Experimental philosophy: testing folk intuitions about free will and moral responsibility

Consuelo Luverà - cluvera@unime.it

Dipartimento di Scienze Cognitive, della Formazione e degli Studi Culturali - Università degli studi di Messina

Abstract

In recent years philosophical concepts like free will and moral responsibility have been investigated through a new approach called experimental philosophy. In this research field, empirical methods are used to survey folk intuitions in order to test natural compatibilism or incompatibilism in ordinary people. The result is a mixed outcome about folk intuitions: some surveys support natural compatibilism, whereas others support natural incompatibilism. This disagreement in the answers can relate to some methodological problems about the possibility to test folk intuitions. Through these problems experimental philosophers have come upon several critics, but maybe they can react by pointing out the triple approach of their research: three different kinds of assignment (experimental analysis, descriptivism and restrictionism) deal with three different approaches in the study of folk intuitions about free will and moral responsibility. We think that experimental descriptivism is at the moment the most profitable project because the results of this investigation seem to open to the other two projects.

Introduction

The old debate about free will has always attracted memorable thinkers involved in philosophical discussions. The notion of freedom is very important because it is commonly linked with the possibility to give one's self and the other people moral responsibility. That is a fundamental evaluation for our personal and social lives. There are many ways of dealing with free will and moral responsibility: you can use a philosophical or theological approach, or also a neuroscientific one, or you can also consider the correlation between free will moral responsibility and law. Another way to deal with those questions is to consider folk intuitions about free will and moral responsibility. That's the approach of the new experimental philosophy: it aims at using empirical methods to investigate the pre-philosophical intuitions in ordinary people, or the so-called non-philosophers.

The attempt to understand what ordinary people think about philosophical problems has recently been a very interesting field of study. That is a very useful kind of investigation also because philosophers often support their theoretical position using commonsensical ideas. But, actually, philosophers don't validate whether their intuitions really agree with what the folk thinks. Now, in experimental philosophy, an investigation of the real pre-philosophical intuitions of common people about free will and moral responsibility through the borrowed methods of experimental psychology, is possible . According to this methodology, some people are introduced into a deterministic or indeterministic scenario. Then the experimenter asks them whether they believe an agent acts in such a scenario of his or her own free will and whether he or she can be thought as morally responsible. The outcome of the survey should say which are the real folk intuitions about those notions. The question here is to determine whether common people support natural compatibilism or natural incompatibilism: that is to identify if non-philosophers think that determinism is in opposition to free will and moral responsibility or that notions are perfectly compatible with a indeterministic world. In fact, while compatibilist philosophers believe we can have free will and we are morally responsible even if causal determinism is true, incompatibilists maintain that the existence of free will entails the fallacy of causal determinism.

Testing folk intuitions

During a philosophical experiment ran in 2005 at the Florida State University (Nahmias *et al*. 2005), Eddy Nahmias and his colleagues tested the folk intuitions of some undergraduates who had not studied the free will problem in order to be not influenced by the knowledge of those philosophical discussions. In this way the experiment could capture the real intuitions of people tested. At first the few familiarity with those concepts was a problem: people didn't understand the concept of "determinism" in the technical way philosophers use it. It seemed that they took for granted that this concept was in contrast with free will, maybe because of the abused sentence "the problem of free will and determinism". In order to compensate for that problem the experimenters proposed a test describing a deterministic scenario without using the word "determinism". The participants read the following scenario:

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 25th, 2150 A.D., twenty years before Jeremy Hall is born. The computer then deduces from this information and the laws of nature that Jeremy will definitely rob Fidelity Bank at 6:00 PM on January 26th, 2195. As always, the supercomputer's prediction is correct; Jeremy robs Fidelity Bank at 6:00 PM on January 26th, 2195. (Nahmias *et al.*, 2005 p.566)

The experimenter asked the participants to respond to some questions. A question asked if, when Jeremy robs the bank, he acts of his own free will. The results indicate that the 76% of participants judged Jeremy robs the bank of his own free will, supporting compatibilist folk intuitions. This outcome proves that the participants had no problem to think that free will is possible also if the world is deterministic and people perform their action freely. Experimenters wondered if the participants judged Jeremy a free agent because the action described in the scenario was a bad action. To avoid this problem they developed two more scenarios in which the action described was good or neutral. In the positive situation Jeremy performs the praiseworthy act of saving a child from a burning building. In the neutral scenario Jeremy decides to go jogging. 68% of the participants said Jeremy saves the child of his free will and 79% said he goes jogging of his own free will. So the kind of action have no effects on judgments of free will. Moreover, 83% judged he was blameworthy for robbing the bank and 88% judged he was praiseworthy in the positive case, the rescue of the child. So the natural compatibilism was also about moral responsibility, not only about free will. However, it is hard to test the common sense intuitions of ordinary people, above all because it is very easy that the answers of non-philosophers are influenced by the way in which the scenario is developed. For example, the supercomputer was a means used to make easier the comprehension of some philosophical concepts like determinism as theorized by Laplace. At first, however, participants seemed not to admit the existence of a supercomputer able to predict all the events in the world. Rather they judge the scenario not plausible. The experimenter participation was necessary so that the participants put aside those ideas and admitted the fictional existence of the supercomputer. But the intervention of the experimenter could have influenced the answers making the percentage of compatibilistic judgments not corresponding with the actual intuitions of participants.

There are many problems in testing the natural compatibilist or incompatibilist intuitions on free will and moral responsibility: for example the way of making the questions can be very important in yielding answers contrasting each other. As we saw, Nahmias arose the question of the effect of the positive or negative action described in the scenario through the participants' answers. The *Performance Error Model* is a hypothesis proposed by Shaun Nichols e Joshua Knobe (2007) about this question. According to this model when the scenario describes actions with a high emotional level people give compatibilist answers, but that happens just because of an affective reaction to the story: actually people's underlying theory

is incompatibilist. So, Nichols and Knobe suggest that strong affective reactions can bias and distort people's judgments and that Nahmias' experimental results can be explained by the affective performance error driving participants' compatibilist intuitions. To test their hypothesis Nichols and Knobe ran an experiment: this time there were two possible scenarios: Universe A, in which everything that happens is completely caused by whatever happened before it, and Universe B, in which almost everything that happens is completely caused by whatever happened before it, except for human actions: "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 it does. On the contrary, in Universe B, decisions are not completely caused by the past, and each human decision does not have to happen the way it does" (Nichols and Knobe, 2007 p. 669). The participants answered to two questions: one high affect question and one neutral affect question. The first one asked whether, in Universe A, a man named Bill, attracted to his secretary, who decides to kill his wife and three children to be with his lover, is fully morally responsible for his action. The second question asked just whether in Universe A people in general are morally responsible for their own action. About the first question the majority of people (72%) gave the compatibilist response, judging Bill fully morally responsible for his homicidal action. But about the second one only a low percentage of participants (5%) said that in Universe A people are morally responsible for their actions: 95% of them answered that in the deterministic Universe A people are not responsible: this outcome supports a natural incompatibilist intuition. The model proposed by Nichols and Knobe suggests that when presented with neutral emotional scenarios most people have incompatibilist rather than compatibilist intuitions. So, they maintain that the actual natural disposition of common people is about incompatibilism. But it is true that in our everyday lives we rarely make moral judgments without appealing to our emotions, as it is quite the opposite. Maybe the Nichols and Knobe's model need for clarification. Nevertheless this study is very important because it focused on the importance of emotion in folk intuitions and in the underlying different psychological process. The results of Nahmias compared with Nichols and Knobe's tests is the opposite: according to Nahmias findings people are natural compatibilist, while according to Nichols and Knobe experiments people are natural incompatibilist.

A problem of experimental philosophy investigating folk intuition about free will and moral responsibility is about the contrasting answers given by the subjects tested in the experimental tasks. The variety in the answers depends on many factors, such as the way of putting the questions, the level of emotion aroused by the story, the concrete or abstract conditions of the scenarios and also the familiarity of the participants with philosophical concepts. Another important reason of the great difference in the folk intuitions on free will and moral responsibility can lay in the cultural differences between people. Unfor-

tunately we don't have many cross-cultural investigations about this issue. The big limitation of the studies that were previously analyzed is the lack of a possible cultural comparison of the folk intuitions about free will and moral responsibility. They were run on American people, all of the same culture.

A recent study conducted by Hagop Sarkissian and colleagues, among which Nichols and Knobe (Sarkissian *et al.* 2010), investigates folk intuitions about free will of people from different cultures. The article is entitled "*Is belief in free will a cultural universal?*" and aims to discover if the findings of experimental philosophy concern the folk intuitions of people in general, or only people from western cultures. People from different cultures could have different intuitions about fundamental questions about the notion of freedom. For example, in a study, Hazel R. Markus e Shinobu Kitayama (1991), maintain that there are important differences between people living in western or eastern society about the word "individual". In the western cultures a person is more independent from the rest of the community and affirms his/her identity in autonomously. On the other hand, in the eastern cultures, an individual is tightly linked with the community to which he/she belongs, and with other behavior and thought. For these reasons Sarkissian and colleagues think that the idea of "acting freely" can have different meanings in different cultures. Their philosophical experiment is a cross-cultural study examining intuitions about free will in subjects from the United States, Hong Kong, India and Colombia. In this experiment the scenarios of Nichols and Knobe (2007) were used, that is the deterministic Universe A and the indeterministic Universe B. The questions were two: the first one asked which universe was more similar to real universe, the second one asked whether in Universe A people are morally responsible for their own action. The majority of participants, despite their culture of origin, answered that Universe B is more similar to ours and that in Universe A people aren't morally responsible supporting, in both cases, incompatibilistic intuitions. This study suggests that the culture of origin is not very important in determining the folk intuitions about free will and moral responsibility revealing a striking degree of cross-cultural convergence and supporting the natural incompatibilism in ordinary people. It is important to underline that, even if participants came from different places and cultures, they were all people exposed to western culture in their lives. It would be interesting to test the intuitions of people which were never influenced by other cultures and to investigate the real degree of cross-cultural convergence in folk intuitions.

Methodological problems and remarks on experimental philosophy

Testing folk intuitions about free will and moral responsibility is a very hard task. Problems concern the methodology used, the comprehension of the philosophical notions in the scenarios proposed, the cultural differences in determining folk intuitions. These problems make the

investigation of folk philosophical concepts very difficult and it is fair to say that, maybe, it's early to have a systematic picture of the folk intuitions about free will and moral responsibility, or if it is actually possible . Nevertheless these studies can be useful for a better comprehension of the psychological mechanisms and the cognitive processes underlying philosophical intuitions and about the role played by the folk concepts in the lives of ordinary people. In the past, philosophers questioned if we are free: they can discuss and speculate; philosophers of today can do a philosophical empirical investigation about concepts as free will and moral responsibility.

Some aspects of this new field of philosophical studies leave it open to attacks, above all from a methodological point of view, but not only. For example Antii Kauppinen remarks experimental philosophers because of their belief that their findings could resolve some ancient philosophical matters (Kauppinen, 2007). Actually, as maintain Nadelhoffer and Nahmias, their ambitions are more unpretentious. Their program can be divided in three projects, each one with a different aim and ambition. They claim that some attacks to experimental philosophy don't concern the whole program, but only single projects: even if at first, experimental philosophy is a reliable field of research taken as a whole (Nadelhoffer and Nahmias, 2007).

The first program, called experimental analysis, concern the systematic investigation of what are folk intuitions about philosophical matters, because those are important in the philosophical debates. Kauppinen remarks that in this project something is wrong with the methodology used to detect the real folk intuitions. The second project, called experimental descriptivism, aims to understand the psychological mechanisms and the cognitive processes underlying folk intuitions, and to investigate the importance of these processes in philosophical questions. So, a target of this branch is also to analyze how the difference between the scenarios proposed to subjects can influence their intuitions. The last project, called experimental restrictionism, is about the caution necessarily dealing with folk intuitions: the idea here is that the great number of factors influencing intuitions represent a problem for the use of intuitions in philosophical debate, it's better to dampen enthusiasm and don't abuse of folk intuitions in philosophy.

The main objection made by Kauppinen is about the methodology used in experimental analysis. He claims that the scenarios and surveys can only inform us of *surface intuitions*, but that they don't help us in the project of finding out the folk concepts, the real *robust intuitions* of common people (Kauppinen, 2007). Robust intuitions can be investigated only when three conditions occur: the subjects tested are competent users of the concepts, the conditions in which judgments are made must be ideal and, finally, you need to avoid pragmatic factors influencing intuitions. About the first condition, when a person is speaking, we usually think that she is a competent speaker about the used concepts. For example, if you are speaking about freedom we usually think that you have a certain knowledge about what you are speaking and that you are

able to use the same concepts conforming the different possible contests and in a certain number of cases. Above all, competent users are those whose application of the concept generally matches the conceptual norms prevailing in the linguistic community. But people are committed to those rules at least implicitly and follow them blindly. Moreover, concept-users have different degree of competence in the use of the conceptual norms. For example a person could define an object as red and another one could say that the same object is dark red, or scarlet, or magenta, etc, because he or she has a better competence in the use of the concept "red". A person could be a competent speaker for some concepts, but not enough to use the concepts in philosophical discussions. According to Kauppinen when just using surveys like experimental philosophers do, it is very difficult if not impossible to determine which speakers are competent users. About the second conditions, Kauppinen maintains that the subjects tested by experimental philosophers could be in ideal conditions during the task: that is the conditions in which judgments are made must be favorable to avoid performance errors, without distorting factor or limits of information, access or ability. For example, we do not treat judgments about colors made in certain poor kinds of lightning. You need ideal conditions to be able to look at robust intuitions of people. But Kauppinen thinks that the experimental set of surveys doesn't offer ideal conditions to participants of the tasks. Finally, even if experimental philosophers are able to test only competent speakers in ideal conditions, that couldn't be enough. In fact, according to Kauppinen it is not easy to separate the contribution of semantic and pragmatic considerations to what people say. He proposed, for example, the sentence "I voluntarily had lunch yesterday" when the word "voluntary" it's not false, but only pragmatically inappropriate because it violates the Gricean pragmatic principle of quality. With the methodology used by experimental philosophers it is impossible separate those two aspects of language (Kauppinen, 2007).

Nadelhoffer and Nahmias respond to the main criticisms put forward by Kauppinen claiming that experimental philosophers tend to be perfectly aware of the potential perturbing factors of their investigations (Nadelhoffer and Nahmias, 2007). They maintain that experimental philosophers do use controlled and systematic experiments to shed light on actual folk intuitions of the participants in the experiments. They believe that their methodological tools are less imperfect than Kauppinen suggests. In fact experimental philosophers use statistical methods and sufficiently large sample sizes. So it is highly unlikely that most participants were inattentive or confused. Moreover the experimenters are committed to avoid participants affected by certain irrelevant factors: for example they are instructed to read the scenarios and questions carefully and have ample time to complete the surveys. Experimental philosophers are used to check whether the subjects have understood the scenarios proposed. If someone of them looks like careless or not competent in the task, he is excluded from the analysis of the results. Moreover, experimental philosophers often ask participants to explain why they responded to questions the way they did, in order to know whether subjects have really understood the scenarios and to improve the future investigations (Nadelhoffer and Nahmias, 2007). Even using these cautions, the potential distorting factors suggested by Kauppinen may be however a problem for experimental philosophers. But the aim of experimental descriptivism is to run controlled and systematic studies primarily to discover psychological factors that drive conflicting intuitions, and so also to discover what could be the possible problematic influences on the responses. In fact the Performance Error Model elaborated by Nichols and Knobe (2007) is just a try to detect the situations that can influence the answers of the surveys. We can say that experimental philosophy, as a whole in its three projects, has an ambitious but actually important target: to investigate folk intuitions also in the complexity of this task and to catch the nuances that we are able to understand.

It's too early to say whether experimental philosophy will be actually useful for the philosophical debate about free will and moral responsibility. That happens also because it is a recent field of research. Some critics claim that it is very difficult to run an accurate and systematic study of folk intuitions about philosophical concepts because of the great number of variables potentially involved in the processes concerning folk intuitions (Paglieri, 2009). However it's also true that experimental philosophers have no intention of declaring any theory as true or false only as regards folk intuitions. Rather they aim to highlight the intuitions of ordinary people about philosophical concepts because many philosophers declare that what they think meets the requirements of the common sense psychology and use this argument to support their thesis. Obviously the data provided by experimental philosophical investigations have to be controlled and accurately examined (Marraffa, 2009).

Future prospects of experimental philosophy

Probably the strength of experimental philosophy lays on the match of three different linked programs: analyzing intuitions, describing underlying mechanisms, restricting excessive enthusiasm. As for criticism towards this method of investigating philosophical concepts, it seems that the descriptivist project is the best conducted and most profitable. In fact the discovery of *which* are the intuitions of all ordinary people about a certain concept (that is the aim of experimental analysis) is a really difficult goal, it's easy to make some mistakes in the evaluation of popular intuitive concepts. Maybe for this reason experimental restrictionism invites to a cautious approach on the study of folk intuitions. But the analysis of *what* underlying intuitions of ordinary people (that is the aim of experimental descriptivism) is the most promising project within the experimental investigation of intuitions about philosophical concepts. This project can be useful to achieve a better comprehension of potential influences on intuitions of the folk and the role played in human lives by freedom and moral responsibility concepts. Moreover these investiga-

tions can help us to understand the way in which our mind works to support or reject a philosophical theory. The descriptivist program is also fundamental in determining the success of the other two projects: it seems impossible to discover which are the intuitions of ordinary people without a previous understanding of the mechanisms through which they emerge and the possible distorting variable and psychological factors.

Experimental philosophy is a recent field of research, but it already provided several data that we ought to consider. Certainly the criticism towards some methodological problems is well grounded. But it seems a new stimulating approach to break the stalemate of free will and moral responsibility debate.

References

KAUPPINEN, A. (2007) «The rise and the fall of experimental philosophy» in *Philosaphical Explorations,* n.10 (2), pp. 95-118.

MARKUS, H., & KITAYAMA, S. (1991) «Culture and the self: Implications for cognition, emotion, and motivation», in *Psychological Review*, n. 98, pp. 224-253.

MARRAFFA, M. (2009) «Lo studio empirico delle intuizioni, ovvero perché la filosofia ha bisogno della scienza (e in particolare della psicologia sociale)», in *Sistemi intelligenti*, n. 2, pp. 317-333.

NADELHOFFER, T., NAHMIAS, E. (2007) «The past and the future of experimental philosophy», in *Philosophical Explorations,* n. 10 (2), pp. 123-149.

NAHMIAS E., MORRIS S., NADELHOFFER T., TURNER J. (2005) «Surveying Freedom: Folk Intuitions about FreeWill and Moral Responsibility», in *Philosophical Psychology* n. 18, pp. 561–84.

NICHOLS, S., KNOBE, J. (2007) «Moral responsibility and determinism: the Cognitive science of folk intuition», *Nous*, n. 41, pp. 663-685.

PAGLIERI, F. (2009) «La filosofia sperimentale: distinzioni e cautele», in *Sistemi intelligenti*, n. 2, pp. 355-369.

Sarkissian, H., Chatterjee, A., De Brigard, F., Knobe, J., NICHOLS, S., SIRKER, S. (2010) «Is belief in free will a cultural universal?», in *Mind and Language*, n. 25 (3), pp. 346-348.