Judgments about moral responsibility and determinism in patients with behavioural variant of frontotemporal dementia: Still compatibilists

Sacha Bourgeois-Gironde, Florian Cova, Maxime Bertoux, Bruno Dubois

Consciousness and Cognition 21 (2012) 851–864

Contents lists available at SciVerse ScienceDirect

Consciousness and Cognition

journal homepage: www.elsevier.com/locate/concog

Judgments about moral responsibility and determinism in patients with behavioural variant of frontotemporal dementia: Still compatibilists ☆

Florian Cova [a,b,*], Maxime Bertoux [c,d,e,f], Sacha Bourgeois-Gironde [b,g], Bruno Dubois [c,d,e,f]

[a] Swiss Centre for Affective Sciences, University of Geneva, Switzerland
[b] Institut Jean Nicod, Ecole des Hautes Etudes en Sciences Sociales, Ecole Normale Supérieure, France
[c] Université Pierre et Marie-Curie (UPMC), Paris 6, France
[d] Institut du Cerveau et de la Moelle Epinière, INSERM/CNRS UMRS 975, France
[e] Institut de la Mémoire et de la Maladie d'Alzheimer (IMMA), Hôpital Pitié-Salpêtrière, France
[f] Centre de Référence Démences Rares, Hôpital Pitié-Salpêtrière, France
[g] CEPERC, Aix-Marseille University, France

ARTICLE INFO

Article history:
Received 25 September 2011
Available online 10 March 2012

Keywords:
Experimental philosophy
Free will
Moral responsibility
Frontotemporal dementia
Emotions
Punishment

ABSTRACT

Do laypeople think that moral responsibility is compatible with determinism? Recently, philosophers and psychologists trying to answer this question have found contradictory results: while some experiments reveal people to have compatibilist intuitions, others suggest that people could in fact be incompatibilist. To account for this contradictory answers, Nichols and Knobe (2007) have advanced a 'performance error model' according to which people are genuine incompatibilist that are sometimes biased to give compatibilist answers by emotional reactions. To test for this hypothesis, we investigated intuitions about determinism and moral responsibility in patients suffering from behavioural frontotemporal dementia. Patients suffering from bvFTD have impoverished emotional reaction. Thus, the 'performance error model' should predict that bvFTD patients will give less compatibilist answers. However, we found that bvFTD patients give answers quite similar to subjects in control group and were mostly compatibilist. Thus, we conclude that the 'performance error model' should be abandoned in favour of other available model that best fit our data.

1. Introduction

Whether moral responsibility and determinism are compatible is a highly debated question among philosophers. 'Compatibilist' philosophers argue that we can be responsible for our actions in a deterministic world (and thus that moral responsibility and determinism are *compatible*) while 'incompatibilist' philosophers claim that determinism is by nature a threat to moral responsibility (and thus that moral responsibility and determinism are *incompatible*). Both sides have used many arguments, most of which ultimately rely on appeal to folk intuitions (i.e. to laypeople's untutored spontaneous judgments about principles or particular cases). This concern about which position is the more intuitive and folk intuitions about

☆ The two first authors contributed equally. This research was supported in part by a Grant from the French Agence Nationale de la Recherche (ANR) (ANR Blanche: SoCoDev). Previous versions of this paper were presented at the Second Workshop of the Experimental Philosophy Group UK in Sheffield and at the European Workshop in Experimental Philosophy in Eindhoven. We thank the organizers and the audience for their helpful comments. We also thank Eddy Nahmias for his comments.

* Corresponding author at: Institut Jean Nicod, Ecole des Hautes Etudes en Sciences Sociales, Ecole Normale Supérieure, France.
E-mail addresses: florian.cova@gmail.com (F. Cova), maximel.bertoux@gmail.com (M. Bertoux), sacha.bourgeois-gironde@univ-provence.fr (S. Bourgeois-Gironde), bruno.dubois@psl.aphp.fr (B. Dubois).

1053-8100/$ - see front matter © 2012 Elsevier Inc. All rights reserved.
doi:10.1016/j.concog.2012.02.004

free will ultimately led 'experimental philosophers' (Cova, 2011; Knobe & Nichols, 2008) to empirically investigate the nature of folk intuitions (Nahmias, Morris, Nadelhoffer, & Turner, 2005; Nahmias, Morris, Nadelhoffer, & Turner, 2006).

However, determining whether laypeople are compatibilists or incompatibilists turned out to be more complicated than expected: it was found that people's answers were likely to vary from expressing compatibilism to expressing incompatibilism depending on how the question was asked and the nature of the case. To account for these conflicting judgments, Nichols and Knobe (2007; see also Nichols 2006) have proposed a 'performance error model', according to which compatibilist judgments are the product of an emotional bias.

In this paper, we put this 'performance error model' to test by studying judgments about moral responsibility and determinism made by patients suffering from a behavioural variant of frontotemporal dementia (bvFTD). People with bvFTD are known to suffer from emotional deficits. So, we argue, the 'performance error model' should predict decreased compatibilist judgments in people with bvFTD. Nevertheless, we present data suggesting that people with bvFTD are no less likely to give compatibilist answers than control participants. We conclude that Nichols and Knobe's 'performance error model' should be abandoned for other models that are more consistent with our results.

2. Folk intuitions about moral responsibility and determinism: three models

2.1. The relevance of folk intuitions to the compatibilism/incompatibilism debate

Imagine a world in which everything that happens is entirely caused by what has happened before according to immutable laws of nature – that is, a world in which everything that happens could be fully explained by the state of this world at an antecedent time and predicted from the knowledge of this antecedent state and of the laws of nature. Such a world is what philosophers call a *deterministic* world.[1] Now, can people be morally responsible for their actions if they live in such a world? 'Compatibilist' philosophers answer 'yes' while 'incompatibilist' philosophers say 'no'. Both opposition positions entail very different account of the nature of freedom and moral responsibility.

Within the philosophical debate about moral responsibility and determinism, appeal to folk intuitions plays an important role (Nahmias et al., 2006). Thus, it is often considered that a theory of moral responsibility that is consistent with folk intuitions has a dialectical advantage, while the burden of proof falls on the shoulders of those who suggest that folk intuitions might be widely mistaken.

This is why both sides typically claim that common sense is on their side, and this is how the debate over the compatibility of moral responsibility and determinism has extended in a debate over the coherence of folk intuitions about moral responsibility with either compatibilism or incompatibilism.[2] For example, on the compatibilist side, Frankfurt (1969) has produced famous cases (since dubbed 'Frankfurt cases') in which we are supposed to have the intuition that an agent is morally responsible in spite of the fact that he had to act the way he did.[3] On the incompatibilist side, 'Manipulation arguments' (see Pereboom, 1995 for an example) have been trying to show that we have both the intuition that manipulated agents have no moral responsibility and that there is no significant difference between manipulated agents and agents living in a deterministic world.[4] And finally, caught in the crossfire, 'revisionist' philosophers consider that our natural conception of moral responsibility is incoherent and needs in fact to be revised (see for example Vargas, 2005). Of course, the plausibility of such a view directly also depends on whether the revisionist's account of the nature of folk intuitions about moral responsibility is correct (Vargas, 2006), so that all three positions are in a way interested in the real nature of folk intuitions about free will and moral responsibility.

2.2. Contradictory results in the experimental philosophy of free will

To settle this dispute, Nahmias and his colleagues (Nahmias et al., 2005; Nahmias et al., 2006) decided to empirically investigate folk intuitions about moral responsibility. They gave participants short vignettes describing agents living in deterministic universe, and asked whether this agent deserved blame for what he has done. Here is an example:

> *Supercomputer case:*
> Imagine that in the next century we discover all the laws of nature, and we build a supercomputer which can deduce from these laws of nature and from the current state of everything in the world exactly what will be happening in the world at any future time. It can look at everything about the way the world is and predict everything about how it will be with 100% accuracy. Suppose that such a supercomputer existed, and it looks at the state of the universe at a certain time on March 25, 2150 AD, 20 years before Jeremy Hall is born. The computer then deduces from this information and the

[1] In this paper, we use 'determinism' in a very specific and laplacian sense according to which determinism implies predictability *de jure*. Philosophers also use 'determinism' in a broader sense that does not necessarily imply such predictability. See for example Van Inwagen (1983).

[2] For a study suggesting that laypeople have intuitions that differ from those of professional philosophers and that folk intuitions cannot be read off philosophers' intuitions, see Schulz, Cokely, and Feltz (2011).

[3] For an empirical investigation of 'Frankfurt cases', see Woolfolk, Doris, and Darley (2006) and Miller and Feltz (2011).

[4] For an empirical investigation of folk intuitions about the Manipulation argument, see Sripada (forthcoming).

laws of nature that Jeremy will definitely rob Fidelity Bank at 6:00 pm on January 26, 2195. As always, the supercomputer's prediction is correct; Jeremy robs Fidelity Bank at 6:00 pm on January 26, 2195.
Imagine such a supercomputer actually did exist and actually could predict the future, including Jeremy's robbing the bank (and assume Jeremy does not know about the prediction). Do you think that when Jeremy robs the bank, he's morally blameworthy for it?

In this particular case, 83% answered that Jeremy was morally blameworthy for having robbed the bank.[5] This means that, in this case, most participants considered that Jeremy could be morally responsible for his actions, while living in a deterministic world. This strongly suggests that people have compatibilist intuitions and similar results were obtained by Nahmias and his colleagues for two other kinds of vignettes.

Nevertheless, things are far from being that simple. In a study by Nichols and Knobe (2007), subjects were presented with the following description of two different universes:

Imagine a universe (Universe A) in which everything that happens is completely caused by whatever happened before it. This is true from the very beginning of the universe, so what happened in the beginning of the universe caused what happened next, and so on right up until the present. For example one day John decided to have French Fries at lunch. Like everything else, this decision was completely caused by what happened before it. So, if everything in this universe was exactly the same up until John made his decision, then it had to happen that John would decide to have French Fries.
Now imagine a universe (Universe B) in which almost everything that happens is completely caused by whatever happened before it. The one exception is human decision making. For example, one day Mary decided to have French Fries at lunch. Since a person's decision in this universe is not completely caused by what happened before it, even if everything in the universe was exactly the same up until Mary made her decision, it did not have to happen that Mary would decide to have French Fries. She could have decided to have something different.
The key difference, then, is that in Universe A every decision is completely caused by what happened before the decision – given the past, each decision has to happen the way that it does. By contrast, in Universe B, decisions are not completely caused by the past, and each human decision does not have to happen the way that it does.

Participants were then asked which one of these two universes was more like ours. Nearly all participants (90%) answered 'Universe B'. Then, participants in the *concrete condition* received the following scenario:

Concrete condition:
In Universe A, a man named Bill has become attracted to his secretary, and he decides that the only way to be with her is to kill his wife and 3 children. He knows that it is impossible to escape from his house in the event of a fire. Before he leaves on a business trip, he sets up a device in his basement that burns down the house and kills his family.
Is Billy fully morally responsible for killing his wife and children?

In this condition, most subjects (72%) gave the compatibilist answer according to which the agent was fully morally responsible. These results are consistent with those obtained by Nahmias and his colleagues. But, let's consider now the *abstract condition*. Participants in this condition had no scenario to read (just the description if Universe A and Universe B) but just received the following question:

Abstract condition:
In Universe A, is it possible for a person to be fully morally responsible for their actions?

In this condition, most subjects (86%) gave the incompatibilist answer, according to which it is impossible for a person living in Universe A to be fully morally responsible. Nichols and Knobe conclude that this shows that it is too simple to claim that people are either compatibilists or incompatibilists: participants' answers can vary depending on how the question is framed.

2.3. Nichols and Knobe's 'performance error model'

How are we to account for these conflicting judgments? While some have argued that abstract and concrete scenarios elicit different psychological mechanisms (Sinnott-Armstrong, 2008), Nichols and Knobe have focused on the affective contrast between the abstract question and the concrete question. Indeed, the concrete question describes a gruesome and revolting crime, while the phrasing of the abstract question is clearly dispassionate. So, it might be that compatibilist answers are emotionally driven and that people are more compatibilist in the concrete case because the situation described (a murder) is emotionally loaded. According to Nichols and Knobe's 'performance error model', compatibilist answers are

[5] Other participants were asked whether Jeremy robbed the bank *of his own free will*. 76% of participants answered that he did. Though questions about free will and moral responsibility are closely related, some philosophers have proposed that the two questions should be distinguished, and that we can be morally responsible without having free will (see Fischer, 2002; for empirical researches on the folk concept of 'free will', see Monroe & Malle, 2010).). In this paper, we leave out the question of free will and focus on moral responsibility. One reason for this is that our main experiment was run in French and that there is no direct equivalent of 'free will' in French (the direct translation is '*libre-arbitre*', but the term is more theologically connoted and much less common in French than the English 'free will').

Table 1
Proportion of compatibilist answers for each condition in Nichols and Knobe's experiment.

	Agent in indeterminist universe (Universe B) (%)	Agent in determinist universe (Universe A) (%)
High affect condition	95	64
Low affect condition	89	23

emotional biases that prevent us to correctly apply our incompatibilist conception of moral responsibility (Nichols & Knobe, 2007).

To support this hypothesis, Nichols and Knobe designed two new conditions. The *low affect condition* was the following:

Low affect condition:

As he has done many times in the past, Mark arranges to cheat on his taxes. Is it possible that Mark is fully morally responsible for cheating on his taxes?

While the *high affect condition* was the following:

High affect condition:

As he has done many times in the past, Bill stalks and rapes a stranger. Is it possible that Bill is fully morally responsible for raping the stranger?

In each condition, for half of the subjects, the question stipulated that the agent was in Universe A while, for the other half, the question stipulated that the agent was in Universe B. Table 1 describes, for each combination, the proportion of participants who answered 'yes'

Participants seemed to have very few compatibilist intuitions in the low affect condition situated in a determinist universe. For Nichols and Knobe, these experiments support the 'performance error model' according to which people have an underlying incompatibilist theory that can be overcome by emotional compatibilist bias.[6]

To sum up: according to Nichols and Knobe, people are naturally incompatibilist. That's why they give incompatibilist answer in the abstract condition and the low-affect condition (the 'low-affect cases'). Nevertheless, emotionally salient moral violations can drive people to be biased to give compatibilist answers, and that is what happens in the concrete condition, the high-affect condition (the 'high-affect cases') and the Supercomputer scenario (see Fig. 1).

2.4. Nahmias and Murray's 'error theory'

Nevertheless, Nahmias and Murray (2010) have recently proposed an 'error theory' according to which most incompatibilist answers are in fact the product of confusion. According to Nahmias and Murray, most people who give incompatibilist answers misunderstand determinism by confounding it with epiphenomenalism or fatalism, i.e. doctrines according to which agents and their mental states do not have any role to play in the generation of their actions (as in cases of manipulations in which the agent is directed by an external force) – an idea they call 'bypassing'.[7] Thus, the reason why so many people are incompatibilist when confronted with Knobe and Nichols' scenarios is that these scenarios are written in such a way that they lead most people to misunderstand determinism as implying 'bypassing', possibly because of the words 'it had to happen'.

To test this hypothesis, Nahmias and Murray (2010) gave to participants a concrete version and an abstract version of Nichols and Knobe's 'Universe A' scenario[8] as well as an abstract and a concrete version of a scenario in which the same universe keeps being re-created. Participants were not only asked if agents in these scenarios deserved praise or blame and acted from their own free will, but they were also asked three questions designed to probe their understanding of what determinism is. Here are sample questions for the concrete version in Universe A (involving Bill killing his wife and children). Participants had to say on a scale how much they agreed or disagreed with the following statements:

- Decisions: Bill's decision to kill his wife and children has no effect on what he ends up being caused to do.
- Wants: What Bill wants has no effect on what he ends up being caused to do.
- Believes: What Bill believes has no effect on what [he/she] ends up being caused to do.
- No control: Bill has no control over what he does.

[6] Note also that the low-affect condition and the high-affect condition are both concrete conditions, which suggests that the difference between the abstract and the concrete condition cannot directly be explained in terms of "abstract versus concrete".

[7] To quote Nahmias and Murray (2010): "What is "bypassing"? The basic idea is that one's actions are caused by forces that bypass one's conscious self, or at least what one identifies as one's "self". More specifically, it is the thesis that one's actions are produced in a way that bypasses the abilities compatibilists typically identify with free will, such as rational deliberation, conscious consideration of beliefs and desires, formation of higher-order volitions, planning, and the like. As such, bypassing might take the form of *epiphenomenalism* about the relevant mental states (i.e., that deliberations, beliefs, and desires are causally irrelevant to action), or it might take the form of *fatalism*—the belief that certain things will happen no matter what one decides or tries to do, or that one's actions *have to* happen *even if* the past had been different. Bypassing suggests that conscious agents have no control over their actions because they play no role in the causal chain that leads to their actions."

[8] That is: the abstract condition and the concrete condition involving Bill killing his wife and children.

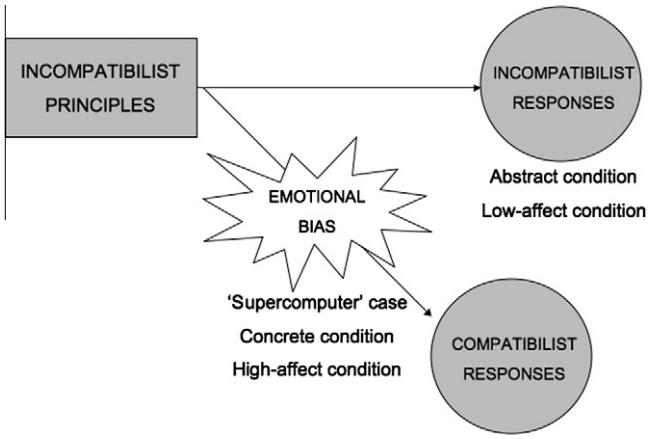

Fig. 1. Nichols and Knobe's 'performance error model'.

For each participant, these questions allowed to determine how much he considered determinism to entail 'bypassing' (for the four first questions, the 'bypassing' answer is agreement; for the last question, it's disagreement).

Nahmias and Murray's results showed that compatibilist intuitions were highly correlated to a good understanding of determinism (that is: an understanding of determinism that do not conflate it with bypassing) and that people who gave incompatibilist answers were far more susceptible to believe that determinism entailed 'bypassing' (that is: the fact that our mental states do not play any role in generating our actions). Furthermore, they found that Knobe and Nichols' description of 'Universe A' scenario (in comparison to the 're-creating universe' scenario) led a great number of participants to think that mental states were 'bypassed' in this universe. Finally, they observed that people in the abstract condition were more prone to believe in 'bypassing' than people in the concrete condition. They conclude that most incompatibilist answers do not reveal a true commitment to incompatibilism, because most of them were the product of a bad understanding of determinism.

To sum up: for Nahmias and Murray, people are naturally compatibilists. That's why most of them give compatibilist answers to the Supercomputer scenario. Nevertheless, some subjects still give incompatibilist answers, because they are mistaken in thinking that determinism entail 'bypassing'. Nichols and Knobe's scenario is badly phrased, and drives more subjects to make this mistake; hence the greater number of incompatibilist answers in the abstract condition and low-affect condition. Nevertheless, in the concrete condition and the high-affect conditions, subjects seem to have a better understanding of determinism and making less mistakes: this is why they give more compatibilist answers in these conditions (see Fig. 2).

But, one might ask, why do people seem to have a better understanding of determinism in the concrete and high-affect conditions? For Nahmias and Murray, the answer is, once more, affect:

> ''it may be that the high negative affect causes participants to neglect the *bypassing* feature of the scenario. In other words, [Nichols and Knobe]'s description of determinism may lead people to make a mistake, which is then ''cancelled out'' in the concrete case—but not in the abstract case—by high negative affect. Hence, we predict that most people will *not* read [Nichols and Knobe's] *concrete* scenario to involve bypassing, which may help to explain why they are generally willing to attribute [moral responsibility] to Bill [for having killed his wife and children]''.

Thus, in Nahmias and Murray's account, affect still acts as a bias, but only as a 'counter-bias': affect can bias people to ignore other biasing features, and this is why people are more prone to error in low-affect conditions and less in concrete high-affect conditions. An error cancels an error.[9]

[9] Note that Nahmias and Murray have also an alternative and compatible account for the difference between the abstract and the concrete condition that doesn't rely on affect. According to them, the concrete condition allows participants to have a better understanding of determinism, in particular by making clear that the agent acts according to their desires: ''Specifically, we believe that judgments about responsibility—including whether agents deserve credit or blame for their actions—will be more reliable if they engage our capacities to think about the beliefs, desires, and intentions of agents (e.g., our ''theory of mind'' capacities), and these are presumably more likely to be engaged when we consider specific agents in specific circumstances. More generally, it may be that people's intuitions are more reliable when they have more details about a scenario, which is likely part of the reason why philosophers construct thought experiments with specific details to probe (or prime) our intuitions. Hence, while we agree with N&K that concrete cases that *also* involve high affect may lead to errors, we do not believe this is a product of concreteness *per se*. Rather, we believe that, in general, concrete cases are more likely to reveal reliable intuitions about [moral responsibility] and [free will] than are abstract cases.'' Nevertheless, this hypothesis is not enough to account for the difference between the low-conflict and the high-affect cases, and that is why Nahmias and Murray still need to postulate the influence of affect as an auxiliary hypothesis.

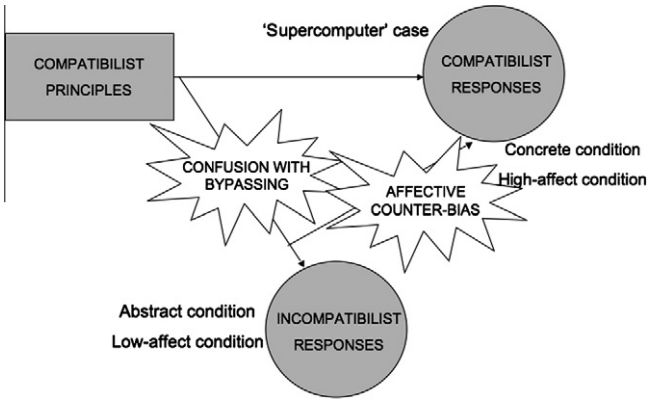

Fig. 2. Nahmias and Murray's 'error theory'.

Nevertheless, this auxiliary hypothesis is not fundamental for Nahmias and Murray's 'error theory'. One might endorse the general framework of their 'error theory', grant that incompatibilist answers are due to misunderstanding, and that subjects are less prone to such mistakes in concrete and high-affect conditions, and still adopt another auxiliary hypothesis concerning why people are better in such cases. Thus, we have the choice between three different kinds of models:

1. Nichols and Knobe's *performance error model (N&K)*, according to which, in absence of affect, people give incompatibilist answers,
2. Nahmias and Murray's *error theory* supplemented by the *affective counter-bias hypothesis (N&Mv1)*, according to which people in ideal conditions are compatibilist (as in the Supercomputer scenario), but can give incompatibilist answers when they take determinism as implying "bypassing", a tendency than can itself be countered by affect,
3. Nahmias and Murray's *error theory* without the hypothesis of a role played by affect *(N&Mv2)*, that is the theory according to which people in ideal conditions are compatibilist, but can give incompatibilist answers when they take determinism as implying "bypassing", a tendency that is diminished in concrete and high-affect cases, for a reason unknown but not linked to affect.

3. How to decide between these accounts: different predictions for patients with emotional deficits

3.1. How to decide between these accounts: a proposal

Now that we have three accounts, how are we to decide between them? We must find a situation in which these three kinds of account make different predictions and test it. One such situation, we suggest, is the putative answers given by patients presenting emotional deficits.

Studies on patients with emotional deficits have been used several times in moral psychology to evaluate hypotheses postulating that a certain kind of answer is due to emotional reactions. For example, Knobe (2003) found that people were likely to judge that an agent intentionally brough about a foreseen side effect, but only when this side effect was morally bad.[10] Nadelhoffer (2006) proposed that attributions of intentionality for bad side effects were the product of a bias, and more precisely of an emotional reaction to the agent's blameworthiness. To test for Nadelhoffer's hypothesis, Young and his colleagues (2006) gave Knobe's scenarios to patients suffering from damages to the ventromedial prefrontal cortex. Indeed, if the pattern of responses was due to emotional reactions, patients with this kind of lesions, because they show emotional deficits, should be less likely to display it. Nevertheless, Young and his colleagues found that the pattern of responses for patients with damages to the ventromedial prefrontal cortex was identical to the one of control subjects. They concluded that the pattern of responses observed by Knobe was not the product of emotional reactions.

Similarly, on the basis of fMRI studies on the resolution of moral dilemmas, Greene et al. (2001); see also Greene, 2008) suggested that utilitarian responses (sacrifice one to save many) were the product of cognitive processes while deontological responses (refuse to sacrifice one, even to save many) were the product of emotional reactions. Thus, this hypothesis led to the prediction that people with emotional deficits and impoverished emotional responses should give more utilitarian responses to moral dilemmas. This prediction was confirmed by researches on patients suffering from a behavioural variant of frontotemporal dementia (Mendez, Anderson, & Shapira, 2005) or from lesions to the ventromedial prefrontal cortex (Ciaramelli, Muccioli, Làvadas, & Di Pellegrino, 2007; Koenigs et al., 2007), thus lending support to Greene's hypothesis (Greene, 2007).

[10] Evidence suggest that this effect, that has been dubbed the 'Knobe effect' or the 'side-effect effect', might not be limited to side effects but could also be observed for means of action (see Cova & Naar, forthcoming).

All these studies rely on the hypothesis that if a certain type of answer is the product of emotional reactions, we might expect patients with emotional deficits to be less likely to give this particular kind of answers. Thus, following this hypothesis, if we take two scenarios (here, the Supercomputer and the concrete condition in which Bill kills his wife and children), we can see our three models making very different predictions for people with emotional deficits:

1. According to *N&K*, people are by default incompatibilists, unless they are biased by emotional reactions. Patients with emotional deficits are less likely to be biased by such reactions, so *N&K* should predict that they would give less compatibilist answers to both the Supercomputer and the concrete condition.
2. According to *N&Mv1*, people are by default compatibilists. Nevertheless, when confronted to Nichols and Knobe's description of determinism, they misunderstand determinism, which leads them to give incompatibilist answers, except for the high-affect case, in which emotional reactions lead them to give the compatibilist answer. Given this account, patients with emotional deficits should be as likely as control subjects to give compatibilist answer to the Supercomputer case, but more likely to give incompatibilist answers in the concrete condition.
3. According to *N&Mv2*, people are by default compatibilists. Nevertheless, when confronted to Nichols and Knobe's description of determinism, they misunderstand determinism, which leads them to give incompatibilist answers, except for the concrete condition, for an unknown reason unrelated to emotion reactions. Thus, according to this account, patients with emotional deficits should not differ in their answer from control subjects, as emotions do not play any role.

Each model predicts a different pattern of responses for patients with emotional deficits. So, all we have to do to adjudicate between these three models is to compare these predictions with answers actually given by patients with emotional deficits.

3.2. A case of emotional deficit: behavioural variant of frontotemporal dementia

Searching for such patients, we found a good example of emotional deficit in the behavioural variant of frontotemporal dementia (bvFTD). bvFTD is a subtype of frontotemporal lobe degeneration, a group of clinical syndromes associated with focal atrophy of frontal and anterior temporal lobes, which also include semantic dementia and progressive nonfluent aphasia (Neary et al., 1998).

bvFTD is characterised by an early alteration in behaviour and personality with emotional blunting, social inappropriateness and loss of insight (Piguet, Hornberger, Mioshi, & Hodges, 2011). Patients suffer from impairment in social cognition and emotional processing, in particular during theory of mind or emotional identification tests (Lavenu & Pasquier, 2005; Lough et al., 2006; Torralva et al., 2007).

These deficits in social cognition and emotional processing appear early in the disease and are specific compared to other neurodegenerative diseases such as Alzheimer Disease (Funkiewiez, Bertoux, de Souza, Lévy, & Dubois, 2012). They are a consequence of the atrophy in orbitofrontal (OFC) and medial prefrontal (mPFC) cortices, two critical cerebral regions involved in the social and emotional cognition that are early and specifically impaired in bvFTD (Boccardi et al., 2005; Broe et al., 2003; Perry et al., 2006; Seeley et al., 2008; Tranfaglia, Palumbo, Siepi, Sinzinger, & Parnetti, 2009).

The role of OFC and mPFC in emotional processing has been widely established through imaging or lesion studies (Hornak et al., 2003; Hornak et al., 2004; Ruby & Decety, 2004). OFC analyses the value of a reward or the affective significance and valence of a stimulus such as a facial or vocal emotions (Rolls, Hornak, Wade, & Mcgrath, 1994), and mPFC is considered as a supramodal emotions representation area (Peelen, Atkinson, & Vuilleumier, 2010). Recent imaging findings in bvFTD made the direct kink between atrophy of OFC and behavioural disturbances (Hornberger, Geng, & Hodges, 2011) or atrophy and blood perfusion decrease of mPFC and emotional processing impairment (Bertoux et al., in preparation; Bertoux et al., submitted for publication).

Thus, bvFTD could be considered as a model of OFC and mPFC dysfunction and emotional impairment and is a very relevant choice to assess emotional implication in moral judgment. By contrast, we decided to recruit mild Alzheimer Disease (AD) patients as a control disease since neither these affective and social deficits were showed in its mild forms (Funkiewiez et al., 2011), nor of the cerebral area from the "emotional circuitry" are impaired in this disease (Rabinovici et al., 2007; Tranfaglia et al., 2009).

4. Controlling for the material

4.1. First control experiment: simplifying Nichols and Knobe' material

However, before actually running the experiment, we faced two worries about the use of Nichols and Knobe's original material: their scenarios, describing and comparing two universes, were too long and complicated for people potentially suffering from cognitive impairments. Moreover, our subjects being French,[11] we had to make sure that our translation and shortening of their material did not alter the effect they discovered.

[11] All experiments presented in this paper have been run on French participants using French material.

Table 2
Aggregated scores for the first control experiment.

	High-affect case	Low-affect case
Read first	6.2	3.8
Read second	5.1	5.3

Our shortened version of the concrete condition (our 'high-affect case') went like this:

High-Affect Case:
Imagine a universe in which everything that happens is completely caused by whatever happened before it. This is true from the very beginning of the universe, so what happened in the beginning of the universe caused what happened next, and so on right up until the present. For example one day John decided to have French Fries at lunch. Like everything else, this decision was completely caused by what happened before it. So, if everything in this universe was exactly the same up until John made his decision, then it had to happen that John would decide to have French Fries.
Thus, in this universe, every decision is completely causedby what happened before the decision – given the past, each decision has to happen the way that it does.
In Universe A, a man named Bill has become attracted to his secretary, and he decides that the only way to be with her is to kill his wife and 3 children. He knows that it is impossible to escape from his house in the event of a fire. Before he leaves on a business trip, he sets up a device in his basement that burns down the house and kills his family.

Our shortened 'low-affect case' was similar, except for the last paragraph, which we replaced with the text of Nichols and Knobe's low-affect condition (Mark cheating on his taxes).

To test for this new material, we recruited 20 participants at the Laboratoire de Sciences Cognitives et Psycholinguistique in Paris,[12] and gave them each of them the two scenarios. 10 received the 'high-affect case' first and 10 received the 'low-affect case' first. After reading each scenario, subjects had to answer the three following questions (always in the same order) on scale ranging from 1 to 7 ('1' being indicated as 'NO' and 7 being indicated as 'YES'):

1. *The 'responsibility' question*: Is Bill morally responsible for the death of his wife and children? [Is Mark morally responsible for having cheated on his taxes?]
2. *The 'blameworthiness' question*: Does Bill deserve blame for the death of his wife and children? [Does Mark deserve blame for having cheated on his taxes?]
3. *The 'punishment' question*: Does Bill deserve a punishment for the death of his wife and children? [Does Mark deserve a punishment for having cheated on his taxes?]

Because answers to the three questions turned out to be highly correlated, we averaged them to obtain an aggregated score and used this aggregated score for our analyses (see Table 2).

A two-factor ANOVA with *case* ('high-affect' or 'low-affect') and *order* ('high-affect first' or 'low-affect first') as factors revealed a significant effect of case ($F(1,18) = 10.8$, $p < 0.01$) and a marginally significant effect of order ($F(1,18) = 0.4$, $p < 0.10$).

In a first time, we only analysed scores for the first scenario received by participants (in order to exclude order effects). As in Nichols and Knobe's original experiments, we found that people obtained higher score for the 'high-affect case' ($M = 6.2$) than for the 'low-affect' case ($M = 3.8$). A Welch t-test proved this difference to be statistically significant ($N = 20$, $t = 3.2$, $df = 17.6$, $p < 0.01$). In a second time, we analysed all scores. Once again, we found that people obtained higher score for the 'high-affect case' ($M = 5.6$) than for the 'low-affect' case ($M = 4.6$). A paired t-test proved this difference to be statistically significant ($N = 20$, $t = 3.3$, $df = 19.0$, $p < 0.01$). Thus, we were able to reproduce Nichols and Knobe's original effect using simpler vignettes.

Though previous researches revealed no order effects for Nichols and Knobe's material (Feltz, Cokely, & Nadelhoffer, 2009), we observed that, overall, people who read the 'high-affect case' first were more likely to give compatibilist answer than those who read the 'low-affect case' first ($M = 5.8$ vs. $M = 4.4$), a tendency revealed marginally significant by our ANOVA. As shown by Table 1 this tendency can be decomposed in a tendency for subjects reading the 'high-affect case' first to give more compatibilist answers to the 'low-affect case' and a tendency for subjects reading the 'low-affect case' first to give less compatibilist answers to the 'high-affect case'. Nevertheless, probably due to our small sample size, only the first tendency reached marginal significance (Welch t-test: $N = 20$, $t = -1.9$, $df = 18.0$, $p = 0.07$).

[12] The age mean was 22.8. 15 participants were women. We intentionally used small groups to check whether the effect could be observed within a small number of participants, given that our sample of patients was likely to be small.
[13] The age mean was 24.7. 13 participants were women.

Table 3
Aggregated scores for the second control experiment.

	High-affect case	Supercomputer
Read first	6.5	6.7
Read second	6.2	5.7

4.2. Second control experiment: checking for order effects

After having verified that our simplified version of the 'high-affect case' generated the same effect than the original case, we had to check for the presence of order effects when this case is given with the Supercomputer case. To test for this, we recruited 20 participants at the Laboratoire de Sciences Cognitives et Psycholinguistique in Paris,[13] and gave them each of them the two scenarios. 10 received the 'high-affect case' first and 10 received the Supercomputer case first. After reading each scenario, subjects had to answer the three same questions than in our first control experiment (the 'responsibility', the 'blameworthiness' and the 'punishment' question).

As for our first control experiment, we averaged responses to the three questions in an aggregated score. We then run an ANOVA using scores as dependent variable and *case* ('Supercomputer' or 'high-affect case') and *order* ('Supercomputer first' or 'high-affect case first) as factors. We found no effect of case ($F(1,35) = 0.1$, $p = 0.74$) or order ($F(1,35) = 0.8$, $p = 0.38$), but found a significant interaction effect between these two factors ($F(1,35) = 4.5$, $p < 0.05$).

As suggested by the results presented in Table 3, this interaction effect might be due to the fact that ratings for the 'Supercomputer case' seems to be more influenced by the order of presentation than ratings for the 'high-affect case'. This might be due to the fact that people reading the 'high-affect case' first have already rated a horrible crime when they read the 'Supercomputer case' and lower their ratings because robbing a bank is less wrong than killing his whole family. On the contrary, people reading the 'Supercomputer' first gives high ratings to this scenario, and must logically continue to give high ratings to the 'high-affect case'. However, the influence of order on ratings for the Supercomputer case only reached marginal significance (Welch t-test: $N = 20$, $t = 1.9$, $df = 9.6$, $p = 0.09$).

Thus, as long as we use small samples in our experiment, there do not seem to be reasons to be worried about the potential interference of order effects.

4.3. Gender effects

Finally, note that an ANOVA with aggregated ratings as dependent variable and gender as a factor revealed a significant gender effect in our first control experiment ($F(1,18) = 7.3$, $p < 0.05$). Overall, women gave more compatibilist answers than men ($M = 5.6$ vs $M = 3.5$), a tendency already observed in the literature (Buckwalter & Stich, forthcoming; Holtzman, in preparation). Nevertheless, no such gender effect was found in our second control experiment (maybe due to a ceiling effect). Post-hoc analyses revealed that, in the first control experiment, difference between genders was only significant for the 'low-affect case' (Welch t-test: $N = 20$, $t = 3.1$, $df = 6.9$, $p < 0.05$)[14] and not for the 'high-affect case', (Welch t-test: $N = 20$, $t = 1.3$, $df = 4.5$, $p = 0.26$), which might explain why we did not observe it in the second experiment, since we did not use the 'low-affect case'.

Each of the accounts we have presented suggest a different account of this difference. Nichols and Knobe's 'performance error model' might explain the gender effect by suggesting that women are more sensible to the affective bias, whether because their affective reaction was higher or they were more likely to be influenced by their affective reaction. The affect-free version of Nahmias and Murray's 'error theory' could propose that women are less likely than men to confound determinism with bypassing. Finally, the original version of Nahmias and Murray's 'error theory', that still gives a role to affect, might choose between these two accounts, or even endorse a mix of both. Anyway, it seems that, as long as do not use the 'low-affect case', gender effects should not be an issue.

5. Frontotemporal dementia and judgments of moral responsibility: Running the experiment

5.1. Participants

Participants were ten patients (5 women and 5 men, 78.0 ± 8,9 years old, range 59–86; disease duration = 3.61 years ± 2.4) who met the criteria for AD (Dubois et al., 2007), twelve patients (5 women and 7 men, 66.5 ± 10.2 years old, range 48–82; disease duration = 3.0 years ± 1.7) who met the criteria for bvFTD (Neary et al., 1998) and ten control subjects (5 women and 5 men, 66.0 ± 7.1 years old, range 55–73).

All patients were evaluated by neurologists with clinical experience in neurodegenerative diseases, and were given a complete neurological and behavioural assessment. This examination confirmed a history of initial progressive decline in social interpersonal conduct and behaviour in bvFTD, with emotional blunting and loss of insight. In AD, it confirmed a history of episodic memory impairment with temporal and spatial disorientation.

[14] *M(women)*=5.2 vs. *M(men)*=2.7.

Table 4
Characteristics of bvFTD and AD patients and neuropsychological data.

	Control subjects	AD	bvFTD
Age	66.0 (7.1)	78.0 (8.9)	66.5 (10.2)
Gender	5 W/5 M	5 W/5 M	5 W/7 M
Educational level	5.2 (1.3)	5.2 (3.0)	5.7 (1.6)
Duration of disease		3.6 (2.4)	3.0 (1.7)
Neuropsychological Assessment			
MMSE (/30)	28.9 (1)	22.7 (3.2)[a]	25.2 (2.7)[a]
FAB (/18)	17.3 (0.9)	14.9 (1.5)[a]	14.3 (1.9)[a]
Facial emotional recognition (%)			46.8 (11.7)[a]
Faux-Pas test			
Faux-Pas detection and explanation (/40)			17.6 (4.5)[a]
Control questions (/20)			18.2 (1.8)[a]

Mean (standard deviation).
[a] Represent a pathologic scores according to normative data on healthy subjects.

All patients underwent a rapid neuropsychological examination that included the Mini Mental State Exam (Folstein, Folstein, & McHugh, 1975) and the Frontal Assessment Battery (Dubois, Slachevsky, Litvan, & Pillon, 2000). In bvFTD patients, we also evaluate theory of mind abilities with the Faux-Pas test (Stone, Baron-Cohen, & Knight, 1998) and facial emotional recognition with 35 faces from Ekman pictures (Ekman and Friesen, 1975) except for two bvFTD patients who were not assessed with these two last tests. Clinical and neuropsychological data are presented on Table 3. Structural MRI and SPECT were performed for all patients and revealed fronto-temporal atrophy and/or hypoperfusion in bvFTD and median-temporal/parietal atrophy in AD. To improve diagnostic accuracy, all patients were clinically followed for at least 18 months.

Patients were not included if they presented any of the following: (1) language complaints (progressive non-fluent aphasia or semantic dementia); (2) systemic illnesses that could interfere with cognitive functioning; (3) vascular lesions on MRI or neurological history suggestive of vascular dementia; (4) motor-neuron disease; (5) major depression; (6) use of anxiolytics or antipsychotics drugs.

This study was conducted at the Institute of Memory and Alzheimer Disease, and in Neurology department (Pitié-Salpêtrière Hospital). All clinical and neuroimaging data were generated during a routine clinical work-up and were extracted for the purpose of this study. Therefore, according to French legislation, explicit informed consent was waived. However, regulations concerning electronic filing were followed, and patients and their relatives were informed that individual data might be used in retrospective clinical research studies.

5.2. Procedure

Each participant received the 'high-affect case' first and the 'Supercomputer case' in second. After each scenario, subjects had to answer the three same questions than in our control experiments (the 'responsibility', the 'blameworthiness' and the 'punishment' question).

bvFTD patients also underwent an emotion recognition test and a theory of mind evaluation. In the emotion recognition test, thirty-five faces from Ekman pictures (Ekman et al., 1975) were presented, and the patient should indicate which emotion was expressed, among a list (presented in the top of the screen). Faces expressed seven different emotions (fear, sadness, disgust, surprise, anger, happiness and neutral). A general recognition percentage was calculated. To assess theory of mind, we used a short (ten stories) version of the Faux-pas test (Stone et al., 1998) in which patient have to detect and explain social inconveniences.

5.3. Results

5.3.1. Neuropsychological examination

First, our groups differed from age ($F(2,37) = 8.5$, $p < 0.01$). AD patients were more aged than bvFTD and control subjects. This age effect is inherent to the diseases we chose, since AD is an elderly disease that mostly appears after 70 years old, while FTD mostly occur near 65 years old. However, they did not differ for educational level ($F(2,37) = 1.0$, $p < 0.60$) and duration disease for patients (U Mann–Whitney, $Z = 0.17$, $p = 0.91$).

Also, our patient groups were matched for general cognitive efficiency (MMS) ($Z = 1.3$, $p = 0.19$) and executive dysfunction (FAB) ($Z = 0.9$, $p = 0.36$).

Finally, bvFTD patients were impaired in facial emotions recognition test and in theory of mind evaluation compared to matched control subjects recruited in a previous study (Funkiewiez et al., 2012) (see Table 4 for a summary of results).

5.3.2. High-affect case

We analysed results for both cases separately. For the 'high-affect case', four ANOVAs with group of subjects as a factor (control subjects, AD patients, and bvFTD patients) revealed no significant effect of pathology either on 'blameworthiness'

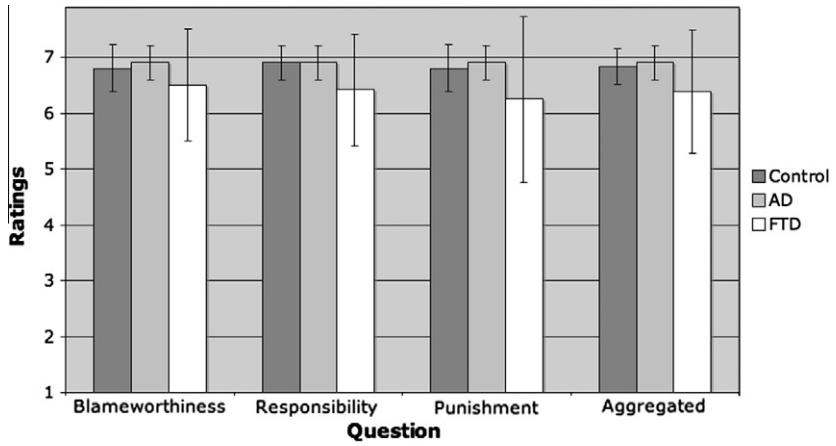

Fig. 3. Mean answers for each question and each group of participants in the 'high-affect case'. Error bars indicate standard deviation.

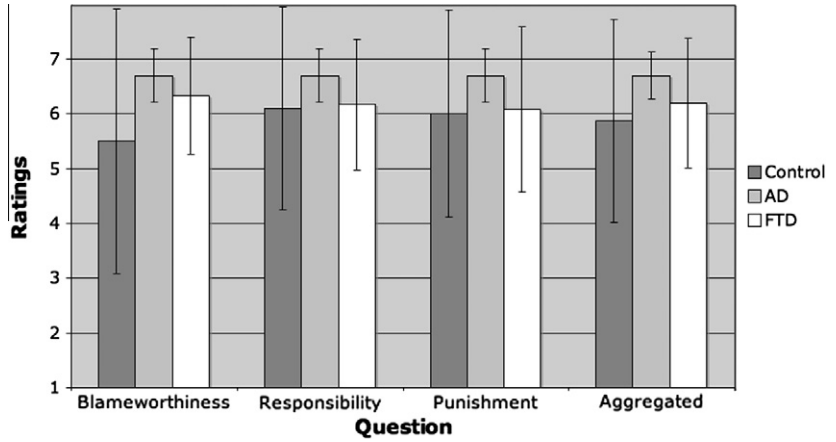

Fig. 4. Mean answers for each question and each group of participants in the 'Supercomputer case'. Error bars indicate standard deviation.

ratings ($F(2,29) = 1.0$, $p = 0.37$), 'responsibility' ratings ($F(2,29) = 2.0$, $p = 0.16$), 'punishment' ratings ($F(2,29) = 1.5$, $p = 0.15$) or aggregated scores ($F(2,29) = 1.7$, $p = 0.21$).

No subjects gave a response below the middle of the scale (4.0) for the 'responsibility' and 'blameworthiness' questions. For the 'punishment' question, only one bvFTD patient gave a response below the midpoint. Thus, subjects in three groups gave mainly compatibilist answers to the 'high-affect case' (see Fig. 3).

5.3.3. Supercomputer case

For the 'Supercomputer case', four ANOVAs with group of subjects as a factor revealed no significant effect on 'responsibility' ratings ($F(2,29) = 1.6$, $p = 0.21$), 'blameworthiness ratings' ($F(2,29) = 0.66$, $p = 0.53$), 'punishment' ratings ($F(2,29) = 0.7$, $p = 0.49$) and aggregated scores ($F(2,29) = 1.1$, $p = 0.36$).

As for the 'high-affect case', subjects in all three groups gave mainly compatibilist answers (see Fig. 4). Only three subjects gave responses below the scale for at least one question. Two were control subjects and one was a bvFTD patient.

6. Discussion

6.1. Implication and interpretation of our results

Our results for both the High-affect and the Supercomputer case seem straightforward: in both cases, bvFTD patients did not give significantly less compatibilist answers than control subjects. In fact, in the Supercomputer case, they even gave more compatibilist answer than control participants (aggregated scores: $M = 5.8$ vs. $M = 6.1$). This goes directly against the predictions that can be drawn from Nichols and Knobe's 'performance error model'. Indeed, for the 'performance error

model', compatibilist responses are the product of an emotional bias that leads people to go against more abstract incompatibilist principles. Thus, the 'performance error model' has for consequences that a subject with impoverished emotional reactions should tend to give less compatibilist answers and more incompatibilist answers. Clearly, this prediction does not match the results we obtained: bvFTD patients with impoverished emotional reactions were not less likely to give compatibilist answers and, overall, gave mostly compatibilist answers. It seems that the 'performance error model' cannot account for these results.

Nahmias and Murray's 'error theory', on the contrary, is fully compatible with these results: people are natural compatibilists and this is why, control subjects or patients, they mostly give compatibilist answers. However, we saw that, to explain Nichols and Knobe's results, Nahmias and Murray needed to add two supplementary hypotheses: first, that Nichols and Knobe's scenario are written in such way that people reading it are biased towards incompabilist answers by understanding determinism as entailing bypassing; second, that this first bias is cancelled by some 'counter-bias' in the high-affect scenario. It is the nature of this 'counter-bias' that we now question: is it really the product of affective reaction, as suggested by Nahmias and Murray?

If this 'counter-bias' is really the product of an emotional response, then it should be absent in patients with emotional deficits, and these patients should tend to give less compatibilist answers in the 'high-affect case' (and in the 'high-affect case' only). What does our result suggest on this point? That it is not the case, so it seems that the 'affective counter-bias' hypothesis in Nichols and Murray's 'error theory' should be replaced by another auxiliary hypothesis for.

A last worry is whether our results can be explained away by certains peculiarities of bvFTD's patients. A first possibility could be that bvFTD patients answer according to a very rudimentary strategy consisting in fully condemning action that are clear moral transgression, determinism notwithstanding, and that this strategy would explain their seemingly compatibilist answers. However, such a hypothesis would be at odd with other findings about bvFTD patients' moral judgments. We already mentioned Mendez et al.'s study of bvFTD patients' answers to moral dilemmas. What they found was that bvFTD patients were more likely to judge morally acceptable the sacrifice of one person to save five others (including in the case in which sacrificing the person amounted to pushing her under a trolley). This pattern of answers is not compatible with the hypothesis according to which bvFTD patients would rely on a very crude strategy of condemning any apparent moral transgression.

A second possibility would be that our results could be explained by the bvFTD patients' deficit in theory-of-mind, as revealed by the Faux-Pas test. However, we do not think so: even if bvFTD had a tendency to overattribute bad intentions and goals to agents (which seems difficult, given that they really have bad intentions and goals), this would not be enough to explain away their compatibilist answers. Indeed, if bvFTD patients had the tendency to consider that an agent is responsible for his action in a deterministic world given that he had the relevant intentions, this would simply mean, once again, that bvFTD patients are compatibilists. Whether one is compatibilist or incompatibilist is not a question a theory-of-mind: it is a question of what one thinks to be the necessary and sufficient component of moral responsibility.

Nevertheless, though such an answer would be enough against an advocate of Nichols and Knobe's 'performance error model', it would still leave one possibility open for an advocate of Nahmias and Murray's 'affective counter-bias'. This advocate could advance the following hypothesis: Nichols and Knobe's scenarios lead people to think that people's intentions have no effect in the production of their action – but this error is corrected by an affective counter-bias in high-affect cases. Surely, bvFTD patients should not benefit from this emotional correction and thus give incompatibilist answers, but their own bias (to overattribute intention) does the same work, and this is why there is no apparent difference between control participants and bvFTD patients in our experiment.

This is indeed a possibility. Nevertheless, we think this hypothesis has a certain disadvantage: as it postulates two different 'counter-bias' to account for our data, it is more costly that an hypothesis that would abandon the idea of an emotional 'counter-bias' to account for control participants and bvFTD patients' answers in the same way. Thus, if such an account existed, it would be more attractive that such a cumbersome (but still plausible) account.[15]

6.2. Beyond emotion: apathetic accounts of intuitions about moral responsibility

So the question is: is there any plausible account of folk intuitions about determinism and moral responsibility who would account for existing data without appealing to emotional reactions? There are two main possibilities for such accounts.

The first possibility is to reject both Nichols and Knobe and Nahmias and Murray's theories to propose a radically novel and alternative account. This is for example what Mandelbaum and Ripley (submitted for publication) have done by advancing the NBAR hypothesis (where NBAR stands for 'Norm Broken, Agent Responsible'). According to them, we have an unconscious belief that whenever a norm is broken, an agent is responsible for breaking the norm. In abstract cases, that belief does not play any role and we, being natural incompatibilists, give incompatibilist answer. Nevertheless, in concrete cases, in which norms are broken, this belief will clash with our natural incompatibilism. People will finally reduce this cognitive dissonance by overriding their incompatibilist intuitions and give compatibilist answers.

[15] Also note that this hypothesis makes the following prediction: bvFTD patients should give more compatibilist answers than control subjects for the 'low-affect case' (cheating on one's taxes). Indeed, in such a case, control subjects should not benefit from the 'affective counter-bias' while patients should still benefit from their 'overattribution counter-bias'.

Though plausible, the NBAR hypothesis faces some difficulty. The first is that, at first sight, it seems unable to explain the difference between the high-affect cases (killing one's family, raping a stranger) in which people give mostly compatibilist answers and the low-affect case (cheating on one's taxes) in which people give mostly incompatibilist answers. Nevertheless, this could be explained by making the hypothesis that cheating on one's taxes is a less salient violation than, say, raping a stranger.

The second difficulty is that Nahmias and his colleagues (2006) have produced a scenario in which participants give mostly compatibilist answers while the agent just goes jogging. As going jogging hardly seems a norm violation, this case poses a problem for the NBAR hypothesis. Nevertheless, this hypothesis is a clear example of account that does not rely on participants' affective reactions and would explain why control participants and bvFTD give the same answers.

Another possibility, though, consists in keeping Nahmias and Murray's error theory but trading their 'affective counter-bias' hypothesis for another auxiliary hypothesis that would not rely on participants' affective reactions. Here is a possibility: that Nichols and Knobe's scenarios lead people to have a wrong reading of determinism. More precisely, they understand that people in Universe A are forced to do what they do. But, in concrete cases, this reading can be corrected if people find it implausible that an agent can be forced to act as described in the vignette. Thus, people might find plausible that someone could be in a situation where she is forced to cheat on her taxes (for example by lack of money) and then stick to the wrong reading of determinism and give incompatibilist answer. But, it might also be that people find hard to imagine a situation in which one would be forced to rape a stranger, and then correct their reading of determinism to the correct one, thus giving compatibilist answers. Though we are currently in the process of testing this hypothesis and cannot assert whether it is the right one or not, it still is an example of how Nahmias and Murray's 'error theory' can be kept while renouncing to use emotional reactions to account for participants' answers.

7. Conclusion

Nichols and Knobe's 'performance error model' and Nahmias and Murray's 'error theory' are the most prominent accounts of patterns in folk intuitions about the relationship between determinism and moral responsibility. Each theory has its own weaknesses: for example, while the 'performance error model' cannot explain why participants still give compatibilist responses to affect-free neutral scenarios (the 'jogger case' in Nahmias et al., 2006), Nahmias and Murray have to forge auxiliary hypotheses to explain why people seem to better understand determinism in the concrete high-affect cases. In this paper, we sought to adjudicate the conflict between those two theories by studying intuitions about determinism and moral responsibility in patients suffering from bvFTD. We found that patients with bvFTD, who suffer from emotional impairment, did not answer differently from control subjects. This suggests, *contra* Nichols and Knobe's 'performance error model', that emotional reactions do not play a key role in generating compatibilist answers. As Nahmias and Murray's 'error theory' is on the contrary perfectly consistent with our results, we conclude that people should prefer their 'error theory' over Nichols and Knobe's 'performance error model'. Nevertheless, we suggest that their auxiliary hypothesis according to which emotions explain the difference between the low-affect and the high-affect case in Nichols and Knobe's experimental paradigm should be changed for an affect-free explanation of this difference.

References

Bertoux, M. L., Habert, M. O., Funkiewiez, A., de Souza, L. C., Volle, E., Dubois, B. (in preparation). Neural correlates of the Social cognitive and Emotional Assessment (SEA) in behavioural variant of frontotemporal lobar degeneration.

Bertoux, M. L., de Souza, L.C., Funkiewiez, A., Leclerc, D., Volle E., Dubois, B. (submitted for publication) Social cognitive and Emotional Assessment (SEA) is a marker of medial and orbital frontal functions: A VBM study in behavioural variant of frontotemporal dementia.

Boccardi, M., Sabattoli, F., Laakso, M. P., Testa, C., Rossi, R., Beltramello, A., et al (2005). Frontotemporal dementia as a neural system disease. *Neurobiology of Aging, 26*, 37–44.

Broe, M., Hodges, J. R., Schofield, E., Shepherd, C. E., Kril, J. J., & Halliday, G. M. (2003). Staging disease severity in pathologically confirmed cases of frontotemporal dementia. *Neurology, 60*, 1005–1011.

Buckwalter, W. & Stich, S. (forthcoming). Gender and philosophical intuition. In Joshua Knobe & Shaun Nichols (Eds.), *Experimental Philosophy* (Vol. 2). Oxford University Press.

Ciaramelli, E., Muccioli, M., Làvadas, E., & Di Pellegrino, G. (2007). Selective deficit in personal moral judgment following damage to ventromedial prefrontal cortex. *Social Cognitive and Affective Neuroscience, 5*, 59–67.

Cova, F. (2011). *Qu'en pensez-vous ? Introduction à la philosophie expérimentale*. Paris: Germina.

Cova, F. & Naar, H. (in press) Side-effect effect without side effects: the pervasive impact of moral considerations on judgments of intentionality. *Philosophical Psychology*.

Dubois, B., Feldman, H. H., Jacova, C., Dekosky, S. T., Barberger-Gateau, P., Cummings, J., et al (2007). Research criteria for the diagnosis of alzheimer's disease: Revising the nincds-adrda criteria. *Lancet Neurology, 6*, 734–746.

Dubois, B., Slachevsky, A., Litvan, I., & Pillon, B. (2000). The fab: A frontal assessment battery at bedside. *Neurology, 55*, 1621–1626.

Ekman, P., & Friesen, W. V. (1975). *Pictures of facial affect*. Palo Alto, CA: Consulting Psychologists Press.

Feltz, A., Cokely, E., & Nadelhoffer, T. (2009). Natural compatibilism v. natural incompatibilism. *Mind & Language, 24*, 1–23.

Fischer, J. (2002). Frankfurt-type examples and semi-compatibilism. In R. Kane (Ed.), *The Oxford handbook of free will* (pp. 281–308). New York: Oxford University Press.

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Researches, 12*, 189–198.

Frankfurt, H. (1969). Alternate possibilities and moral responsibility. *Journal of Philosophy, 66*, 829–839.

Funkiewiez, A., Bertoux, M. L., de Souza, L. C., Lévy, R., & Dubois, B. ((2012). The SEA (Social cognition and Emotional Assessment): A clinical neuropsychological tool for early diagnosis of frontal variant of frontotemporal lobar degeneration. *Neuropsychology, 26*, 81–90.

Greene, J. (2007). Why are VMPFC patients more utilitarian? A dual-process theory of moral judgment explains. *Trends in Cognitive Sciences, 11*, 322–323.

Greene, J. (2008). The secret joke of Kant's soul. In W. Sinnott-Armstrong (Ed.). *Moral psychology* (Vol. 3, pp. 35–79). Cambridge: MIT Press.
Greene, J., Sommerville, R., Nystrom, L., Darley, J., & Cohen, J. (2001). An fMRI investigation of emotional engagement in moral judgment. *Science, 293*, 2105–2107.
Holtzman, G. (in preparation) Why are men incompatibilists?.
Hornak, J., Bramham, J., Rolls, E. T., Morris, R. G., O'Doherty, J., Bullock, P. R., et al (2003). Changes in emotion after circumscribed surgical lesions of the orbitofrontal and cingulate cortices. *Brain, 126*, 1691–1712.
Hornak, J., O'Doherty, J., Bramham, J., Rolls, E. T., Morris, R. G., Bullock, P. R., et al (2004). Reward-related reversal learning after surgical excisions in orbito-frontal or dorsolateral prefrontal cortex in humans. *Journal of Cognitive Neuroscience, 16*, 463–478.
Hornberger, M., Geng, J., & Hodges, J. R. (2011). Convergent grey and white matter evidence of orbitofrontal cortex changes related to disinhibition in behavioural variant frontotemporal dementia. *Brain*.
Knobe, J. (2003). Intentional action and side effects in ordinary language. *Analysis, 63*, 190–193.
Knobe, J., & Nichols, S. (2008). *Experimental philosophy*. New York: Oxford University Press.
Koenigs, M., Young, L., Adolphs, R., Tranel, D., Cushman, F., Hauser, M., et al (2007). Damage to the prefrontal cortex increases utilitarian moral judgments. *Nature, 446*, 908–911.
Lavenu, I., & Pasquier, F. (2005). Perception of emotion on faces in frontotemporal dementia and alzheimer's disease: A longitudinal study. *Dementia and other Geriatric Cognitive Disorders, 19*, 37–41.
Lough, S., Kipps, C. M., Treise, C., Watson, P., Blair, J. R., & Hodges, J. R. (2006). Social reasoning, emotion and empathy in frontotemporal dementia. *Neuropsychologia, 44*, 950–958.
Mandelbaum, E. & Ripley, D. (submitted) Explaining the abstract/concrete paradox in moral psychology: the NBAR hypothesis.
Mendez, M., Anderson, E., & Shapira, J. (2005). An investigation of moral judgment in frontotemporal dementia. *Cognitive and behavioral neurology, 18*, 193–197.
Miller, J., & Feltz, A. (2011). Frankfurt and the folk: an empirical investigation. *Consciousness and Cognition, 20*, 401–414.
Monroe, A., & Malle, B. (2010). From uncaused will to conscious choice: the need to study, not speculate about people's folk concept of free will. *Review of Philosophy and Psychology, 1*, 211–224.
Nadelhoffer, T. (2006). Bad acts, blameworthy agents, and intentional action: some problems for juror impartiality. *Philosophical Explorations, 9*, 203–219.
Nahmias, E., Morris, S., Nadelhoffer, T., & Turner, J. (2005). Surveying freedom: Folk intuitions about free will and moral responsibility. *Philosophical Psychology, 18*, 561–584.
Nahmias, E., Morris, S., Nadelhoffer, T., & Turner, J. (2006). Is incompatibilism intuitive? *Philosophy and Phenomenological Research, 73*, 28–53.
Nahmias, E., & Murray, D. (2010). Experimental philosophy on free will: an error theory for incompatibilist intuitions. In J. Aguilar, A. Buckareff, & K. Frankish (Eds.), *New waves in philosophy of action*. London: Palgrave-Macmillan.
Neary, D., Snowden, J. S., Gustafson, L., Passant, U., Stuss, D., Black, S., et al (1998). Frontotemporal lobar degeneration: A consensus on clinical diagnostic criteria. *Neurology, 51*, 1546–1554.
Nichols, S. (2006). Folk intuitions on free will. *Journal of Cognition and Culture, 6*, 57–86.
Nichols, S., & Knobe, J. (2007). Moral responsibility and determinism: The cognitive science of folk intuitions. *Noûs, 41*, 663–685.
Peelen, M. V., Atkinson, A. P., & Vuilleumier, P. (2010). Supramodal representations of perceived emotions in the human brain. *Journal of Neuroscience, 30*, 10127–10134.
Pereboom, D. (1995). Determinism al dente. *Noûs, 29*, 21–45.
Perry, R. J., Graham, A., Williams, G., Rosen, H., Erzinclioglu, S., Weiner, M., et al (2006). Patterns of frontal lobe atrophy in frontotemporal dementia: A volumetric mri study. *Dementia and other Geriatric Cognitive Disorders, 22*, 278–287.
Piguet, O., Hornberger, M., Mioshi, E., & Hodges, J. R. (2011). Behavioural-variant frontotemporal dementia: Diagnosis, clinical staging, and management. *Lancet Neurology, 10*, 162–172.
Rabinovici, G. D., Seeley, W. W., Kim, E. J., Gorno-Tempini, M. L., Rascovsky, K., Pagliaro, T., et al (2007). Distinct mri atrophy patterns in autopsy-proven alzheimer's disease and frontotemporal lobar degeneration. *American Journal of Alzheimers Disease and Other Dementia, 22*, 474–488.
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.
Ruby, P., & Decety, J. (2004). How would you feel versus how do you think she would feel? A neuroimaging study of perspective-taking with social emotions. *Journal of Cognitive Neuroscience, 16*, 988–999.
Schulz, E., Cokely, E. & Feltz, & A. ((2011). Persistent bias in expert judgments about free will and moral responsibility: A test of the expertise defense. *Consciousness and Cognition, 20*, 1722–1731.
Seeley, W. W., Crawford, R., Rascovsky, K., Kramer, J. H., Weiner, M., Miller, B. L., et al (2008). Frontal paralimbic network atrophy in very mild behavioral variant frontotemporal dementia. *Archive of Neurology, 65*, 249–255.
Sinnott-Armstrong, W. (2008). Abstract + Concrete = Paradox. In J. Knobe & S. Nichols (Eds.), *Experimental philosophy* (pp. 209–230). New York: Oxford University Press.
Sripada, C. S. (in press). What makes a manipulated agent unfree? *Philosophy and Phenomenological Research*.
Stone, V. E., Baron-Cohen, S., & Knight, R. T. (1998). Frontal lobe contributions to theory of mind. *Journal of Cognitive Neuroscience, 10*, 640–656.
Torralva, T., Kipps, C. M., Hodges, J. R., Clark, L., Bekinschtein, T., Roca, M., et al (2007). The relationship between affective decision-making and theory of mind in the frontal variant of fronto-temporal dementia. *Neuropsychologia, 45*, 342–349.
Tranfaglia, C., Palumbo, B., Siepi, D., Sinzinger, H., & Parnetti, L. (2009). Semi-quantitative analysis of perfusion of brodmann areas in the differential diagnosis of cognitive impairment in alzheimer's disease, fronto-temporal dementia and mild cognitive impairment. *Hellenic Society of Nuclear Medicine, 12*, 110–114.
Van Inwagen, P. (1983). *An essay on free will*. Oxford: Clarendon Press.
Vargas, M. (2005). The revisionist's guide to moral responsibility. *Philosophical Studies, 125*, 399–429.
Vargas, M. (2006). Philosophy and the folk: On some implications of experimental work for philosophical debates on free will. *Journal of Cognition and Culture, 6*, 239–254.
Woolfolk, R., Doris, J., & Darley, J. (2006). Attribution and alternate possibilities: Identification and situational constraints as factor in moral cognition. *Cognition, 100*, 283–301.
Young, L., Cushman, F., Adolphs, R., Tranel, D., & Hauser, M. (2006). Does emotion mediate the effect of an action's moral status on its intentional status? Neuropsychological evidence. *Journal of Cognition and Culture, 6*, 291–304.