

**UNIVERSITAT DE BARCELONA**

**Faculty of Philosophy**

**Master in Analytic Philosophy**

**Research Paper:**

**GRIFFITHS' PSYCHOEVOOLUTIONARY THEORY OF BASIC EMOTIONS: IS THE  
AUTOMATIC APPRAISAL MECHANISM INFORMATIONALLY ENCAPSULATED?**

Student: Matilde Aliffi

Supervisor: Prof. David Pineda

Academic year: 2013/2014

## **Griffiths' psychoevolutionary theory of basic emotions: is the automatic appraisal mechanism informationally encapsulated?**

### **Abstract**

Griffiths argues that a system which triggers the emotional response, named automatic appraisal mechanism (AAM), is informationally encapsulated (Griffiths, 1997). After proposing a clarification of the AAM using Shea's taxonomy (Shea, 2013), I will claim that Griffiths' inference to the best explanation in favor of the informational encapsulation of the AAM is not compelling. I will present empirical evidence (Paquette et al., 2003) that is incompatible with the thesis of the informational encapsulation of the AAM in order to cast doubts on Griffiths' explanation, and I will propose an alternative one. My alternative explanation will be that the AAM is synchronically impenetrable, and I will affirm that it is preferable over Griffiths' one because is less theoretically demanding, and moreover in accordance with empirical evidence that shows the possibility of diachronic cognitive penetration of the AAM. I will conclude by claiming that this revision can provide also a better account of the irrationality of recalcitrant emotions.

**Keywords:** psychoevolutionary theory of emotions, cognitive penetrability, recalcitrant emotions.

### **Introduction**

Griffiths argues that a system which triggers the emotional response, named automatic appraisal mechanism (AAM), is informationally encapsulated (Griffiths, 1997). This approach is interesting because it offers a clear explanation of why recalcitrant emotions that conflict with the agent's evaluative judgment are possible. However, I will claim that this account is in tension with empirical evidence that shows that recalcitrant emotion can be overcome by cognitive-behavioral therapy (Paquette et al., 2003). In this paper I will firstly propose a clarification of the AAM using Shea's taxonomy (Shea, 2013), in order to understand better the framework of the discussion. Then I will discuss Griffiths' argument in favor of the informational encapsulation of the AAM, and I will motivate why it is not compelling. Since Griffiths' argument is an argument to the best explanation, my strategy will be finding an alternative explanation that is preferable than Griffiths' one, namely that the AAM is synchronically impenetrable. I will argue that my alternative explanation will be preferable over Griffiths' one because less theoretical demanding, and in accordance with empirical

evidence (Paquette et al., 2003), which shows the possibility of diachronic cognitive penetration of the AAM.

I will structure the paper as follows: in the first section I will introduce Griffiths' psychoevolutionary theory of basic emotion, and in the second section I will present the notion of informational encapsulation of the AAM. In the third section I will offer a clarification regarding what counts as memory of the AAM. I will discuss Griffiths' argument in the fourth and the fifth section. In the sixth section I will propose my revision of Griffiths' argument. In the seventh section I will conclude with the advantages of my approach.

### **1. Psychoevolutionary theory of basic emotions**

In my paper I will concentrate on a particular theory of the emotion, Griffiths' psychoevolutionary theory of basic emotions. Griffiths thinks that there are six basic emotions, which are surprise, anger, fear, disgust, sadness and joy. In a nutshell Griffiths' psychoevolutionary approach considers that basic emotions are evolutionary adaptations. Emotions are evolutionary adaptations in the sense that they have been selected throughout the evolution of the species for bringing about effective behavior to cope with the challenge that the environment poses to the organism. For instance, fear is an adaptive response to danger, because it was selected in order to promote fight or flight response to dangerous situations or disgust is an adaptive response to noxious stimuli, because it lowers the risk of ingesting infected food. The emotions that have been left out, like for instance piety, pride or love, are not considered *basic* because according to Griffiths they are better accounted as social constructions<sup>1</sup> rather than evolutionary adaptations. In my work I will concentrate on Griffiths' account of basic emotions.

The psychoevolutionary approach has its roots in Darwin's theory of evolution and in his work about the expression of emotions (Darwin, 1872, 1965). Darwin's studies influenced Paul Ekman's empirical research on the expression of emotions (Ekman, 1972, 1977) that constitutes the ground of Griffiths' theory (Griffiths, 1997). In particular, following Ekman (Ekman, 1977), Griffiths argues in

---

1 Claiming that emotions are social construction means that emotions are a cultural phenomenon, in other words emotions are culturally learned and transmitted.

favor of the existence of so called affect-programs<sup>2</sup>, that is, neural circuits that are probably located in the hypothalamus and associated regions (Griffiths, 1990). According to Griffiths affect-programs are<sup>3</sup> basic emotions, and there is one distinct affect-program for each basic emotion. He also claims that there is an automatic appraisal mechanism (henceforth, AAM), a system that has the function of triggering affect-programs (Griffiths, 1997). When an affect-program is activated, it displays complex, coordinated and automatic patterns of emotional responses. Affect-programs are complex because they are composed by several features: facial and vocal changes, musculoskeletal response, changes of the autonomic nervous system and in the level of hormones. For instance, when the affect-program of fear is activated, the heart rate increases, the body begins to sweat and so on. Those changes are coordinated because those responses occur in recognizable sequences and automatic because they do not need a conscious direction to unfold.

Affect-programs can be triggered by two different systems, the high-level cognitive system and the AAM. The high-level cognitive system is defined as the computational system «in which people use information of the sort they verbally assent to (traditional beliefs) and the goals they can be brought to recognize (traditional desires)<sup>4</sup> to guide relatively long-term actions and to solve theoretical problems» (Griffiths, 1997). Consider Jill that starts to think of possible dangerous outcomes of her action. In this situation the affect-program of fear can be triggered by the high-level cognitive system as a result of her chain of thoughts.

The AAM is the second system that triggers the emotions. According to Griffiths, it is able to operate independently from high-level cognitive processes. When it receives perceptual information, it evaluates rapidly and in an automatic way if it deserves an emotional response or not. Consider as an example Jill who walking on a hiking trip through the forest suddenly spots a big animal, as a reaction to which a fear response occurs.

In line with the psychoevolutionary approach, Griffiths argued that the reason why the AAM has

---

2 Griffiths' main argument in favor the existence of affect-programs is the presence of emotional responses that depends upon an innate factor. For the related discussion see Griffiths, (Griffiths 1990).

3 This identification causes a conceptual revision: basic emotions are identified with affect-programs, and not with the interpretation that the agent gives of his physiological state. That means that patterns of emotional responses are not caused by the emotion but *are* the emotion.

4 Beliefs and desires are here conceived in accordance with the representational theory of mind. They are mental states that have intentionality, in the sense that they are *about* something. We can say that S believes (or desires) that p if S is appropriately related to a mental representation (with a belief-relation, a desire-relation) whose propositional content is that p.

been retained is that it gives an evolutionary advantage, because it is faster than the high-level cognitive system and it can evaluate the situation independently from it. That might give an evolutionary advantage because it permits to accept data which contradicts beliefs, making us to consider data hostile to our current beliefs (Griffiths, 1990). Moreover, according to Griffiths, the AAM has also the capacity to learn associations between stimuli and an emotional reactions; this learning capacity is advantageous because ecosystems are different, so each organism can acquire associations that are typical of its environment.

## 2. Informational encapsulation of the AAM

The notion of informational encapsulation was introduced by Jerry Fodor, relatively to input systems (Fodor, 1983). Fodor conceived the architecture of mind structured with three components: transducers, input systems and central processors. Transducers are processes that have as input the information that comes from the environment (*distal stimuli*) and as output a representation (*proximal stimuli*). Proximal stimuli is then computed by the input system, and then central processors have the function process the output of input systems in order to to fixate beliefs. Consider as an example the case in which Mary sees a red cube in front of her; she firstly receives the information that comes from the environment, which is translated by transducers into the proximal stimuli (the image on the retina), then her (visual) input system computes the proximal stimuli in information available to her cognitive system.

Now, the thesis of informational encapsulation is a restriction on the flow of information into the input system. More specifically, according to Fodor input systems are informationally encapsulated if and only if the following condition holds:

(IE) Given a set of input, processing within the input system cannot be causally affected by information stored in other cognitive systems.

This property is considered the essential feature<sup>5</sup> of modularity (Fodor, 1983, p. 71). That means

---

<sup>5</sup> The other features of modularity are domain-specificity, mandatory operation, limited central accessibility, fast

that if the notion of informational encapsulation fails, then a system would cease to be seen as modular to an interesting extent.

The AAM, according to Griffiths, has some of the features of Fodor's input systems that characterize modularity (Griffiths, 1997, 1990). In this paper I will not discuss Griffiths' argument for establishing all the features of modularity that Griffiths attributes to the AAM. I will rather concentrate on Griffiths' claim that the AAM is informationally encapsulated.

In order to transfer the claim of informational encapsulation of the input systems to the AAM, Griffiths needs to show that the AAM is analogous to an input system under the relevant respects. There is an important difference between the AAM and input systems, that is that while the output of the input system is a mental representation, the output of the AAM is behavioral (the pattern of emotional responses of the selected affect-program). However, in accordance with Griffiths, I think that the possibility to transfer the claim of informational encapsulation of input system is still warranted by the fact that an input system and AAM are both computational systems.

### **3. Further clarification on what counts as AAM**

Prior to a deeper analysis of Griffiths' argument of the informational encapsulation of the AAM, I want to stress the need of further clarification of what counts as an AAM, if we want Griffiths account to be consistent with the claim that the AAM is informationally encapsulated. As I have said, according to Griffiths, the AAM has a form of memory that stores information about classes of stimuli which merit the emotional response. When the AAM receives perceptual information, then the system compares these stimuli to “memories” previously assessed as meriting the emotional response. Griffiths said that he does not mean to «imply a physiological separation between the memory and the decision-making system, [...] the two might be physiologically identical». According to Griffiths, part of this memory stores information that is innate, like the information that loss of balance should activate a fear response (Watson, 1930), but other associations

---

processing, shallow output, fixed neural architecture, characteristic and specific breakdown patterns and characteristic and ontogenetic pace and sequencing. Those properties normally go in a cluster, but it is not necessarily that all those properties are satisfied for a system to be modular. For a detailed description of this feature see Fodor (Fodor, 1983). For a critical discussion about the relation between those properties see Prinz (Prinz, 2006 a).

between classes of stimuli and emotional reactions are learnt (Griffiths, 1997, p. 88). Specifying the characteristics of memory is important for Griffiths' claim that the AAM is informationally encapsulated. The reason is that if the AAM relies on the memory which belongs to the high-level cognitive system, we would have top-down influences on the working of the AAM. A top-down influence is the effect of previously stored information on a psychological process. Consider the situation in which Mary believes that foreigners are dangerous and this belief  $b$  is stored in her memory. If by seeing a foreigner, the AAM will trigger fear drawing on a piece of information, say the belief  $b$ , that was stored in the high-level cognitive system, then Mary's AAM would not be informationally encapsulated.

If we want to maintain the thesis of the informational encapsulation of the AAM, the AAM's memory should be not coincident with the memory available to the high-level cognitive system. For instance the AAM's memory should be stored within the AAM instead. If this is the case then when it receives an input, the AAM will compare it with previously stored information that does not come from the high-level cognitive system. There will still be a top-down effect, because there will be a causal influence of previous stored information (in the AAM's memory) on the AAM's process. However in this case the background information is stored within the AAM, and then there would not be a case of cognitive penetration of the AAM.<sup>6</sup> The thesis of cognitive impenetrability could be maintained, at the price to draw a distinction between the memory of the AAM and the memory of the high-level cognitive system.

In my opinion, it is useful to follow Nicholas Shea's taxonomy (Shea, 2013) in order to elucidate what is a top-down influence within the system, and to characterize better what it is meant by having a memory that is embodied in the AAM. In order to do this I will firstly define what it is meant by occurrent representation. An occurrent representation is a piece of prior information that is tokened in order to causally affect a psychological process. Imagine for instance that you think that full skirts are uncomfortable. This representation is stored in your memory, but this thought does not always come into your mind. When it does, for instance when you are deciding which skirt you want to buy, the representation is tokened in order to causally affect your decision process and it is thus occurrent. Now we can present the distinction between the concepts of implicit and explicit (mental) representation. On the one hand an *implicit representation* is a disposition to

---

<sup>6</sup> This is acknowledged by Fodor (1983) and Macpherson (2013).

transition between two or more occurrent representations that can have no influence on subsequent processing except via the representations between which the disposition subsists. Consider a psychological process that makes transitions from a light contrast map to representation of where the edges of the object lie. This system encodes the implicit representation that edges of objects tend to occur where there are discontinuities in the light contrast map, because this information affects subsequent processing but only in virtue of its effect in producing the representation of the location edges<sup>7</sup>. On the other hand, an *explicit representation* is an occurrent representation that can affect many pieces of subsequent processing. The representation of the location of edges is an example of an explicit representation because it can affect subsequent processing like object discrimination and object categorizations<sup>8</sup>.

We can adapt Shea's framework to our discussion about what counts as the AAM's memory.

Consider the case in which the AAM was disposed to make transitions from the stimuli of a snake to the activation of the affect-program of fear. This disposition implicitly encodes the information that snakes tend to occur when there is danger. However, this information would be an implicit representation, because it has effect on subsequent processing, but that influence is only in virtue of producing its effect on AAM's activation of fear. Conversely, the belief that snakes are dangerous is an explicit representation, because it can act as input to the processes of thinking, reasoning, planning and so on. A way of distinguishing top-down effects that are embodied in the AAM from top-down effect that are outside the system is asking respectively whether the causal impact of background information is implicit or whether it is caused by the tokening of explicit representations outside the process.

#### **4. Griffiths' argument in favor of informational encapsulation of AAM**

Griffiths' argument in favor of the informational encapsulation of the AAM consists in presenting a plausible case in which a person has a recalcitrant emotion and in providing the best explanation of why recalcitrant emotions exist. An emotion is said to be recalcitrant when it occurs even if it

---

7 This example is due to Shea (2013).

8 This example is due to Shea (2013).



conflicts with the agent's evaluative judgments. Typical cases of recalcitrant emotions are cases in which subjects are afraid of something regardless the fact that they think that they are safe. Cases of recalcitrant emotions include phobias. Recalcitrant emotions are quite common, and they do not only involve fear responses. Consider, for instance, cases of people who having had a terrible nausea after having eaten spoiled food are disgusted by seeing the same kind of food, even if they know that the food is fresh and not noxious. Or, alternatively, consider that a person P feels anger towards somebody S, thinking that S has done some offense to him, and that, when P realizes S did not do any offense to P, S still feels anger to S<sup>9</sup>.

Griffiths' argument is an argument to the best explanation<sup>10</sup>, and proceeds as follows:

(P1): There are cases of recalcitrant emotions (emotions that persist against the judgment of the subject).

(P2): The best explanation of why recalcitrant emotions are possible is claiming that the AAM is informationally encapsulated.

(C): The AAM is informationally encapsulated.

In this argument, Griffiths considers evidence of recalcitrant emotions (P1). The paradigmatic case considered by him is a situation in which a subject sees an earthworm and its fear response is activated even if he thinks that earthworms are not dangerous. Since the occurrence of recalcitrant emotions is quite uncontroversial, he does not present empirical evidences in favor of P1.

What in general is considered more problematic in the literature about emotions is why these emotions occur, and in particular why the emotional response *persists* against the judgement and the will of the subject. Theories of emotions have to be able to provide a compelling explanation of why recalcitrant emotions exist, but that is often problematic, in particular for cognitivist theories like judgementalism. That is because, if emotions are evaluative judgments, as judgementalism proposes, then recalcitrant emotions cannot be explained without introducing a conflict between

---

9 See Gorenstein for a psychological discussion of cases like this (Gorenstein et al., 2007).

10 As all inferences to the best explanation, the argument is not sound because the premises do not guarantee the conclusion. In inferences to the best explanation, if some evidences occur and the explanation *e* is the best that we have, this explanatory success is a mark of the truth of *e* (but the truth of *e* does not necessarily follow from the premises).

two contradictory beliefs. For judgementalists fearing something is at least in part as judging that something is dangerous. So, in order to explain why, for instance, an agent fears earthworms while judging that worms are not dangerous judgementalists should say that the subject unconsciously believes that worms are dangerous but he consciously asserts to the belief that they are not. That would be problematic for two reasons: because it does not seem that the subject is falling in a contradiction and because the explanation violates logical charity attributing an unconscious judgment to the subject in contradiction to what he knows (Brady, 2009)<sup>11</sup>.

Conversely, Griffiths argued that the fact that recalcitrant emotions are possible is explained clearly by the informational encapsulation of the AAM (P2). If the AAM were informationally encapsulated, then the AAM triggers the emotion without taking into account beliefs that are stored in the high-level cognitive system for evaluating the stimulus. So the explanation of why it is possible that a subject fears earthworms while thinking to be safe would be that the AAM will activate the fear-affect-program without taking into account beliefs about the safeness of the situation. That would explain why recalcitrant emotions are possible without attributing irrationality to the subject. Accepting that this explanation is the best explanation that we currently have of this phenomenon, it is possible to abduce that the AAM is cognitively encapsulated.

This second premise requires as a background assumption the existence of an AAM. A possible way to criticize Griffiths' argument would be denying its existence. However, I will not concentrate on this possibility and I will grant to him this assumption.

Even if Griffiths' explanation of why recalcitrant emotions occur is better than the judgementalist one, I think that it has nevertheless some problems. In what follows I will cast doubts on Griffiths' argument, challenging his premise P2. My strategy will be finding an alternative explanation of why recalcitrant emotions occur, and arguing that it is better than the claim that the AAM is informationally encapsulated. Firstly I will point out another possible explanation that I will call P2' that has less theoretical costs and that is thus more adequate than P2. Secondly, I will raise doubts also on P2' proposing a third possible explanation P\* that is better than P2 and P2'.

---

11 There are other cognitive explanations of why there are recalcitrant emotions that does not imply that the subject is falling into a contradiction. For instance, neojudgementalist theories like Roberts (Roberts, 2003) that sees emotions as concern-based construals, where construals are intended as impressions or imaginations and they impinge in subject's concerns. However also this account is very problematic. For a discussion of it see Brady (Brady, 2009) and Tappolet (Tappolet, 2012).

Griffiths' argument in favor of the informational encapsulation of the AAM has the theoretical cost to imply a separation between the AAM and the other cognitive systems. I think that it is not necessary to have a cognitively encapsulated system in order to account for cases of recalcitrant emotions. A cheaper alternative would be the following:

(P2') The best explanation of why recalcitrant emotions are possible is claiming that the AAM is cognitively impenetrable.

The notion of cognitive impenetrability was introduced by Pylyshyn who claimed that a system is cognitively impenetrable if it «cannot be influenced by such purely cognitive factors as goals, beliefs, inferences, tacit knowledge and so on» (Pylyshyn, 1980). On the contrary, a system is cognitively penetrable if «the function it computes is sensitive, in a semantically coherent way, to the organism's goals and beliefs, that is, it can be altered in a way that bears some logical relation to what the person knows» (Pylyshyn, 1999: 343). The notion of informational encapsulation and the notion of cognitive impenetrability are related, as Fodor acknowledged (Fodor, 1983), however they are not co-extensional. Following Robbin (Robbin, 2010) it is possible to say that the relation between informational encapsulation and cognitive impenetrability is «as genus to species». On the one hand, if a system is informationally encapsulated, then it is also cognitively impenetrable; on the other hand the converse is not true. A system could be cognitively impenetrable, for instance, but not informationally encapsulated if it is encapsulated only relative to beliefs but not to causal influence of representations that come from the high-level cognitive systems. The empirical evidence such as the McGurk effect (McGurk, 1976), seems to show this possibility regarding the auditory system. The McGurk effect is an illusion that occurs when the content of the auditory perception is modified by the visual information. For instance, consider a video in which subject's lips pronounce the syllable “ga”, while in fact the video is dubbed on the sound of the phoneme “ba”. People affected by the McGurk effect combine those different information and report that they hear the phoneme “da” as “gaba” and “baga” instead of “ba”. McGurk effect is quite generally considered as evidence in favor of the claim that auditory speech is encapsulated relative to beliefs but not relative to vision. Here there is a case in which, if this interpretation is

correct, auditory perception is not informationally encapsulated because visual information could affect the processing of the stimuli. However auditory perception might still be cognitively impenetrable. The concept of informational encapsulation and the concept of cognitive impenetrability are thus not co-extensional, since the former includes the latter, while the converse is not true. The argument that Griffiths brings forward supports the claim that the AAM is informationally encapsulated as much as it supports the cognitive impenetrability of the AAM; however, since the claim that AAM is informationally encapsulated is more costly than the claim that the AAM is cognitively impenetrable, I think that recalcitrant emotions could be more economically explained by abducting (P2') rather than (P2). If Griffiths wants to deny this point, he also has to provide a further argument in favor of the informational encapsulation of the AAM.

Moreover, (P2) and also (P2') might still be not compelling, if they were incompatible with the explanation of other empirical evidence regarding recalcitrant emotions. Consider yourself on a roller-coaster. You feel fear, even if you are secured to the coaster car by a resistant lap bar and safety belts, and you know that you are perfectly out of danger. This is a very likely example of a recalcitrant emotion. Now it seems also plausible to think that, through effort and learning you could cease to be scared while being to the same roller-coaster. And if that should work, it would seem to reveal a sort of plasticity in our system that triggers emotions. Now, I think that for our discussion it could be more important to understand if this plasticity occurs, and if it could be the effect of a causal influence of beliefs on the AAM's process. Somebody, however, might object that this situation looks more like a plausible example than empirical evidence in favour of the possibility to overcome recalcitrant emotions.

In order to overcome this weakness, in what follows I will concentrate on empirical evidence that involves phobic subjects. Phobias are anxiety disorders that affect people who experience persistent and intense fear when confronted with phobogenic stimuli. Since in phobias the emotion of fear is persistent, it is possible to consider them as cases of recalcitrant emotion. Even if probably the majority of recalcitrant emotions are not cases of phobia, in this paper I will mention prevalently studies on phobias, because those are the instances of recalcitrant emotions that have been studied more thoroughly, and because Griffiths chooses recalcitrant fear as a paradigmatic example for his argument.

Some studies show that people who suffer of arachnophobia, for instance, are able to overcome the phobia by undergoing therapy (Paquette et al., 2003), (Hauner et al., 2012). In particular I will concentrate on Paquette et al.'s study. This is an empirical study which shows that recalcitrant emotions could be overcome by therapy that is at least in part cognitive, since therapy employed in the study was the cognitive-behavioral therapy (CBT). Cognitive-behavioral therapy consists indeed in combining an exposure-based treatment to the stimuli that trigger the emotional response with an education for correcting beliefs regarding it.

Paquette's study was conducted on a group of twelve subjects that had phobia of spiders, and on a control group of thirteen healthy subjects. All subjects were scanned with a functional magnetic resonance imaging (fMRI)<sup>12</sup> while viewing a film that depicted spiders and another one depicting butterflies, in order to subtract the brain activity when viewing butterflies from the activity measured when seeing spiders. Phenomenologically, while viewing the film phobic subjects displayed fear reactions, but normal subjects did not display any reaction. At the subpersonal level, phobic subjects presented an activation of the right dorsolateral prefrontal cortex and a parahippocampal activation, that was absent in normal subjects. The dorsolateral prefrontal cortex activation was interpreted as the attempt of the phobic subjects to use cognition to lower and control the emotional reaction, where the parahippocampal activation is considered to be an automatic activation of memories that generate spider phobia. During four consecutive weeks, phobic subjects met their therapist that used cognitive-behavioral therapy to help them in overcoming the phobia.

In these sessions, subjects were firstly exposed to exercise books containing pictures of spiders, then they were gradually exposed to film about living spiders, finally they were exposed to real spider and it was asked to them to touch the tarantula. The exposure was guided by the therapist and accompanied with education for correcting misbeliefs about spiders. Specific data about which beliefs were corrected or induced in patients during the therapy are not available; however we know that therapists in the experiment used cognitive-behavioral therapy, so we can say that this education was conducted as cognitive-behavioural practice requires. In particular cognitive behavioral therapists ask to focus on one's own body and thoughts, and they challenge patients' cognitions, regarding for instance their ability to cope with the emotion, such as thoughts like “I

---

12 An fMRI is an imaging technique that measures brain's activity by the changes in blood flow.

won't be able to handle this situation". They also try to eliminate catastrophic thinking of the patients, such as thoughts like "the spider will bite me". Moreover they explain how avoidance and escape behaviour<sup>13</sup> reinforce the phobia and they provide notions about ways of controlling their thought and information on what they have to do during the exposure part of the therapy (Begić, 2010).

This therapy was considered effective since during the fourth session all subjects were able to touch a real spider without reporting fear reaction and after the fourth session, the brain activation pattern of phobic subjects was similar to the one noted in normal control subject, without frontal or hippocampal activity.

In order to fit this empirical evidence in the discussion, I need to see if Griffiths would consider this as a case of recalcitrant emotion that is overcome. Let us assume that his theory about two different systems that trigger emotion is correct. Griffiths would consider that a recalcitrant emotion is overcome if the affect-program is not activated. In order to know if the affect-program is not activated it is not enough to rely on subject's report, because in his account emotions are patterns of emotional responses (affect-programs) and the subject's interpretation of his physiological state is not taken into account in determining if the emotion occurs and which emotion is activated.

However, I think that the study here presented would be accepted by Griffiths as a case in which recalcitrant emotions are overcome. That is because the subject's report of the absence of fear reaction was correlated with data obtained with fMRI about the activation of brain areas. Assuming that in phobic subjects the AAM activates the affect-program of fear, and that the AAM does not activate it in non-phobic subjects, it is possible to interpret the absence of frontal and hippocampal activity as evidence in favour of the fact that the AAM does not trigger any emotional response.

This empirical evidence is not in conflict with (P1), because it does not deny that recalcitrant emotions occur, but it denies just that they are *always* persistent. The question now is to explain why they cease to be persistent. That might depend on what sort of causal influence affects the process of the AAM. On the one hand, if conscious or unconscious beliefs do not have a causal influence on the system, then the AAM could still be considered cognitively impenetrable, but it remains to be explained why the AAM ceases to activate the emotional response. On the other

---

13 By avoidance behavior it is meant the patient's attempt of avoiding phobogenic stimuli.

hand, if beliefs causally affect the process, then it will be needed to investigate if the system is cognitively impenetrable.

I will start by considering the first alternative. The fact that beliefs causally affect the AAM does not follow simply from the fact that recalcitrant emotions are overcome, because the AAM could cease to activate the fear response also because the background information embodied in its process changes. If that is the case, as we said in the third section, then the system could still be considered cognitively impenetrable. A way to overcome recalcitrant emotions along this line was also considered by Griffiths. He cursorily considered the possibility that cases of recalcitrant emotions could be overcome by suggesting a strategy of counterconditioning to cope with those emotions (Griffiths, 1990, p. 184). Classically, counterconditioning therapy was conceived to work non-cognitively, just by creating a new association with the stimulus that gradually lowers the fear response to it. If this new association is just a creation of a new disposition of the system to move from a certain representation to another, as the counterconditioning therapy seems to require, then it does not follow necessarily that background information outside of the AAM is tokened to causally affect the stimuli processing. To put it in Shea's terminology, it would need just an implicit representation. I think that Griffiths here is assuming this view about how conditioning works.

However, it might be the case that even conditioning therapy involves the tokening of explicit representations. For instance it could be that some expectations, like beliefs about the probability of a certain event, causally affect the processing of the stimuli. This is suggested for instance by Southworth and Kirsh's empirical study about the role of expectancy in fear reduction of agoraphobia (Southworth and Kirsh, 1988). In their study, two groups of agoraphobic people are treated by exposure therapy<sup>14</sup>. The subjects in the high expectancy group were told that they were receiving a treatment that had been demonstrated to be effective, while the subject of the low expectancy group were told that they would receive the therapy after participating a two weeks assessment period. The results reveal that higher-level expectation can causally affect the quality of the success of a therapy. If that interpretation is correct, then the AAM would be widely cognitively penetrable, both when the AAM learns and when it unlearns an association by exposure therapy. For the sake of the argument, I will grant to Griffiths the assumption that a therapy that uses just

---

14 A kind of conditioning that works by exposing the phobic subject to the relevant stimulus, in order to create a new association between the stimulus and the absence of danger.

conditioning does not involve tokening of explicit representations. However, the therapy used in Paquette's study is not just a conditioning therapy, because the therapy has also a more “cognitive” part. In the next section I will develop the argument in favour of a causal influence of a belief in the AAM' process.

## 5. Argument from the causal influence of a belief

This argument runs as follows:

A1) Empirical evidences show that recalcitrant emotions could be overcome by an effort (or a therapy) that is at least in part cognitive.

A2) The cognitive part of the therapy has a causal impact on the AAM's process.

-----

C) There is a causal influence of a belief that affects the AAM's processing.

Since Paquette et al.' study (Paquette et al., 2003) shows that an instance of recalcitrant emotion (fear of spiders) could be overcome by a therapy that is at least in part cognitive, A1 is acceptable.

Let's move to second premise A2. To begin with, let us assume that Griffiths' theory of the existence of the AAM is correct. That assumption is consistent with the findings of the study since the interpretation of the activation of the parahippocampal area seems to be consistent with the activation of an automatic mechanism which triggers the emotional response. Now it has to be seen if in this experiment the cognitive part has a causal impact on the automatic system that triggers emotions. Paquette et al.'s interpretation of the results of the experiment is that CBT reduces phobic avoidance by a deconditioning process but also by the psychotherapeutic technique for re-framing beliefs and negative thinking. This, according to Paquette's interpretation has the merit of rendering obsolete the activation of the brain regions that were associated with the phobic response (Paquette et al., 407). Further studies need to be done in order to strengthen this interpretation. However, in general there is reason to think that the cognitive part of the cognitive-



behavioral therapy is causally efficacious in reducing or overcoming the phobia.

I can motivate this claim further by using a *reductio ad absurdum*. Let us assume that the cognitive part of the therapy of phobic subjects is causally inert. If we maintain that recalcitrant emotions are triggered by the AAM, the only case in which we could say that the cognitive part is causally inert is by reducing the impact of the cognitive-behavioral therapy in overcoming recalcitrant emotions to the impact of a behavioural therapy. By behavioural therapy I intend a conditioning therapy<sup>15</sup> in which the patient is exposed to the phobogenic stimuli in order to create a new association between fear elicitors and the absence of danger, where he does not receive also a psychotherapeutic (or cognitive) aid. If the cognitive part is causally inert, then we should conclude that cognitive-behavioral therapy is equally effective as behavioural therapy.

However, there are empirical studies which suggest that cognitive-behavioral therapy is more effective than just behavioural therapy in treating some cases of agoraphobia (Salkovskis et al., 2006), (Mattick et al., 1989). As those studies reveal, cognitive therapy is then in general causally efficacious. So, even if further studies are needed, there is reason to think that the cognitive part of the therapy is causally efficacious in overcoming the phobia also in the cases considered in Paquette et al.'s study. Since recalcitrant emotions can be overcome at least partially through the cognitive part of a cognitive-behavioral therapy, I can conclude that there is a causal influence of a belief that causally affects the AAM's process.

## 6. Is the AAM cognitively impenetrable?

The above mentioned empirical findings are incompatible with the thesis of the cognitive impenetrability of the AAM, because I have argued that there is an influence of an explicit representation that affects the AAM's processing<sup>16</sup>. It is then problematic to explain how (P2) and also (P2') are consistent with the empirical evidence. Given these considerations, I think that it is needed an alternative explanation of recalcitrant emotion, better than (P2) and (P2').

Prior to presenting this alternative explanation (P\*), I will begin by considering that the above

---

15 As I outlined in the fourth section.

16 Note that those empirical findings are also incompatible with the thesis of the cognitive encapsulation of the AAM.

mentioned empirical evidence do not show an *immediate* causal influence of beliefs on the process of triggering emotions. That is because subjects do not cease to have recalcitrant emotions as soon as they acquire the contrasting belief, but they in general need to undergo a process of therapy that requires different sessions to be effective. So there is a sense according to which the AAM is still cognitively impenetrable, but only for a limited amount of time.

This distinction between these two concepts was initially introduced by Paul Churchland in a philosophical discussion with Fodor about cognitive penetration in perception (Churchland, 1988). In particular Churchland tried to undermine Fodor's view on cognitive impenetrability of perception, by focusing on evidence of perceptual learning. With perceptual learning, perception develops different capacities, and Churchland claimed that some of them would involve cases of cognitive penetration. Instead of discussing his rather complex argument, I will focus on his distinction made in perception literature between synchronic and diachronic cognitive penetration. In general a perceptual process is synchronically penetrated if a belief penetrates the processes quickly and very easily, while a perceptual process is diachronically penetrated if its perceptual processing is reconfigured through time.

It is possible to introduce a reference to time in the discussion about the AAM's possibility to overcome a recalcitrant emotion, saying that synchronic cognitive penetration of the AAM occurs when, given a certain input, the AAM *instantly* ceases to trigger the affect-program as a result of a causal influence of a belief. On the other hand, diachronic penetration of the AAM occurs if the AAM ceases to trigger the affect-program after a period of training.

The difference between these two kinds of cognitive penetration is not clear cut because there is not an agreed criterion about where drawing the line. In any case, the learning process of overcoming emotions could count as a clear case of diachronic cognitive penetration, since the AAM ceases to activate the fear response after some session of cognitive-behavioral therapy (Paquette et al., 2003).

Once this distinction is drawn, Paquette et al. empirical finding is incompatible with the claim that the AAM is diachronically impenetrable, but it is not incompatible with the claim that the AAM is synchronically impenetrable. It is also incompatible with the claim that the AAM is informationally encapsulated, because I argued that in the empirical evidences considered, the AAM is causally

affected by the tokening of an explicit representation. Thus, P2 and P2' are not compelling because they face these problems, and because it is possible to find a better explanation of recalcitrant emotions, that is the following:

P2\* The best explanation of why recalcitrant emotion are possible is claiming that the AAM is synchronically cognitively impenetrable.

This new explanation could provide a convincing reason of why recalcitrant emotions occur. According to P2\* those emotions happen because, when a certain association regarding the dangerousness of a certain stimulus is built into the AAM, this association cannot be eliminated instantaneously by the opposite judgment. However, this does not mean that the system has to be in general cognitively impenetrable or informationally encapsulated. An instance of that is the possibility of overcoming a recalcitrant emotion thanks to, among other causes, a cognitive influence. P2\* is a better explanation than P2 and P2', because it is less theoretically demanding, and moreover it is consistent with the empirical data that suggest a causal influence of a belief on the process that triggers the emotion.

Griffiths' argument is then not compelling because his second premise P2 fails, since I have found a better explanation that is less theoretically demanding and fits better the empirical evidence. If he wanted to deny this point, I think he should provide an additional argument in favor of the informational encapsulation of the AAM. Moreover I have proposed some empirical data in favor of diachronic cognitive penetration of the AAM. I think that whoever wants to maintain the claim that the AAM is informationally encapsulated has to propose a different explanation of these empirical data. Accepting the claim that the AAM is diachronically cognitively penetrable requires also re-thinking the concept of modularity of the AAM. That is because if the notion of informational encapsulation fails, then it fails also the essential feature of modularity. One possibility would be outlining a different notion of quasi-modularity along the lines that Prinz proposes (Prinz, 2006 b). However, I will not develop this point in this paper, and I will leave it for future work.

My account does not entail that recalcitrant emotions can always in principle be overcome by cognitive-behavioral therapy. The above mentioned evidence shows that it is likely that recalcitrant

emotions can in principle be overcome by cognitive-behavioral therapy, however it does not establish that it is *always* the case. Consider the hypothesis that there are some emotion elicitors that are innate, and suppose that losing support is an innate elicitor of fear. Now, consider a situation in which subjects lose balance and feel fears while thinking to be safe (for instance while being in an attraction an amusement park). It might be the case that these subjects undertake cognitive-behavioral therapy in order to eliminate their recalcitrant emotion, but that in fact they cannot stop the activation of fear when they lose balance. That could be due to the association so deeply built into their AAM that it is almost impossible to change; as a result, they can just learn how to lower the emotional reaction of fear. Even if that situation will happen, since my account does not entail that all recalcitrant emotions can in principle be overcome, it will not be undermined.

However, if the majority of emotion elicitors were innate, it would be possible to consider evidence of diachronic cognitive penetration more as an exception to the rule of the informational encapsulation of the AAM, rather than a counterexample to it. That would be problematic for my account, however I think it is possible to reply to this objection by saying that it is widely accepted that most of the emotion elicitors are learned, as Griffiths agrees (Griffiths, 1997, p. 89).

One of the virtues of my revised version of the inference to the best explanation is that it fits well within the psychoevolutionary framework. It is possible to say that the synchronic cognitive impenetrability of the AAM gives an evolutionary advantage since it permits to accept data which contradicts belief. That would be adaptive because it might be the case that our beliefs regarding a certain stimuli are wrong, and the cost of failing to respond to danger might lead to death, while feeling fear might save our lives. However, the fact that AAM could be diachronically cognitively penetrated in general would not threaten that possibility, because the system still remains synchronically impenetrable. Additionally, the possibility that the AAM is diachronically penetrable (possibility that P2\* leaves open) would also be adaptive. Imagine the situation in which you have a system that when it faces a certain stimulus triggers fear, even if you think that this stimulus is not dangerous. This fear response makes your life difficult, because it arises against your will, and this hinders you; moreover imagine that your conscious or unconscious belief cannot make anything for changing that response. The only thing you can do is just try to change *indirectly* your disposition to

feel fear, for instance breathing slowly or undergoing a counterconditioning therapy, because your high-level cognitive system cannot have a causal impact on those processes. And imagine another situation in which you have also the possibility to learn how to change your disposition to feel fear also by using your belief, in addition to other indirect ways. The second situation would be clearly better, because it would allow you to reduce more quickly the waste of energy involved in having repeatedly cases of recalcitrant fear. Therefore the possibility that cognition can have a diachronic causal impact on the process of the AAM could still be consistent with Griffith's evolutionary framework.

## **7. Irrationality of recalcitrant emotions**

I think that my revision has also the merit to provide a better account of the sort the irrationality that is involved in recalcitrant emotions. It is widely agreed that recalcitrant emotions involve a sort of irrationality (Tappolet, 2012), (Brady, 2007, 2009). That is grounded in the intuition that recalcitrant emotions have to be overcome, by changing the emotional response. However, if the AAM were informationally encapsulated, it would be difficult to explain why there is some irrationality involved in recalcitrant emotions. One possibility would be claiming that the irrationality resides in the contrast between the content of the AAM's evaluation, and the content of the subject's belief. While beliefs are rationally assessable states, the AAM's evaluation does not seem to be rationally assessable as well. That is because under the hypothesis of the informational encapsulation of the AAM, the AAM is cognitively impenetrable: beliefs cannot causally affect the AAM's process. If beliefs cannot causally affect the AAM's process, it is doubtful why the evaluation of the AAM should be considered rationally assessable. If the system is informationally encapsulated, how could the output of the system be rationally assessable? The evaluation of the AAM seems more a-rational, just as the experience of a subject under the effect of the Müller-Lyer illusion is a-rational in seeing two lines at different length, while thinking that they have the same length.

It would be still possible to find alternative explanations that are coherent with the assumption of the cognitive impenetrability of the AAM, maybe as Brady (Brady, 2009). In any case, I think that my

revision is a nice solution to this problem: rejecting the claim of the informational encapsulation of the AAM in favor of the more modest claim of the synchronic impenetrability of the AAM, and accepting evidence that shows that the AAM could be diachronically penetrable is enough to render the AAM's evaluation rationally assessable to an interesting extent. If our cognition and our thoughts can influence the AAM's evaluation, then there is in a way the possibility, and thus in a way the responsibility to get rid of recalcitrant emotions. The source of the irrationality could be then the misfiring of recalcitrant emotions. Our epistemic *ought* would then be directed towards ridding ourselves of them.

### **Conclusion**

I have argued that Griffiths' argument in favor of the informational encapsulation of the AAM is not compelling, since I find that an alternative explanation of why recalcitrant emotions exist which does not require the claim that the AAM is informationally encapsulated. This is a modest revision: claiming that the AAM is synchronically impenetrable it is still consistent with the claim that the AAM is informationally encapsulated. However, I have shown that there is empirical evidence like Paquette et al.' study (Paquette et al., 2003) that could be interpreted as showing that the AAM is diachronically penetrated. While the claim that the AAM is informationally encapsulated is incompatible with the interpretation of these empirical evidences, my revision is able to account for them. Whoever wants to maintain the claim that the AAM is informationally encapsulated should then re-describe alternatively this evidence. I concluded that my revision, and also data about the diachronic cognitive penetrability of the AAM fit well with the psychoevolutionary theory of emotion and that the possibility of diachronic penetrability is advantageous in providing an account of the irrationality involved in recalcitrant emotions.

## References

- Brady M. S., (2007), *Recalcitrant Emotions and Visual Illusions*, *American Philosophical Quarterly*, 44:3, pp. 273 - 284.
- Brady M. S., (2009), *The irrationality of recalcitrant emotions*, *Philosophical Studies*, 145, pp. 413-430.
- Cosmides L., Tooby J., (2000), *Evolutionary Psychology and the Emotions*, in Lewis, M., Haviland-Jones, J.M., *Handbook of Emotions*, (2<sup>nd</sup> ed.), Guilford, New York.
- Darwin C., (1872), *The Expressions of Emotions in Man & Animals* ( 1<sup>st</sup> ed.), Philosophical Library, New York.
- Ekman P., (1972), *Emotions in the Human Face*, Pergamon Press, New York.
- Ekman P., (1977), *Biological and cultural contributions to body and facial movement*, in John Blacking (ed.), *The Anthropology of the Body*, Academic Press, London.
- Fodor J. A., (1983), *The modularity of mind*, MA: MIT Press, Cambridge.
- Fodor J. A., (2000), *The mind doesn't work that way: the scope and the limits of computational psychology*, MIT Press, Cambridge.
- Gorenstein E. E. et al., (2007), *Cognitive-behavioral therapy for Reduction of Persistent Anger*, *Cognitive and Behavioral Practice*, 41, pp. 168-184.
- Griffiths P. E., (1990), *Modularity and the Psychoevolutionary Theory of Emotion*, *Biology and Philosophy*, 5:2 , pp. 175- 196.
- Griffiths P. E., (1997), *What Emotions Really Are: The Problem of Psychological Categories Science and Its Conceptual Foundations*, University of Chicago Press.
- Jokić-Begić N., (2010), *Cognitive-Behavioral Therapy and Neuroscience: Towards Closer Integration*, *Psychological Topics*, 19:2, pp. 235-254.
- Mattick R. P. et al., (1989), *Exposure and Cognitive Restructuring for Social Phobia: A Controlled study*, *Behavior Therapy* 20, pp. 3-23.
- McGurk H., MacDonald J.,(1976), *Hearing lips and seeing voices*, *Nature*, 264, pp. 746-748.
- Paquette V., et al., (2003), *“Change the mind and you change the brain”: effects of cognitive-behavioral therapy on the neural correlates of spider phobia*, *Neuroimage*, 18, pp. 401-409.

- Pylyshyn Z. W., (1980), *Computation and cognition: issues in the foundations of cognitive science, The behavioral and brain sciences*, 3, pp. 111-169.
- Pylyshyn Z. W., (1999), *Is vision continuous with cognition? The case for cognitive impenetrability of visual perception*, *Behavioral and Brain Sciences*, 22:3, pp. 341- 423.
- Prinz J. J., (2006 a), *Is the mind really modular?*, in R. Stainton, ed., *Contemporary Debates in Cognitive Science*, (pp. 22–36). Blackwell, Oxford.
- Prinz J. J., (2006 b), *Is Emotion a Form of Perception?*, *Canadian Journal of Philosophy*, 32 (Supplement), pp. 137- 160.
- Robbins P., (2010), *Modularity of Mind*, *The Stanford Encyclopedia of Philosophy*, Edward N. Zalta(ed.), URL = <<http://plato.stanford.edu/archives/sum2010/entries/modularity-mind/>>.
- Salkovskis P. M., et al., (2006), *Belief disconfirmation versus habituation approaches to situational exposure in panic disorder with agoraphobia: A pilot study*, *Behavior Research and Therapy* 45, pp. 877-885.
- Shea N., (2013), *Distinguishing Top-Down From Bottom-Up Effects*, in S. Biggs, M. Matthen, D. Stokes (eds.) *Perception and Its Modalities* (OUP).
- Southworth S., Kirsh I., (1988), *The role of expectancy in exposure-generated fear reduction in agoraphobia*, *Behavioral Research and Therapy*, 26: 2, pp. 112- 120.
- Hauner K. et al., (2012) *Exposure therapy triggers lasting reorganization of neural fear processing*, *Pnas*, 109:31, pp. 9203- 9208.
- Tappolet C., (2012), *Emotions, perceptions, and emotional illusions*, in C. Calabi (Ed.), *The crooked oar, the moon's size and the Kanizsa triangle: Essays on perceptual illusions*, Cambridge, MIT Press.
- Watson J. B., (1930), *Behaviorism*, W. W. Norton, New York.