

Priming Effects and Free Will

Ezio Di Nucci (Universität Duisburg-Essen)

Abstract I argue that the empirical literature on priming effects does not warrant nor suggest the conclusion, drawn by prominent psychologists such as J. A. Bargh, that we have no free will or less free will than we might think. I focus on a particular experiment by Bargh - the 'elderly' stereotype case in which subjects that have been primed with words that remind them of the stereotype of the elderly walk on average slower out of the experiment's room than control subjects – and I show that we cannot say that subjects cannot help walking slower or that they are not free in doing so. I then illustrate how these cases can be reconciled and normalized within a Davidsonian theory of action to show that, in walking slower, subjects are acting intentionally. My argument applies across various experiments, including those of goal priming. In the final section I argue that the only cases in which priming effects are efficacious are so-called Buridan cases.

There is increasing interest, amongst psychologists, for the consequences of behavioural priming on the idea of free will. There is evidence that, through priming, agents can be made to automatically do certain things, or take certain attitudes, or choose one thing over another – without their realizing the influence of the priming or their altered actions, attitudes, or choices: 'social norms', 'emotions', 'goals', 'stereotypes', and 'social behaviours' are some of the things that can be influenced through priming (Bargh 2006: 147).

This kind of challenge to free will must be distinguished from the philosophical consequences that Libet (1985) drew from his neurological findings; but it shares, with Libet's, the presumption to demonstrate empirically that there is no free will; or, at the

very least, that people have less free will than they ordinarily think (Bargh (2008: 147) himself likens his project to Libet's).

In this paper I look at a famous example of this kind of challenge to free will, Bargh's: showing that the empirical data does not demonstrate or even suggest that there is no free will or even less free will than unwise laymen might suppose. Bargh is very explicit about the overall working hypothesis of his research: "there is no need to posit the existence of free will in order to explain the generation of behavioral inputs, and there is no need to posit free will in order to explain how those (unconscious) impulses are sorted out and integrated to produce human behavior and the other higher mental processes" (2008: 148).

1. Bargh's 'elderly' priming

Bargh's experiment is as simple as it is stunning: a group of people is, without their realizing, primed with the stereotype of the elderly through what subjects think is a purely linguistic scrambled-sentence task. The test contains a disproportionate amount of words related to the 'elderly' stereotype: "worried, Florida, old, lonely, grey, selfishly, careful, sentimental, wise, stubborn, courteous, bingo, withdraw, forgetful, retired, wrinkle, rigid, traditional, bitter, obedient, conservative, knits, dependent, ancient, helpless, gullible, cautious, and alone" (Bargh et al. 1996: 236).

Remarkably, in leaving the experiment's room subjects tend to walk slower than a control group whose test did not contain a disproportionate amount of words related to the 'elderly' stereotype: "participants primed with the elderly-related material subsequently behaved in line with the stereotype – specifically, they walked more slowly down the hallway after leaving the experiment" (Bargh and Chartrand 1999: 466)¹. Participants do not show any awareness of their walking slower than 'normal' or of their being presented with a 'bias' list of words: "No participant expressed any

¹ According to my own calculations, the group primed with the elderly stereotype does on average a speed of, approximately, 4.2 km/h, while the control group does 4.8 km/h.

knowledge of the relevance of the words in the scrambled-sentence task to the elderly stereotype” (Bargh et al. 1996: 237).

The idea is that had a subject not been primed with the ‘elderly’ stereotype, she would have walked faster down the hallway. And that therefore the subject has been ‘made to’ walk slower in such a way that her control over her own behaviour has been diminished or taken away. We might take this to mean that a subject from the ‘elderly’ group isn’t as free as a subject from the control group. I don’t think that the inference from these findings to diminished responsibility or free will is warranted. Furthermore, I think that subjects from the ‘elderly’ group have as much control over their walking speed as subjects from the control group. Therefore I don’t think that this experiment has anything to say about free will.

I don’t mean to deny the counterfactual according to which, had subjects not been primed, they would have walked faster. But I don’t think that anything interesting, with regards to free will, follows from that counterfactual. It’s not at all clear that ‘elderly’ subjects can’t help walking slower; that is, walking, on average, at 4.2 km/h. They are free to increase (or indeed decrease) their speed at will: suppose, for example, that the agent receives an emergency phone call, or that the fire alarm goes off, or that she starts to wonder whether she has left her phone in the experiment’s room.

Since my argument is that priming experiments have no implications on free will, I must say something about the concept of free will. There is no philosophical (or otherwise) consensus on what free will is; still, there are no available conceptions of free will that are challenged by the literature on priming; therefore my argument does not depend on endorsing a particular conception of free will. One important divide in the literature is between those conceptions of free will according to which an agent acted freely only if she could have done otherwise (*libertarianism*, see, for example, van Inwagen (1983)) and those conceptions according to which whether or not one could have acted freely is independent of whether one could have done otherwise (*compatibilism*, see Frankfurt (1969) and Fisher (1994)), and it rather depends on the

relevant reason to action relationship obtaining (reasons-responsiveness is, for example, what Fisher & Ravizza (1998) call it).

Whichever side of this important divide you happen to fall on, you ought not to think that there are consequences for free will to be found in the priming literature. On the one hand this is because, as I will show (Section 2), the priming literature does not suggest that agents could not have done otherwise when they act under the influence of priming; on the other hand because, as we will also see (Section 3), priming influences do not interfere with the obtaining of the relevant relation between reasons (or intentions) and actions.²

2. Overcoming the priming

There is some evidence that, were for example the fire alarm to go off, the ‘elderly’ stereotype would have no influence at all on how quickly ‘elderly’ subjects would reach the exits as opposed to ‘control’ subjects. In a similar experiment by Macrae and Johnston (1998), subjects were primed with the stereotype of ‘helpfulness’, and then put in a situation in which they could have picked up a pen that the experimenter pretended to accidentally drop. The results match the ‘elderly’ experiment: subjects who had been primed with the ‘helpfulness’ stereotype tended to pick up the pen more often than subjects in the control group. But Macrae and Johnston added an element: sometimes the pen was working fine, and sometimes it was leaking. And they found that when the pen was leaking, there was no registered effect of the ‘helpfulness’ stereotype: primed subjects no longer tended to help more often than control subjects.

This is not meant to deny the link between the elderly stereotype and subjects’ walking pace. It is just a way of showing that the elderly stereotype doesn’t constrain subjects so as to limit their control, leeway, or, indeed, responsibility. Take a case in point: suppose that the hallway in question is quite long, and that at the end of it a

² Here I won’t be able to discuss free will at any length, but I have done so elsewhere: Di Nucci 2010a, 2011a, and 2011b.

primed subject can see a struggling senior. And suppose that the primed subject does not get there in time to help the senior, whose heart failure reaches an irreversible phase. Suppose, finally, that had the primed subject made it to the end of the corridor sooner, that would have greatly increased the chances of saving the senior's life.

We may think that the primed subject is partially to blame for not getting to the senior in time. And I don't think that whether or not we know about the experiment and the stereotype's effect makes any difference on our judgement, nor it ought to make any difference to our ascriptions of responsibility. That is, the stereotype's effect does not justify the subject who didn't get to the senior in time. If the subject was in a position to get there in time, we will understandably blame them for being too late. If, on the other hand, the subject couldn't have made it there in time, we will likely feel as though we cannot in any way blame the subject. And the truth of each conditional will be independent from the stereotype's effect, which is therefore redundant. The only way in which the stereotype might make a difference is if we want to conclude, from the experiment, that whether or not the subject can make it there in time to save the senior depends on whether the subject has been primed. But that would be much too strong a conclusion to draw from the 'elderly' experiment, as Macrae and Johnston's (1998) results suggest.

3. Primed behaviour and intentional action

Subjects are unaware of the effects of the stereotype, and they don't walk slower *on purpose*. That means that we can *make* subjects do something without their knowing it and that we can *make* subjects do something that they don't *mean* to do. Here I show that, nevertheless, they act intentionally. We can see this through standard causal accounts of intentional action. Take Davidson's (1963) original formulation: an agent S has intentionally A-ed only if S had a certain pro attitude (a desire, for example) and a belief that A-ing would satisfy that pro attitude; and this pro attitude plus belief caused S to A.

A point of caution: here I am not equating 'intentional action' with 'free action': indeed, we often act freely without acting intentionally. Indeed, because of the loose connection between intentional action and free will, showing that Bargh's agents do not act intentionally would not show that they are not acting freely. But I don't even think that the claim that Bargh's agents don't *intentionally* walk down the hallway at 4.2 km/h is true. Indeed, it seems to me that if control subjects walk at 4.8 km/h intentionally, then primed subjects must also walk at 4.2 km/h intentionally. Because the same Davidsonian rationalization that applies to the former also applies to the latter. Control subjects, supposedly, walk down the hallway at 4.8 km/h because they have some pro attitude towards leaving the building where the experiment took place within a reasonable time-frame, and a belief that walking down the hallway at that pace will satisfy their pro attitude (they might not know exactly what speed they are walking at, but they will roughly know their pace and whether it will satisfy their goals). Given this rationalization, control subjects walk down the hallway at 4.8 km/h intentionally.

The same goes for primed subjects, since the same rationalization applies. Just like control subjects, primed subjects will also have, supposedly, some pro attitude towards leaving the building within a reasonable time-frame. And, just like control subjects, primed subjects too will have some belief according to which the pace they are walking at satisfies their pro attitude. So, again, since primed subjects walk down the hallway at 4.2 km/h because of the relevant pro attitude plus belief, primed subjects walk down the hallway at 4.2 km/h *intentionally*.

But, it may be objected, primed subjects don't walk at 4.2 km/h because of their pro attitude plus belief; they walk at 4.2 km/h because of the priming effects of the 'elderly' stereotype. Here we might distinguish between causal explanations that are also rational explanations (such as rationalizations, at least according to Davidson), and causal explanations that are not also rational explanations, such as the priming effects of the 'elderly' stereotype. Both the 'pro attitude plus belief' explanation and the 'priming' explanation tell us why the agent walks at 4.2 km/h. But only the former offers a consideration in the light of which the agent walks at 4.2 km/h. Namely, only the former

offers us the agent's reasons for walking at 4.2 km/h. The latter does not: the latter only tells us of one of many causal conditions that have played a role in the agent's walking at 4.2 km/h.³

Had there not been the priming effect, the agent would have not walked at 4.2 km/h down that hallway. But, similarly, had the corridor been different, or had the experiment taken place somewhere else, or had the agent not been selected for the experiment, or had she had different legs, the agent would have not walked at 4.2 km/h down that hallway. These, together with the priming effect, are some of trillions of causal factors that play a role in the agent's walking down the hallway at 4.2 km/h. And, as long as they don't interfere with the agent's plans, their presence within the causal history of that action does not have anything to say about the intentional character of that action.

The effects of the 'elderly' stereotype on the primed group as opposed to the control group can be compared to the difference it would make were one group to have to walk down a corridor with a slight acclivity, as opposed to the other group walking down a perfectly flat corridor. Predictably, the former group would walk slower than the latter. Does that mean that the former group is not free in walking slower than they would have otherwise done on a perfectly flat corridor? I don't think so: had they any reason not to walk at the pace they do, they would increase or decrease their pace.

Take the following case: two comparable groups of students are subject, on consecutive days, to the linguistic test employed in Bargh's experiment, but the test contains no bias (just like with Bargh's control group). When students come out of the experiment room, unbeknownst to them the experimenters measure how long it takes them to walk down the corridor. On the first day the corridor is perfectly flat. On the second day the corridor has been modified so as to present a slight acclivity (one so slight that the first group wouldn't have noticed the difference had they been again, we might suppose). We find that the group which takes the test on the second day walks on average slower down the corridor than the group which takes the test of the first day.

³ Here I can't discuss the causal theory of action in any detail, but see: Di Nucci 2008, 2009, and 2010b.

Has the free will of any member of either group been affected or diminished? I don't think so. Would we find any of these new results challenging, frightening or even interesting? I don't think so. Now I would contend that the reason why these latter results might not be perceived to be as frightening as Bargh's is that we still hang on to a dualism of sorts, so that an intervention upon what is perceived to be the agent's psychology or mind is frightening in a way in which an intervention upon the agent's body or environment isn't.

But that isn't crucial here; what's more important for my argument is that these new sets of experiments are conceptually equivalent to Bargh's; and still nobody would think that these experiments represent a challenge to free will. My point is that priming ranks along indefinite other influences (natural or otherwise) which have some effect on human behaviour: the inclination and consistency of a path will, statistically, make a difference to the average speed of those who walk upon it; similarly will the things that a mind has entertained in the recent past (like priming) and the things that a body has entertained in the recent past: imagine if one group was made to take a longer route into the experiment's room than the other; they might be slightly more tired and they might therefore walk out of the experiment's room slower, on average, than the other group. None of those things has ever bothered us with relation to free will. They ought not to start bothering us now either.

A related point: if one argues that the primed group is unfreely walking slower than they would have otherwise had, one ends up having to conclude, counterintuitively, that the primed group is doing 4.2 km/h unfreely while the control group is doing 4.8 km/h freely. But it looks as though subjects from the primed group have the same relationship with their pace as subjects from the control group: they are satisfied with their pace and they have no reason to change it. Furthermore, as suggested by Macrae and Johnston's (1998) results, were they to have any reason to change their pace, subjects from the primed group would be able to do so as effectively and effortlessly as subjects from the control group: that is, the 'elderly' stereotype would be ineffective.

I have not shown that the ‘elderly’ stereotype does not have an effect on the primed subjects’ walking pace; but I never meant to do that. As I said at the outset, I do accept the basic counterfactual according to which had subjects not been primed, they would have walked faster. What I have shown is that the ‘elderly’ stereotype only has an effect within the subject’s will and reasons. As in, the ‘stereotype’ does not *make* subject do things that they have reason not to do – it does not make them do things unfreely or against their will. Subjects walk out of the room at 4.2 km/h because that satisfies their reasons: it is enough to get home for dinner, say; or it does not interfere with their 5pm appointment. Were 4.2 km/h to interfere in any way with what subjects planned to do or had reason to do or wanted to do, Macrae and Johnston’s (1998) results suggest that subjects would not walk at that pace.⁴

4. *Buridan cases*

In conclusion, there is an interesting philosophical consequence to my argument: the effects of the priming influences that emerge from Bargh’s experiments appear to be reduced to so-called Buridan cases. These are scenarios in which more than one option satisfies the agent and the agent has no reason to choose any particular one of the options which all equally satisfy her. Suppose that you are in Tesco’s, standing in front of multiple rows of identical, undamaged chopped tomato cans. They look all the same; they are all the same price; they are all equally easy to reach. You need a tomato can for your dinner, and therefore you are under rational pressure to buy *a* tomato can; but because *any* tomato can will satisfy your dinner plans, you don’t have any reason to buy any particular tomato can rather than another. While you are under rational pressure to

⁴ It has been suggested to me that another one of Bargh’s (1996) experiments might be more difficult to accommodate for my argument. In this one, subjects who have been primed with a rudeness stereotype are more likely to interrupt the experimenter than control subjects. I think that this experiment is conceptually equivalent to the elderly one: primed subjects are more likely to A than control subjects, where A can be both ‘walking slower than 4.5 km/h’ and ‘interrupt’. The same goes for so-called goal-priming (Bargh et al. 2001; Fitzsimons & Bargh 2003): the experimenters argue that they can even prime goals, but comparing the ‘elderly’ experiment analysed in this paper and the ‘high-achievement’ goal-priming experiment reveals that they are structurally equivalent and that my arguments here on the ‘elderly’ experiment also apply to the ‘high-achievement’ experiment.

buy t_1 v t_2 v t_3 v ... t_n , you have no reason to buy t_1 rather than any other. Effects such as the ‘elderly’ stereotype are restricted to Buridan cases: when you have no reason against the performance ‘suggested’ by the stereotype, the stereotype might have an effect. When you have reason to do something other than what the ‘stereotype’ suggests, the stereotype is not efficacious. Recall the struggling senior’s scenario: since you have reason to rush to the struggling senior, the ‘elderly’ stereotype won’t hold you back, and you will rush there as fast (or as slow) as subjects from the control group – as suggested by Macrae and Johnston’s (1998) results.

The pace of the primed group is just like the tomato cans: 4.2 km/h satisfies primed subjects just as much as 4.3 km/h or 4.8 km/h would have. They have no reason to do 4.2 km/h rather than 4.3 km/h or 4.8 km/h. They have reason to do whatever pace does not interfere with their plans. This suggests that far from being a threat to free will, effects such as the ‘elderly’ stereotype are a potential solution to the Buridan puzzle: how do I choose between many different tomato cans that will all satisfy my plans or between many different speeds that will all satisfy my plans? I don’t need to choose: the automatic and unaware effects of priming will choose for me.⁵

One might object that walking at 4.2 km/h or 4.8 km/h might look like choosing between the 27th and the 43rd tomato can. But that picking up the pen or not picking up the pen is no Buridan alternative: those are two different things. This would be a misunderstanding of Buridan cases: the point of Buridan cases is not that all the different alternatives are *actually* identical; the point is that they are all equal as far as the agent is concerned. So the different tomato cans might be factually different in some important respect, but as long as each tomato can will similarly satisfy the agent’s plans, all tomato cans are equal as far as the agent’s goals are concerned. All tomato cans are, in this sense, *rationally* equal.

⁵ In this paper I didn’t mean to defuse the potential *political* relevance of priming: representative democracies are founded upon partial and limited knowledge. And it is a consequence of partial and limited knowledge that there are plenty of alternatives between which agents have no rational way to choose. In this environment, then, any way of conditioning which of these subjectively equal alternatives an agent chooses is potentially dangerous.

Therefore what I have argued is not that priming effects such as the ‘elderly’ stereotype or the ‘helpfulness’ stereotype are efficacious only in making agents choose one of various factually identical alternatives. What I have argued is that they are efficacious only in making agents choose one of various alternatives that are all equal as far as the agent is concerned because each of them satisfies the agent’s plans. So whether or not the agent has been primed might influence which of two factually different options the agent takes, say helping someone or not helping someone. But, crucially, it will only do so if the agent doesn’t mind whether she helps or doesn’t.

References

- Bargh, J.A. (2006), 'What Have We Been Priming All These Years?', *European Journal of Social Psychology*, 36: 147-168.
- Bargh, J.A. (2008), 'Free Will is Un-natural', in Baer, J., Kaufman, J.C., & Baumeister, R.F. (eds.), *Are We Free?* Oxford University Press.
- Bargh, J.A., Chen, M., & Burrows, L. (1996), 'Automaticity of Social Behavior: Direct effects of trait construct and stereotype activation on action', *Journal of Personality and Social Psychology* 71: 230-244.
- Bargh, J.A. & Chartrand, T.L. (1999), 'The Unbearable Automaticity of Being', *American Psychologist* 54: 462-479.
- Bargh, J.A., Gollwitzer, P.M., Lee-Chai, A., Barndollar, K., & Trötschel, R. (2001), 'The Automated Will: Nonconscious Activation and Pursuit of Behavioral Goals', *Journal of Personality and Social Psychology* 81: 1014-1027.
- Bargh, J.A. & Fitzsimons, G.M. (2003), 'Thinking of You: Nonconscious Pursuit of Interpersonal Goals Associated With Relationship Partners', *Journal of Personality and Social Psychology* 84: 148-164.
- Davidson, D. (1963), 'Actions, Reasons, and Causes', *Journal of Philosophy* 60: 685-700.
- Di Nucci, E. (2008), *Mind Out of Action*. VDM Verlag.
- Di Nucci, E. (2009), 'Simply, false', *Analysis* 69 (1): 69-78.
- Di Nucci, E. (2010), 'Refuting a Frankfurtian Objection to Frankfurt-Type Counterexamples', *Ethical Theory and Moral Practice* 13 (2): 207-213.
- Di Nucci, E. (2010), 'Rational constraints and the Simple View', *Analysis* 70 (3): 481-486.
- Di Nucci, E. (2011), 'Frankfurt counterexample defended', *Analysis* 71 (1): 102-104.

- Di Nucci, E. (2011), 'Frankfurt versus Frankfurt: a new anti-causalist dawn', *Philosophical Explorations* 14 (1): 117-131.
- Fischer, J.M. (1994), *The Metaphysics of Free Will*. Blackwell.
- Fisher, J.M. and Ravizza, M. (1998), *Responsibility and Control*. Cambridge UP.
- Frankfurt, H. (1969), 'Alternate Possibilities and Moral Responsibility', *Journal of Philosophy* 66: 829-839.
- Libet, B. (1985), 'Unconscious cerebral initiative and the role of conscious will in voluntary action', *Behavioural and Brain Science* 8: 529-566.
- Macrae, C.N. & Johnston, L. (1998), 'Help, I Need Somebody: Automatic Action and Inaction', *Social Cognition* 16: 400-417.
- Van Inwagen, P. (1983), *An Essay on Free Will*. Oxford UP.