that different neural mechanisms underlie dispositions to successfully solve the particular instrumental learning task or its component. For example, Blair (2006) argues that the difference between passive avoidance tasks and the acquisition phases of response reversal tasks where psychopaths are successful is due to the fact that different brain regions are being utilized. However, a less central role can be assigned to the specific sub-personal mechanisms in fixing the psychological capacities that ground similarity of the scenarios relevant for what is epistemically accessible at the personal level.

The argument that certain considerations are counterfactually epistemically accessible to psychopaths can be based on the assumption that the personal level can abstract away from the specific mechanisms that might fail to operate in specific instances. In accordance with this line of reasoning, what is relevant is that *the person* is capable of recognizing certain considerations. In fact, there are experiments that show that psychopaths are as sensitive as non-psychopathic controls to the means relevant for solving the task. For

example, in one pack gambling task when participants are able to view their net earnings and when they are required to pause before making a choice, psychopaths solve the task as well as non-psychopaths (cf. Koenigs & Newman, 2013, p. 96). Similarly, in passive avoidance tasks when the avoiding of the punishing stimuli was made explicit by, for example, introducing money as a reward that can be won, psychopaths' performance was normalized (ibid., p. 95).

We respond that these experiments show that when psychopaths perform as non-psychopaths, there are not only differences in the sub-personal mechanisms involved, but also differences at the personal level. Therefore, these experiments do not offer relevantly similar scenarios. Most of the experiments differ from those where psychopaths fail in certain respects; in rewards, priming of attention, or simply in the instructions that are explicitly given to the participants. Which sub-personal mechanisms will activate personal level states depends on the nature of the task and the instructions provided to the participants (cf. Brazil et al., 2013a).

This is especially shown in experiments where psychopaths fail to respond to certain stimuli when the task primes the automatic and non-voluntary processes of attention shifting, but manage to respond to the same stimuli when the task relevant information is perceived by effortfully adjusting attention to the relevant contextual stimuli. For example, this is demonstrated in the Flanker task where the participant has to respond to a target stimulus and ignore the distracter, flanked, stimuli. In the letter flanker task, for example, the participant is shown letters A or S that demand left response or letters G or H that demand right response. In the congruent condition the target letter is surrounded by the flankers that demand the same response (e.g. AAASAAA). In the incongruent condition the target letter is surrounded by flankers that demand incompatible response (e.g. GGGAGGG). When the task is cued by first indicating where the target stimulus will

appear, psychopaths showed less distractor interference than non-psychopaths. However, when the initial cue highlighted both places, where the target and the distractor stimuli were supposed to appear, then both groups of participants displayed roughly the same interference effects (Zeier, Maxwell, & Newman, 2009). This indicates that when psychopaths perceive the stimuli, which in this case involve effortful allocation of the attention, they display normal reactions in the decision-making process (see Koenigs & Newman, 2013, p. 99). Here it must be noted that in the Flanker task, psychopaths' disregard of the secondary, flanker information enables them to perform better than the non-psychopathic participants.

To sum up, psychopaths seem to be oblivious to certain aspects of the situation and therefore they often seem not to will the means for their ends. However, this alleged conclusion is not evidentially supported. The emerging picture of a psychopath as portrayed by the instrumental learning tasks considered so far involves a type of person who, compared to non-psychopaths, has abnormal functioning of the decision-making system (ibid.). But the literature on which Maibom (2005) relied seems to support only the conclusion that those abnormalities pertain to information detection and not to willing the accessible means that are necessary and/or sufficient for accomplishing some end. However, there are some recent studies that are worth considering in assessing this issue. In the next section we introduce these researches and discuss what conclusion could be drawn from them. Moreover, based on our view of this recent experimental work we advance some recommendation for future experimental studies.

5. Other evidence on the instrumental learning of psychopaths

Certain studies show that the differences between psychopaths' and non-psychopaths' propensity to solve the instrumental tasks sometimes lie in the condition in which the task

requires engagement in effortful, intentional adaptation of behaviors (see e.g. von Borries et al., 2010; Brazil et al., 2009; Brazil et al., 2013a). For example, in a probabilistic response reversal study done by Inti Brazil and colleagues, the authors confirmed their prediction that when the psychopathic participants are not aware of the learning component of the task, they perform equally well as non-psychopaths (Brazil et al., 2013a).[4] However, when the task primes active and intentional monitoring of the probabilistic relationship between the cue and the stimuli thereby priming active learning, then the psychopaths do less well than the control group. The difference between the performances in the two components of the task seem to be underpinned by two distinct mechanisms, as indexed by the EEG and fMRI studies (Brazil et al., 2009; O'Connell et al., 2007; Orr & Carrasco, 2011; see also note 8). As argued before we think that the failures in the second condition are due to the inaccessibility or lower accessibility of the error information to the psychopath's awareness that is relevant for optimally solving the reversal component of the task. However, other studies seem to put pressure on our interpretation.

Recent studies appear to show that ventromedial prefrontal cortex patients (VM patients from now on) on a conscious level often recognize the correct strategies on the gambling and reversal tasks but continue to act disadvantageously (Rolls et al., 1994; Bechara et al., 1999). In particular, in relation to the issue of instrumental rationality, there are very impressive reports of what these patients say about their performances:

> On the first reversal trial patients generally made some objection or comment that clearly showed their awareness that the contingencies had changed (for example, "they've switched!", or "it's changed over"), and the same was true when extinction began. Some patients seemed to try to instruct themselves to respond in a way that would yield points, but without success. (Rolls et al.,

> 1994, p. 1521)

These data could be plausibly interpreted as showing that the information relevant for the task is actually accessible to VM patients, but they do not use it. On this basis, it could be argued that they exhibit instrumentally irrational behavior. Since there is evidence that incarcerated psychopaths also exhibit ventromedial cortex deficits (Blair, 2008) one could argue by analogy that the same consideration applies to psychopaths.

Now it has to be acknowledged that grounding expectations about psychopaths on information from VM patients may not be warranted. The two conditions differ. While VM patients show increase in *reactive* aggression, the distinctive feature of psychopathy is the pervasiveness of *instrumental* aggression (cf. Blair, Mitchell, & Blair, 2005, pp. 93-94).[5] So even though psychopaths and VM patients may in certain respects have similar neural deficits their ultimate difference in behavioral patterns may not warrant the analogical reasoning.

However, Brazil and colleagues appear to suggest the availability of the relevant information to psychopaths (see e.g. von Borries et al., 2010; Brazil et al., 2009; Brazil et al., 2013a). In one of their more theoretical papers they claim that psychopathic traits:

> play an important role in the active implementation of available information to guide changes in behavior. This suggests that impairments in associative learning previously found in clinical psychopathy might also be (partly) due to a deficiency in *using* reinforcement information appropriately to drive behavior (…). (Brazil et al., 2013b, p. 8, emphasis in the original)

Similarly, they claim that:

> the previous findings from our laboratory suggest that in individuals with psychopathy (…) impairments are present when adaptation relies on

> intentional use of available information. (Brazil et al., 2013a, p. E14)

Explicit mentioning of the *available information* that is not *used* by individuals with psychopathic traits might indicate that the task relevant information is actually accessible to individuals with psychopathy but that they simply do not use it in decision-making process. This reading might be taken as contradicting MacCoon and Newman's (2006) claim cited above. If this were the case, then we could plausibly say that psychopaths are instrumentally irrational because they do not intend to adopt the available means necessary and/or sufficient for accomplishing their ends.

However, Brazil and colleagues (Brazil et al., 2013a, 2013b) do not provide evidence for the conclusion (nor that seems to be in their focus) that psychopaths are aware of the relevant information when solving the task and do not respond to it as if they are subject to some motivational or motor-executive problems. They maintain that the findings on VM patients support their prediction that psychopaths will show impairments in response reversal task when the participants are aware of the task instructions, i.e. when the nature of the task promotes intentional, controlled behavior adaptation (Brazil et al., 2013a, p. E14; see also Brazil et al., 2013b). From this we can infer that VM patients and psychopaths are similar with regard to instrumental learning tasks because they are *aware* of the task requirements and instructions. But this similarity does not rely on the further assumption that when solving the task psychopaths are also aware of the relevant information or that they recognize the errors but cannot make themselves to use the information to change their responses appropriately.

Furthermore, on a close inspection of these studies, it seems that the proper interpretation of the notion of 'using the available information' does not contradict MacCoon and Newman (2006) and it does not warrant the conclusion that psychopaths do

not will the necessary and/or sufficient accessible means for their ends. In von Borries et al. (2010), Brazil et al. (2009), (2013a), and (2013b), it is reported that psychopaths have problems in monitoring when a performance error occurs and then to use that error information in adapting their behavior to successfully solve the task. As opposed to the original studies on instrumental learning, these more recent studies report a more specific deficit that psychopaths exhibit in comparison to non-psychopaths.

When performing some act it is important to recognize when an error has been made in order to change the behavior and successfully perform the intended action. This process of error monitoring has two components that are reminiscent of the general dual processing theories in cognitive psychology (Ullsperger & von Cramon, 2006). The first component refers to the early stage of error processing, which happens automatically, usually without awareness that the error was committed. Normal consequence of early error processing is the slowing down of the behavior that was being performed. The second component refers to *error signaling*, which involves "an intentional, much slower, and complex process based on conscious error recognition." (Brazil et al., 2009, p. 138). In general we can say that while the first component detects error at the non-conscious level and adapts behavior automatically, the second component normally involves conscious error recognition and intentional adaptation of the behavior.[6]

Psychopaths have problems in the second component of error monitoring, while on the automatic level they seem to perform normally (Brazil et al., 2009; Brazil et al., 2013a). This distinction between early and late error processing allows us to give an interpretation of the notion 'available information that is not being used' as it applies to psychopaths. There are two ways in which error information in response reversal tasks seems to be available to psychopaths (Brazil et al., 2013a). First there is the trivial sense in which the available information is just *there*, and from the third person perspective we know this

because the performance error factually occurred. But that does not mean that the error was available from the perspective of the agent (the psychopath) who is trying to solve the task.

In the second sense that is relevant for our purposes, the error information is available because it might be (and sometimes is) detected by the automatic, early error processing system. But in this case it must be noticed that there is no implication that the information which is available in the second sense is also available in the sense in which the individual *consciously recognizes* it in a representational format that could be used for intentional adaptation of the behavior.

To support our case, we would like to advance a possible hypothesis on how the information is available to psychopaths and why they do not use it. Psychopaths do not use the *available* information because, due to an impairment in the later intentional error monitoring system, i.e. it is not accessible to them at the level at which the error signal could be recognized as such.[7] Namely, by using EEG studies (see note 6) Brazil et al. (2009) have shown that psychopathic participants are on average *less aware* of the errors that they make in the task. We have, thus, a reason to think that the task relevant information is on average *less available* to the psychopaths.[8]

This is in line with MacCoon and Newman's (2006) suggestion that non-psychopath's range of accessibility of the task relevant information is greater than the psychopath's. Therefore, we conclude that if epistemically accessible means are those that are actually available to the agent, the recent experiments on instrumental learning do not show that psychopaths have impaired instrumental rationality when the accessibility relation is taken into consideration.

However, we are aware that our conclusion might be more directly supported or refuted by experiments of a different type. Deciding on the issue whether the performance of psychopaths in instrumental learning task shows that they are instrumentally irrational

requires considering personal level mechanisms. Specifically, the presence of verbal reports during the tasks, as of those of VM patients mentioned above, could show the psychopaths' awareness of the means available to pursue the end of the task. Their failure to adopt these means might then be an indication of impairment in instrumental rationality.

6. Conclusion

Investigating the rationality of psychopaths offers a particular instance of the general challenge of relating philosophical accounts of this notion to specific empirical results concerning the performance of certain classes of individuals in experimental conditions. Heidi Maibom, and other philosophers who have followed her seminal work, have found that experiments about instrumental learning in incarcerated psychopaths support the conclusion that they are instrumentally irrational. By investigating philosophically the requirements of this latter notion and their relation to the observed behaviors and their current scientific explanations, we have challenged this reasoning.

Moreover, in the last section we have advanced our hypothesis according to which informational inaccessibility stands at the root of psychopath's problems in solving the instrumental tasks, and not the failure to translate relevant information into action, that is the more general problem of motor execution. And lastly, on the basis of studies involving VM patients, we have proposed to test this hypothesis by including verbal reports that could show what information psychopathic participants are aware of, and how they construe the situation when they try to solve the relevant instrumental learning task.

Notes

[1] We thank one of the anonymous reviewers for pressing us to make more precise the empirical case for the inaccessibility of the relevant information to the psychopath and the way in which this type of inaccessibility is pertinent for evaluations of rationality.

[2] We are very grateful to one anonymous reviewer for suggesting, besides other useful stylistic improvements, the notion of relevantly similar scenario to set out our argument more perspicuously.

[3] It is important to stress that sub-personal, unconscious and automatic perceptual states normally provide inputs to perceptual states of which we are aware, such as the fact that visual processing of information provides me with an input that I can recognize as my table and consequently classify it as falling under the concept of being light grey. This perceptual state can induce me to integrate that information by forming a rational belief that the table in front of me is light grey.

[4] The fact that participants are not aware of the learning component of the task means that they do not get the instruction that there is a probabilistic relation between the cue and the stimulus to which they were supposed to respond or withhold the response. They do not also get the information that this relation will change in the reversal phase. This condition was supposed to prime automatic learning of the connection between the cue and the stimuli, and then the task was to automatically learn the new relation between the cue and the stimuli when the reinforcement relations change.

[5] Reactive aggression refers to aggressive behavior that is triggered by the agent's frustration and impulsivity, i.e. lack of forethought. Instrumental aggression refers to the utilization of aggression and ensuing violence as a means to a certain goal. Usually it involves cold calculation and planning. We do not claim that psychopaths do not exhibit reactive aggression, they most certainly do. Here we stress the important point that instrumental aggression distinguishes, almost as a defining feature, psychopathy from other disorders that involve aggression. See also Maibom (2005), where it is maintained that there are significant attentional differences in VM patients and psychopaths.

[6] In the neuroscientific literature the difference between early unconscious processing of errors and late conscious signaling of errors has been drawn in terms of EEG studies (they were also confirmed by fMRI studies). The waveform related to the first component of error processing is termed error-related negativity (ERN). The second component is related to waveform termed error positivity (PE). The brain region that is believed to underpin these processes involves anterior cingulate cortex (ACC) and related areas (see O'Connell et al., 2007).

[7] Brazil et al. (2013b) leave it open whether people with psychopathic traits have deficiencies in using the relevant information because of the deficiencies in the sensory domain, motor domain or in the interaction between the two domains (see ibid., p. 8). However, Brazil (personal communication) endorses our interpretation of the studies done by him and colleagues. Furthermore, he also agrees that the more probable explanation of the psychopathic deficiencies, which is in line with IES and RMH, lie in the sensory domain and not in some other domain that pertains to motivation and/or motor-execution.

[8] If the task relevant information (i.e. the performance errors that psychopathic participants make in carrying out the tasks) is unavailable to psychopaths' conscious awareness then that could explain their deficits in instrumental learning and similar tasks. For example, it is hypothesized by O'Connell et al. (2007, p. 2578) that error awareness, unlike non-conscious error detection, "may engender broader adaptations of a performance strategy that are likely to result in longer term changes in behaviour." Similarly Orr & Carrasco (2011, p. 5891) contend that "[c]onsciously perceived errors would be more significant, and more likely to lead to correcting one's erroneous actions, than unperceived errors."

This is a preprint version of the paper: Jurjako, M. and Malatesti, L. 2016. Instrumental rationality in psychopathy: implications from learning tasks. *Philosophical Psychology*, 29(5): 717-731.

References

Aaltola, E. (2014). Affective empathy as core moral agency: Psychopathy, autism and reason revisited. *Philosophical Explorations, 17*, 76-92.

Bechara, A., Damasio, A. R., Damasio, H., & Anderson, S. W. (1994). Insensitivity to future consequences following damage to human prefrontal cortex. *Cognition, 50*, 7–15.

Bechara, A., Damasio, H., Tranel, D., & Damasio, A. R. (1999). Deciding advantageously before knowing the advantageous strategy. *Science, 275*, 1293-1295.

Bermúdez, J. L. (2005). *Philosophy of psychology*. New York/Oxford: Routledge.

Blair, J. R. (2006). Subcortical brain systems in psychopathy: The amygdala. In C. P. Patrick (Ed.), *Handbook of Psychopathy* (pp. 296-312). New York, London: Guilford.

Blair, J. R. (2008). The amygdala and ventromedial prefrontal cortex: functional contributions and dysfunction in psychopathy. *Philosophical Transactions of the Royal Society B, 363*, 2557-2565.

Blair, J. R., & Mitchell, D. G. (2009). Psychopathy, attention and emotion. *Psychological Medicine, 39*, 543-555.

Blair, J. R., Colledge, E., & Mitchell, D. (2001). Somatic markers and response reversal: is there orbitofrontal cortex dysfunction in boys with psychopathic tendencies? *Journal of Abnormal Psychology, 29*, 499–511.

Blair, J. R., Mitchell, D. G., & Blair, K. (2005). *The psychopath: Emotion and the brain*. Oxford: Blackwell Publishing.

Blair, K. J., Morton, J., Leonard, A., & Blair, J. R. (2006). Impaired decision-making on the basis of both reward and punishment information in individuals with psychopathy. *Personality and Individual Differences, 41*, 155-165.

Brazil, I. A., de Bruijn, E. R., Bulten, B. H., von Borries, A. K., van Lankveld, J. J., Buitelaar, J. K., et al. (2009). Early and late components of error monitoring in violent offenders with psychopathy. *Biological Psychiatry, 65*, 137-143.

Brazil, I. A., Maes, J. H., Scheper, I., Bulten, B. H., Kessels, R. P., Verkes, R. J., et al. (2013a). Reversal deficits in individuals with psychopathy in explicit but not implicit learning conditions. *Journal of Psychiatry and Neuroscience*, *38*, E13-20.

Brazil, I. A., Hunt, L. H., Bulten, B. H., Kessels, R. P., de Bruijn, E. R., & Mars, R. B. (2013b). Psychopathy-related traits and the use of reward and social information: a computational approach. *Frontiers in Psychology*, *4*, 1-11.

Broome, J. (2013). *Rationality through reasoning*. Oxford: Wiley-Blackwell Publishing.

Cooke, D. J. & Michie, C. (1999). Psychopathy across cultures: America and Scotland compared. *Journal of abnormal psychology*, *108*, 55-68.

Damasio, A. R. (1994). *Descartes' error: Emotion, rationality and the human brain*. New York: Putnam.

Davidson, D. (2001). *Essays on actions and events*. Oxford: Oxford University Press.

Fisher, L., & Blair, J. R. (1998). Cognitive impairment and its relationship to psychopathic tendencies in children with emotional and behavioural difficulties. *Journal of Abnormal Child Psychology*, *26*, 511-519.

Forth, A. E., Kosson, D. S., & Hare, R. D. (2003). *The Psychopathy Checklist: Youth version*. Toronto: Multi-Health Systems.

Fowler, K. A., & Lilienfeld, S. O. (2013). Alternatives to Psychopathy Checklist-Revised. In K. A. Kiehl, & W. P. Sinnott-Armstrong (Eds.), *Handbook on psychopathy and law* (pp. 34-57). Oxford: Oxford University Press.

Glenn, A. L., & Raine, A. (2014). *Psychopathy: an introduction to biological findings and their implications*. New York: New York University Press. Hare, R. (2003). *The Psychopathy Checklist-Revised*. Toronto: Multi-Health Systems.

Hare, R., & Neumann, C. S. (2010). Psychopathy: assessment and forensic implication. In L. Malatesti, & J. McMillan (Eds.), *Responsibility and psychopathy* (pp. 93–124). Oxford: Oxford University Press.

Hiatt, K. D., & Newman, J. P. (2006). Understanding psychopathy: The cognitive side. In C. Patrick (Ed.), *Handbook of psychopathy* (pp. 334–352). New York: Guildford Press.

Kant, I. (1785/1993). *Grounding for the metaphysics of morals*. Translated by J. W.

Ellington. Indianapolis/Cambridge: Hackett Publishing.

Kennett, J. (2006). Do psychopaths really threaten moral rationalism? *Philosophical Explorations*, *9*, 69–82.

Kennett, J. (2010). Reasons, emotion, and moral judgment in the psychopath. In L. Malatesti, & J. McMillan (Eds.), *Responsibility and psychopathy: Interfacing law, psychiatry and philosophy* (pp. 243-260). Oxford: Oxford University Press.

Kiehl, K. A., & Sinnott-Armstrong, W. P. (Eds.) (2013). *Handbook on psychopathy and law*. Oxford: Oxford University Press.

Koenigs, M., & Newman, J. P. (2013). The decision making impairment in psychopathy: Psychological and neurobiological mechanisms. In W. P. Sinnott-Armstrong, & K. A. Kiehl (Eds.), *Handbook on psychopathy and law* (pp. 93-106). Oxford: Oxford University Press.

Kolodny, N., & Brunero, J. (2013). Instrumental rationality. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Summer 2015 ed.). Retrieved from: http://plato.stanford.edu/archives/fall2013/entries/rationality-instrumental/

Lapierre, D., Braun, C. M., & Hodgins, S. (1995). Ventral frontal deficits in psychopathy: neuropsychological test findings. *Neuropsychologia, 33*, 139-151.

Lykken, D. T. (1957). A study of anxiety in the sociopathic personality. *Journal of Abnormal and Social Psychology, 55*, 6–10.

MacCoon, D. G., & Newman, J. P. (2006). Content meets process: Using attributions and standards to inform cognitive vulnerability in psychopathy, antisocial personality disorder, and depression. *Journal of Social and Clinical Psychology, 25*, 802-824.

Maibom, H. (2005). Moral unreason: The case of psychopathy. *Mind and Language, 20*, 237–257.

Maibom, H. (2010). Rationalism, emotivism, and the psychopath. In L. Malatesti, & J. McMillan (Eds.), *Responsibility and psychopathy: Interfacing law, psychiatry and philosophy* (pp. 227-241). Oxford: Oxford University Press.

Malatesti, L., & McMillan, J. (Eds.). (2010). *Responsibility and psychopathy: Interfacing law, psychiatry and philosophy*. Oxford: Oxford University Press.

Millar, A. (2004). *Understanding people: Normativity and rationalizing explanation*. Oxford: Oxford University Press.

Mitchell, D. G., Colledge, E., Leonard, A., & Blair, J. R. (2002). Risky decisions and response reversal: Is there evidence of orbitofrontal cortex dysfunction in psychopathic individuals? *Neuropsychologia, 40*, 2013–2022.

Moul, C., Killcross, S., & Dadds, M. R. (2012). A model of differential amygdala activation in psychopathy. *Psychological Review*, *119*, 789-806.

Newman, J. P. (1998). Psychopathic behavior: An information processing perspective. In D. J. Cooke, & A. E. Forth (Eds.), *Psychopathy: Theory, research and implications for society* (pp. 81-104). Dordrecht: Kluwer.

Newman, J. P., Patterson, M. C., & Howland, E. W. (1990). Passive avoidance in psychopaths: The effects of rewards. *Personality and Individual Differences, 11*, 1101–1114.

Newman, J. P., Patterson, M. C., & Kosson, D. S. (1987). Response perseveration in psychopaths. *Journal of Abnormal Psychology, 96*, 145-148.

Nichols, S. (2004). *Sentimental rules: On the natural foundation of moral judgment*. Oxford: Oxford University Press.

O'Brien, B., & Frick, P. J. (1996). Reward dominance: Associations with anxiety, conduct problems, and psychopathy in children. *Journal of Abnormal Child Psychology, 24*, 223-240.

O'Connell, R. G., Dockree, P. M., Bellgrove, A., M., Kelly, P., S., et al. (2007). The role of cingulate cortex in the detection of errors with and without awareness: A high-density electrical mapping study. *European Journal of Neuroscience, 25*, 2571–2579.

O'Neill, O. (2001). Consistency in action. In E. Millgram (Ed.), *Varieties of practical reasoning* (pp. 301-329). Cambridge, MA.: MIT Press.

Orr, J. M., & Carrasco, M. (2011). The role of the error positivity in the conscious perception of errors. *The Journal of Neuroscience, 31*, 5891-5892.

Patrick, C. J. (Ed). 2006. *Handbook of psychopathy*. New York/London: Guildford Press.

Prinz, J. (2006). The emotional basis of moral judgments. *Philosophical Explorations*, 29–43.

Rolls, E. T., Hornak, J., Wade, D., & McGrath, J. (1994). Emotion-related learning in patients with social and emotional changes associated with frontal lobe damage. *Journal of Neurology, Neurosurgery, and Psychiatry*, *57*, 1518-1524.

Samuels, R. & S. Stich. (2004). Rationality and psychology. In P. Rawlings, & A. R. Mele (Eds.), *The Oxford handbook of rationality*. pp. 279-300. Oxford: Oxford University Press.

Sinnott-Armstrong, W. (2014). Do psychopaths refute internalism? In T. Schramme (Ed.), *Being amoral: Psychopathy and moral incapacity* (pp. 187-208). Cambridge MA.: MIT Press.

Ullsperger, M., & von Cramon, Y. D. (2006). How does error correction differ from error signaling? An event-related potential study. *Brain Research*, *1105*, 102-109.

von Borries, A. K., Brazil, I. A., Bulten, B. H., Buitelaar, J. K., Verkes, R. J., & de Bruijn, E. R. (2010). Neural correlates of error-related learning deficits in individuals with psychopathy. *Psychological Medicine*, *40*, 1559-1568.

Williams, B. (1981). Internal and external reasons. In B. Williams, *Moral luck* (pp. 101-113). Cambridge: Cambridge University Press.

Zeier, J. D., Maxwell, J. S., & Newman, J. P. (2009). Attention moderates the processing of inhibitory information in primary psychopathy. *Journal of Abnormal Psychology, 118*, 554-563.