

Editors' note

The paper

Nenad Natan Popovic: *New Experiments On Psychological Egoism*. KRITERION – Journal of Philosophy, 30:3, pp.65-76.

is retracted from *KRITERION – Journal of Philosophy*. The reason for this decision is that we have been informed that it was simultaneously submitted to several journals, including our journal, on August 13, 2016. This is clearly against the rules for submitting to *KRITERION – Journal of Philosophy*.

Christian J. Feldbacher-Escamilla & Alexander Gebharter

New Experiments On Psychological Egoism



NENAD NATAN POPOVIC

Abstract

Psychological egoism is the doctrine that all of our actions are ultimately motivated by self-interest, even including seemingly altruistic actions. Although this view is commonly referred to as “vacuous” or “trivially true” because of its immunity to counterexamples, I argue that there are possible observations that can refute this thesis. I describe the experiments, identify the results that could in principle falsify the strongest possible version of psychological egoism and formulate a probabilistic argument against this view. Ultimately, I conclude that a “trivially true” formulation of psychological egoism is an empirical theory that is in all likelihood false.

Keywords: *psychological egoism, altruism, self-interest, unselfish behavior*

1 Introduction

Anyone who has ever taught a philosophy course that includes the doctrine of psychological egoism can easily imagine the following dialogue taking place in the classroom. In fact, the excerpt below is a brief description of an exchange of arguments that many of us have witnessed numerous times. Let (DEF) stand for a defender of psychological egoism, and let (OPP) stand for an opponent of this view:

(OPP): Psychological egoism must be false. We don't always act out of self-interest. There are many people who donate their time, effort and money to helping others.

(DEF): No, that does not show that they are not acting out of their self-interest. If you volunteer or give money to charity, perhaps

you are doing it for tax purposes or to have something to put on your resume; perhaps you just want to feel good about yourself, or perhaps you know that you would feel bad if you didn't help; maybe you want others to think highly of you because you are helping people in need; or maybe you just can't stand to see others suffer, so by helping them you are helping yourself. I can't tell which explanation is true, but it must be one of these. So, people who help or care about others are still ultimately motivated by their self-interest.

(OPP): What about people who risk their own lives or their own well-being to save others? What about a parent who jumps in front of a car to save their child? A soldier who throws himself on a grenade to save others?

(DEF): Perhaps they know that they couldn't live with the fact that they didn't try to do something, or maybe they identify their self-interest with the well-being of others. Either way, it still does not mean that they are not motivated by self-interest. In those circumstances people simply choose what they think is the lesser of the two evils at the time when they are acting.

(OPP): What if I just give you my wallet and tell you to keep the money. Believe me, I don't want to do that at all. But if I did it just to prove a point, would that count as behavior contrary to psychological egoism?

(DEF): No, it wouldn't. If you did it because you wanted to show that psychological egoism is false, then that would be your self-interest and you would still be acting in accordance with this view.

(OPP): So there couldn't possibly be a case in which a person doesn't act out of their self-interest, then. If a person did help others and thereby promoted his own welfare, then he is self-interested, while if in the same circumstances he helped others but sacrificed his welfare to do so, then he is equally self-interested. If that is what psychological egoism is stating, then the theory is vacuous. You are just defining actions in terms of self-interest, and self-interest itself is defined so broadly that absolutely anything can be categorized under it. Of course there can't be a counterexample to this theory.

While it is possible to distinguish between several different formulations of psychological egoism,¹ my focus will be on a version of this view that has been implicitly referred to in the previous dialogue. Thus, by psychological egoism I shall understand the view that all of our actions are ultimately motivated by self-interest, and that even seemingly altruistic motives are in fact just instances of motives intended to further our own self-interest. More specifically, my target will be the following thesis:

(PE): For every agent S and for every action A, S will willingly perform A at t0 only if S believes at t0 that performing A will promote S's self-interest.

Psychological egoism is notoriously difficult to disprove if the notion of "self-interest" is understood in the way in which it was used by (DEF) in the previous dialogue.² As it has been illustrated, no matter what kind of counterexample we try to offer, it seems that there is always an explanation (more accurately, a disjunction of possible explanations) available to the supporter of (PE). It looks as if this theory is therefore immune to any scenario and that it cannot be refuted by means of any conceivable counterexamples. Perhaps this point is best summarized by Feinberg:

'Until we know what they [psychological egoists] would count as *unselfish* behavior, we can't very well know what they mean when they say that all voluntary behavior is *selfish*. And at this point we may suspect that they are holding their theory in a "privileged position"—that of immunity to evidence, that they would allow no *conceivable* behavior to count as evidence against it. What they say then, if true, must be true in virtue of the way they define—or re-define—the word "selfish." And in that case, it cannot be an empirical hypothesis.'³

I take it that this is the primary reason why this version of psychological egoism is commonly referred to as a "trivially true," "vacuous," "uninteresting", "unfalsifiable in principle" or just "trivial." This is also the primary reason why most of the literature on this subject instead focuses on other, "more interesting" formulations of psychological egoism.⁴

I will argue that, contrary to this common misconception, there are possible observations that can refute (PE). I will specify some experiments that can be performed in order to test this theory empirically and identify the outcomes that would show that the thesis under discussion is false. Once this result is established, it will become apparent that psychological egoism cannot be a trivial thesis, even when it is defined in a way which includes seemingly altruistic acts. Rather, my ultimate conclusion will be that (PE) is false in all likelihood, although we may have to wait for the experiments to be carried out in order to get final evidence.

2 *The (PE) Experiments*

Suppose that a group of subjects is invited to participate in a study involving the use of polygraphs. Each participant will receive an honorarium for answering a number of questions administered by a polygraph specialist. Suppose also that the subjects who have been chosen are not familiar with the techniques used in this process.

For any subject S, a set of *irrelevant* questions and *control* questions is administered first.

After this phase has been completed, the subjects are then asked if they would like to donate a small amount of their honorarium to some charitable organization. The subjects are told that they are under no obligation of any kind to make a donation, nor that it is expected of them to do so in any sense. If S refuses to make a donation, the test is over and no further questions are asked. If, on the other hand, S decides to make a donation, the test continues. S is then told to confirm the donation by touching the screen of a monitor several times, until the final confirmation window appears informing S that his donation will be processed as soon as he presses a green “YES” button on the screen. However, this button is not available and can only be enabled by the specialist administering the test. At this time, S is presented with a set of *relevant* questions that pertain directly to S’s possible motives for making a donation. This set of questions consists of claims that would usually be used by a supporter of psychological egoism, and S is supposed to provide an answer to each of them. Suppose that the set of relevant questions contains the following:

Q1: Do you want to make this donation because you are expecting some benefit from it?

- Q2: Do you want to make this donation because you want to feel good about yourself?
- Q3: Do you want to make this donation because you believe you will get any pleasure out of making this donation?
- Q4: Do you want to make this donation because you would feel bad if you did not do it?
- Q5: Do you want to make this donation because you believe that you would regret it later if you did not do it?
- Q6: Do want to make this donation because you want to impress the test administrator in some sense?
- Q7: Do you want to make this donation because you want to impress someone else later on when you tell this story?
- Q8: Do you believe that something negative will happen to you if you do not make a donation?
- Q9: Do you believe that something positive will happen to you if you make this donation?
- Q10: Is it in your self-interest to any degree to make this donation?

Suppose that the test administrator is instructed to enable the button that allows S to make a donation if and only if S gives a negative answer to all control questions Q1 – Q10 and passes the polygraph test for each of those questions. If both of these conditions are satisfied, the administrator will immediately enable the confirm button and the donation will be processed as soon as S touches the screen. If S gives an affirmative answer to any of the control questions Q1 – Q10, the administrator will not enable the confirm button but will inform S that there was a problem with the software and that S's donation cannot be processed at this time. Similarly, S will not be in a position to make a donation if he fails the polygraph test for any of the control questions that he provided a negative answer to.

If this experiment is performed and S ends up making a donation, we will be provided with evidence that (PE) is false since S has *probably* acted in a purely altruistic manner.⁵ If several subjects end up making a donation the evidence will get proportionally stronger, and if a significant portion of all tested subjects passes the test (PE) will be outright refuted.⁶

We have thus identified outcomes that would show that (PE) is false, which means that psychological egoism is actually an empirical thesis that can be tested in principle by means of experimentation. While we may have to wait for the results of those experiments to come in before we can proclaim our final verdict, there are at least some indicators that we should expect (PE) to be refuted by those results. If (PE) were true, then is highly improbable that anyone in the world could ever truthfully give a negative answer to all of the questions (Q1) – (Q10).⁷ Yet, there is something very peculiar about this result since we can undoubtedly envision that occurring, which already furnishes us with a strong reason to suspect that (PE) is in fact false. Also, there clearly are numerous individuals who believe that genuinely altruistic acts are possible, and it seems reasonable to expect that they would truthfully deny any questions suggesting the contrary. This also indicates that a significant portion of the subjects should be expected to pass the test and thus provide us with relatively strong evidence against (PE).

3 *The Ultimate Objection*

The experiments that have been described provide opponents of (PE) with a threatening argument against this doctrine. We can thus easily imagine the initial dialogue between (OPP) and (DEF) extend in this direction:

(OPP): What about people who genuinely and sincerely believe that their actions are not motivated by self-interest? What if we conduct experiments using polygraph and it turns out that a significant number of people can pass this test? Would that not constitute ultimate evidence against psychological egoism?

At this point, a supporter of (PE) seems to be left with only two options. He can either concede that the aforementioned experiments are relevant and thus expose the doctrine he is defending to refutation, or he can resort to the following argument:

(DEF): Our motives for acting are not infallibly known. Just because people believe that they are not acting selfishly, it does not follow that they are not acting selfishly. Perhaps they are motivated by subconscious mechanisms and are not aware of their deep egoistic motives; perhaps they are self-deceived or simply mistaken about their real motives. Either way, the experiments

can only show that some people may believe that they aren't acting out of self-interest, but they can't demonstrate that people aren't acting out of self interest. Therefore, the experiments described are irrelevant. They can't possibly show that psychological egoism is false.

Is the reply given by (DEF) satisfactory and can we simply dismiss the previous argument against (PE) on the grounds that people can be unaware or mistaken about their own reasons for acting? Should we conclude that the experiments are altogether irrelevant to the status of (PE) since we may not infallibly identify the factors that motivate our actions? I will argue that the argument given by (DEF) is fundamentally flawed and that, although the main premise is true and our motives for acting are not infallibly known, the conclusion does not follow.

4 *The Ultimate Reply and the Intuitive Appeal of the Probabilistic Argument*

As mentioned previously, it would be outright unreasonable to deny the first premise of the argument. Of course it is possible that people can be mistaken about their own motives. However, it is prima facie implausible to claim that *all* people who *ever* pass the lie detector test *must* hold false beliefs about their own motives. Yet, this is exactly what (DEF) has to maintain in order to dismiss the experiments.

So, while this experiment does not rule out the possibility of subjects being mistaken about their own motives and while it is a far cry from concluding that all subjects who pass this test are de facto motivated by unselfish reasons, we must be reminded that it is the supporters of (PE) that are arguing for a universal thesis. To falsify (PE) we only need one instance of behavior that is not motivated by self-interest, and the argument simply states it is *improbable* that all of those who pass the lie detector in the circumstances described are mistaken about their own motives.

Additionally, it seems intuitively clear that the results of the experiments cannot be irrelevant. To illustrate this, suppose that out of a group of a thousand subjects only one ends up making a donation. In this case it would not seem as unusual to argue that a small percentage of individuals can be mistaken about what really motivates their actions. However, suppose that the experiments show that more than 50% of people who are tested in fact end up in that category. At this

point, making the same claim as before would already start to sound farfetched, as we would have to maintain that most people are either self-deceived or mistaken about their own motives in order to preserve (PE).

Even worse for the supporter of (PE), this argument could be pushed further. Suppose that altruistic ideas become culturally dominant at some time in the future, to the point that nearly 100% of the population believes that we should help others even if it does not advance our own self-interest, and that nearly everyone actually acts on that principle most of the time. In that case it would not be hard to imagine that nearly 100% of the tested subjects would pass the polygraph test. Yet, supporters of (PE) would still have to attribute false beliefs to almost everyone and argue that virtually all humans are systematically mistaken about their motives with respect to most of their actions. At least, this should be considered a highly undesirable consequence of (PE).

In conclusion, it seems clear that our intuitions may at least somewhat fluctuate depending on the number of people who end up passing the polygraph test. There is an intuitive difference between one ending up in this category, versus someone or nearly everyone honestly believing that their actions were not motivated by self-interest. It seems clear that the actual results of the experiments bear at least *some* relevance, and therefore the experiments and the probabilistic argument against (PE) cannot be entirely dismissed based on the claim that our reasons for acting are not infallibly known. So, although the main premise of the argument made by (DEF) is true, its conclusion does not follow.

Additionally, the case against the argument made by (DEF) can be further strengthened by focusing on the circumstances of the subject. According to the objection, subjects who pass the polygraph test *must* hold false beliefs about their own motives, presumably because they are mistaken, self-deceived or unaware of their deep egoistic motives.

However, we must be reminded that the setup of the experiments includes questions that have been specifically designed to address these concerns. According to the setup, subject S has been explicitly asked about possible egoistic motives. Subject S has thus been placed in the position to be reflective about his true reasons for acting as he considers those questions and how they relate to his actions. If S was mistaken or failed to even consider any alleged egoistic motive that he might have while he was performing the action, he surely became aware of any such motive during this part of the experiment and reflecting on the questions

will increase the likelihood of identifying its potential relevance.

Perhaps this may not be enough for the subject to uncover just about any deep egoistic motive that he might have, but it certainly minimizes the likelihood that *all* subjects who pass the test hold false beliefs about their motives. Besides, the setup of the experiment can easily be further amended to neutralize the force of even this objection. For example, we can stipulate that the subjects must undergo a session with an expert who specializes in helping the subjects uncover their deep or hidden motives before answering the questions. By virtue of this setup that can be even further amended if needed, it is *prima facie* unlikely that all subjects who end up giving negative answers are thus motivated by hidden reasons that are inaccessible to them.

Furthermore, the polygraph test itself is designed to detect any unsettling pattern that the subject might exhibit. As such, it should be particularly troubling for people who are sensitive or worried about a certain subject or self-deceived about their real motives. Consider someone who would find it painful to know that what he sees as purely altruistic motives are in fact just instances of egoistic motives, and yet who is by the nature of the experiment suddenly placed in the position to directly process this possibility. It seems fairly likely that it would be detected during the test. As we have seen previously, rather than being mistakenly classified as false negatives, subjects who fall in any such category are likely to produce a disproportional number of false positives.

In summary, if subject S is wrong about his own motives, it is either because S is unaware, mistaken or self-deceived about his real motives. If S is unaware or mistaken about his true motives, the questions will likely make him aware of it; if S is self-deceived or refuses to acknowledge his true motives, the polygraph test will likely detect it. Therefore, it is highly improbable that all subjects who pass the test are wrong about their own motives. Therefore, it is highly likely that (PE) is false.

The previous argument tried to establish that the individuals who pass the test are unlikely to be mistaken about their motives, but it is important to note that the argument is fatal to (PE) practically as long as there is any non-zero probability that any individual subject who passes the test is speaking truly. For example, suppose that each individual subject who passes the test is 90% likely to be mistaken. By simple probabilistic calculations, if 7 subjects pass the test on this assumption it is already more than 50% probable that at least one of them is actually speaking truly; if 22 subjects pass the test the probability of (PE) being false is already above 90%. So, even if we disregard the

previous points and allow for the sake of the argument that subjects are likely to be self-deceived or otherwise mistaken about their own motives, even a relatively small number of confirmations will still allow us to conclude that it is highly improbable that (PE) is true. The only way for a supporter of (PE) to escape this result is to maintain that that it is *impossible* for the subjects to speak truly in those circumstances, or if the experimental results actually fail to produce sufficient evidence.

In conclusion, it cannot be denied that our reasons for acting are not infallibly known. This fact, however, does not constitute sufficient ground for dismissing the probabilistic argument against (PE) and the experiments intended to provide the evidence. There is a clear intuitive appeal that the results of the experiments cannot be completely irrelevant, and it is highly improbable that *all* subjects who pass the polygraph test *must* hold false beliefs. The premise of the argument provided by (DEF) is true, but the conclusion simply does not follow. Consequently, (PE) turns out to be an empirical claim that can be tested by means of observation-based experiments rather than being a “vacuous” thesis. Far away from being “trivially true,” psychological egoism in the end is just a “sitting duck” waiting to be demonstrated false once and for all.

Notes

- 1 For a summary of possible interpretations of psychological egoism see [8, pp. 501-503].
- 2 [1] argues that the experiments done in the field of social psychology provide empirical evidence against psychological egoism and for the empathy-altruism hypothesis. See [11], [12], and [5] for criticisms of Batson’s conclusions; ([1]) also see [2] for replies. [13] argue against psychological egoism based on evolutionary theory. For criticism of Sober and Wilson ([11]) see [12]. For traditional philosophical arguments see [4], SermonXI and [7, p. 166]. While these arguments have merits that have widely been recognized and while they are effective against many formulations of psychological egoism, a defender of (PE) can still insist on taking the same route as described in the initial dialogue and argue that no genuine counterexample to the theory has been presented.
- 3 [6, p. 503]
- 4 For the most part, this seems to apply to both critics and supporters of the view. For example, [9] makes a distinction between weak and strong psychological egoism and argues that the former is true while the latter is false, although neither formulation is identical to (PE). There are exceptions, though: [3] defends a version of psychological egoism that is very similar to (PE) on the grounds that it is the best explanation for human actions.
- 5 A defender of (PE) will surely argue that we can be mistaken about our own motivation and that our reasons for acting are not infallibly known. Although

it should be perfectly clear that I am offering a probabilistic argument against (PE), I will specifically address this objection in the next section.

6 According to a study conducted by the National Research Council ([10, p. 214]) for subjects untrained in countermeasures polygraph tests can discriminate lying from truth-telling at rates well above chance, though well below perfection. Also, while the National Research Council study ([10, Appendix I]) questions reliability of polygraph testing, notice that the main problem lies with producing a large number of false positives. So, far more truth-telling people test as deceptive and fail the test than the other way around. While this can be seen as a flaw of the method, it will actually significantly increase the overall value of evidence against (PE) obtained by means of our experiments. Polygraph testing may not be useful to identify a spy out of a group of thousand people since hundreds of innocents will also fail the test, but it will be very useful for our experiments since the test is likely to mistakenly classify a genuinely altruistic motive as egoistic. For example, if each one of a hundred subjects who are tested acts altruistically and if the assumed accuracy rate of the method is 80% we should only expect to get 80 confirmations of altruistic subjects or at best a few more. If fifty out of a hundred subjects act altruistically, we will likely get only 40 confirmations. If there are ten subjects who acted altruistically, we should expect to see 8 confirmations and so on. Given the two results established by the National Research Council, even a single case of altruistic behavior recorded by means of this experiment should count as very strong evidence against (PE).

7 Of course, there might be some overlooked selfish benefit not included in Q1 – Q10. If any such motive is identified, the questions should be adjusted accordingly.

Nenad Natan Popovic
 Miami Dade College
 USA
 <npopovic@mdc.edu>

References

- [1] C. D. Batson. *The Altruism Question: Toward a Social-Psychological Answer*. Hillsdale, NJ: Lawrence Erlbaum Associates, 1991.
- [2] C. D. Batson. *Altruism in Humans*. New York: Oxford University Press, 2011.

- [3] S. Berman. “A Defense of Psychological Egoism.” In: *Desire, Identity and Existence: Essays in Honor of T.M. Penner*. Ed. by N. Reshotko. Academic Printing and Publishing, 2003, pp. 143-157.
- [4] J. Butler. “Fifteen Sermons Preached at the Rolls Chappel”. In: *British Moralists, 1650–1800*. Ed. By D. D. Raphael. Indianapolis: Hackett, 1991, pp. 325-386.
- [5] R. B. Cialdini, S. L. Brown, B. P. Lewis, C. Luce and S. L. Neuberg. “Reinterpreting the Empathy-Altruism Relationship: When One Into One Equals Oneness”. In: *Journal of Personality and Social Psychology* 73(3) (1997), pp. 481-494.
- [6] J. Feinberg. “Psychological Egoism”. In: *Reason and Responsibility*. Ed. by J. Feinberg and R. Shafer-Landau. Belmont, CA: Wadsworth, 15th edition, 2013, pp. 501-512
- [7] D. Hume. “Of Self-Love”. In: *An Enquiry Concerning the Principles of Morals*. Ed. by T. L. Beauchamp. Oxford University Press, 1998, Appendix II.
- [8] H. LaFollette. “The Truth in Psychological Egoism”. In: *Reason and Responsibility*. Ed. by J. Feinberg. Belmont, CA: Wadsworth, 7th edition, 1989, pp. 500-507.
- [9] M. Mercier. (2001). “In Defence of Weak Psychological Egoism”. In: *Erkenntnis* 55 (2001), pp. 217-237.
- [10] National Research Council of the National Academies, Committee to Review the Scientific Evidence on the Polygraph, Board on Behavioral, Cognitive and Sensory Sciences, Committee on National Statistics and Division of Behavioral and Social Sciences and Education. *The Polygraph and Lie Detection*. Washington, D.C.: The National Academies Press, 2003.
- [11] E. Sober and D. S. Wilson. *Unto Others: The Evolution and Psychology of Unselfish Behavior*. Cambridge, MA: Harvard University Press, 1998.
- [12] S. Stich, J. M. Doris and E. Roedder. “Altruism”. In: *The Moral Psychology Handbook*. Ed. by J. M. Doris and the Moral Psychology Research Group. Oxford University Press, 2010, pp. 147-205.