Minds and Machines (forthcoming)

Habits, Priming, *Aliefs* and the Explanation of Mindless Action

Ezio Di Nucci (University of Copenhagen)

There is a growing body of evidence on the influences of automatic and unconscious processes on our actions. Here I introduce some representative examples of this growing body of evidence, chosen so as to form a diverse group of related *mindless* phenomena: habits, skills, priming and nudges (Section 1). I then argue that this evidence challenges traditional belief-desire-based approaches in the philosophy of action (Sections 2 and 3). I further discuss a recently proposed solution to this challenge, Gendler's *Alief*, finding it wanting (Section 4). I conclude by sketching my own alternative solution, based on the old story of Buridan's ass (Section 5).

1. The empirical case for automaticity

In a fascinating study on eating habits, consumption of fresh and stale popcorn at the cinema by habitual cinema-going popcorn consumers was tested: habitual popcorn consumers ate just as much one-week-old stale popcorn as fresh popcorn (Neal, Wood et. al. 2011). Those without the habit of eating popcorn at the movies ate more fresh popcorn than stale popcorn. Neal, Wood et al. refer to this sort of habitual behaviour as 'automatic', 'non-goal-dependent', and say that it is not under 'intentional control' (which once they refer to as "personal control" (2011: 9)).

Neal, Wood et. al. (2011) found that either an unusual environmental context (eating popcorn in a meeting room instead of a cinema) or a novel way of carrying out the habit (eating with the non-dominant hand) disrupted habitual performance, resulting in even the subjects with the cinema-going popcorn-eating habit eating more fresh than stale popcorn. This is taken to confirm their hypothesis that "habits should not be *activated* automatically outside of their typical performance context and should not be *executed* automatically when responses are performed in novel ways" (2011: 1). This suggests, then, that habits are normally both activated and executed *automatically*. What does it mean that habits are activated and executed automatically? Here is again Wendy Wood: "Cue-response associations are basic to habit performance. By these cognitive associations, habits can be performed automatically, with little thought or effort" (2012: 980).

Habits are not the only kind of automatic performance. In walking down a flight of stairs, you are ill advised to look at your steps: should you do that, you will likely trip. When taking cash from an ATM, you are ill advised to think about your pin, just type it in. These are not the results of experiments in behavioural psychology, rather just some of those things an average person learns in walking this earth; but these could as well have been experimental results: expert golfers are similarly ill advised to take too long to put or concentrate on their swing.

> In an experiment, novices and expert golfers were studied under two conditions: they had either only up to three seconds for each putt or all the time they wanted. Under time pressure, as mentioned, novices performed worse and had fewer target

> hits. Yet surprisingly, experts hit the target more often when they had less time than when they had no time limit. In a second experiment, players were either instructed to pay attention to their swing or distracted by an unrelated, second task (counting tape-recorded tones). When they were asked to pay attention to their swing, as one might expect, novices did better than when they were distracted. Yet with experts, it was again the opposite. When experts concentrated on their swings, their performance decreased; when experts' attention was distracted, their performance actually improved (Gigerenzer 2007: 33; for details on the two experiments see Beilock et al. 2002 & Beilock et al. 2004).

These experiments suggest that more time to think, more attention, and more concentration worsen the performance of experts while they improve the performance of novices. Given the same type of performance, its tokens are importantly different when performed by experts and novices; and this difference seems to depend on the relevant skills that the experts have acquired and that the novices (still) lack. Skills, then, are peculiar in that not only does their exercise run to successful completion without needing thought, attention, or concentration; but the successful completion of an exercise of skill would be hindered by thought, attention, or concentration.[1]

When it comes to our habits and daily routines, we are the experts. The average person is, normally, an expert at climbing stairs and typing her own pin number just like the professional golfer is an expert at putting. The connection between skills on the one hand and habits and routines on the other isn't mysterious: both skills and habits develop through time by practice and repetition, which suggests that mastering a practice has to do with freeing cognitive resources which are no longer necessary (and which can become counterproductive, as we have just seen).

> Habits develop when people give a response repeatedly in a particular context and thereby form associations in memory between the response and recurring context cues (Neal, Wood et al. 2011: 1).

Priming offers us a further example of automatic behaviour. One of the classic examples is walking slower (than a control group) out of the experiment's room after having taken a linguistic test containing 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).[2]

A very similar one, presented in the same seminal paper (Bargh et al. 1996), has subjects primed with a 'rudeness' stereotype who go on to interrupt the experimenter more often than two other groups of subjects (one primed with a 'politeness' stereotype and a control group). As I discuss this case in some detail throughout the paper, I think it is important to

[1] Here – specifically on skills - see also Montero (2016).

[2] A lot of the priming experiments have not passed replications over the last few years, and there is growing talk of the whole priming literature being discredited. Having said that, there have also been recent successful replications. I don't want to minimize this issue, which is obviously very important and all the more so for an article like this one which uses a lot of evidence and examples from the priming literature. But as far as I can see there is no consensus that priming experiments should no longer be taken seriously and I am myself not really in a position to judge on that. For those interested in this issue, there is a very informative article in *The Chronicle of Higher Education*, with references to all the relevant literature including successful and unsuccessful replications: http://chronicle.com/article/Power-of-Suggestion/136907/.

be specific about the content of this experiment. For the rudeness priming the following words were used: "*aggressively, bold, rude, bother, disturb, intrude, annoyingly, interrupt, audaciously, brazen, impolitely, infringe, obnoxious, aggravating,* and *bluntly*" (234). For the politeness priming the following words were used: "*respect, honor, considerate, appreciate, patiently, cordially, yield, polite, cautiously, courteous, graciously, sensitively, discreetly, behaved,* and *unobtrusively*" (234). For the control group the following words were used: "*exercising, flawlessly; occasionally rapidly, gleefully, practiced, optimistically, successfully, normally, send, watches, encourages, gives, clears,* and *prepares*" (234).

There are many other similar cases: "showing suitably primed subjects a picture of a library leads them to speak in quieter tones; showing them an image of an elegant dining room—or exposing them to the smell of soap—leads them to eat more neatly. Subliminal visual priming with an image of an African-featured face leads subjects to respond more aggressively to certain sorts of provocation. Priming subjects with thoughts of their (achievement-oriented) mother leads them to persist longer at word-find tasks; priming them with thoughts of a friend makes them more likely to help a stranger" (Szabo Gendler 2008: 659-660).

The elderly and rudeness stereotypes type of priming belong to what is now in the psychological literature commonly referred to as so-called 'concept priming', as opposed to 'goal priming'. An example of the latter is provided by an experiment where participants are primed with a high-performance goal by having them exposed to the following words: "*win, compete, succeed, strive, attain, achieve,* and *master*. In the neutral priming condition, these words were *ranch, carpet, river, shampoo, robin, hat,* and *window*" (Bargh, Gollwitzer et al. 2001: 1016). Those primed with a high-performance goal did then much better than the control group on a subsequent intellectual task.

Something similar happens with so-called *nudges*:

> Carolyn is the director of food services for a large city school system... One evening, over a good bottle of wine, she and her friend Adam, a statistically oriented management consultant who has worked with supermarket chains, hatched an interesting idea. Without changing any menus, they would run some experiments in her schools to determine whether the way the food is displayed and arranged might influence the choices kids make. Carolyn gave the directors of dozens of school cafeterias specific instructions on how to display the food choices. In some schools the desserts were placed first, in others last, in still others in a separate line. The location of various food items was varied from one school to another. In some schools the French fries, but in others the carrot sticks, were at eye level. From his experience in designing supermarket floor plans, Adam suspected that the results would be dramatic. He was right. Simply by rearranging the cafeteria, Carolyn was able to increase or decrease the consumption of many food items by as much as 25 percent. (Thaler & Sunstein 2008: 1)

In a certain sense, priming effects could be considered the psychological correlate of nudges: they alter behaviour by altering an individual's psychology, rather than the environment. The *choice architecture* (that's how Thaler and Sunstein talk about nudges) of priming is more fundamental, it could be said, because it shapes the *agent* of choice rather than the *object* of choice.[3] On the other hand, just like with the habits experiments that we have just looked at,

[3] This is admittedly no more than a useful simplification – as priming also uses the environment to affect the agent, which is also a description that could be applied to nudges. Thanks to a referee for pointing this out.

nudges appear to affect behaviour by intervening on the environmental cues that trigger the behavioural responses in automatic and habitual actions.[4]

In the next section I introduce traditional approaches in the philosophy of action that, I argue in the section after next, are challenged by the empirical evidence on automaticity just presented.

2. Traditional action theory

To avoid unnecessarily overcomplicating my discussion I will focus throughout the paper only on the experiments on priming and habits, and in particular on the 'interrupting' case and the 'popcorn' case. To interrupt someone else is something that one may do both intentionally and unintentionally. How does traditional action theory account for this difference? Let us look at a traditional account due to Donald Davidson. On Davidson's account some action A is intentional under description *d* only if that action was caused by a primary reason of the agent comprising of a pro attitude towards actions with a certain property, and a belief that action A, under description *d*, has that property.[5]

What Davidson says about 'primary reasons' helps us understand their role for intentional action: "Such a reason gives minimal information: it implies that the action was intentional" (1980: 6); also: "To know a primary reason why someone acted as he did is to know an intention with which the action was done" (1980: 7). The relation between Davidson's 'primary reasons' and *intentions* shows that Davidson's account is the reductive counterpart of the so-called *Simple View* of intentional action: E φ-s intentionally only if E intended to φ: 'For me intentionally to A I must intend to A . . . I will call this the Simple View' (Bratman 1987: 112; on the Simple View see also McCann 1991, 2010, 2011 and Di Nucci 2009, 2010).

What would the traditional account of intentional action say about the case of interrupting in the priming experiment? Well, it certainly does not look as though we could say, on the traditional account, that the experimentee interrupts the experimenter unintentionally. There is no cognitive gap (missing belief or false belief) upon which we could found the judgment that the experimentee does not intentionally interrupt. It is not as if the experimentee does not notice that the experimenter is involved in conversation with someone else.

Here it may be that the popcorn case is not equivalent. One could suggest the following: cinema-goers are intentionally eating popcorn but unintentionally eating *stale* popcorn.[6] This is modelled on Davidson's classic (1980: 84-85) example of intentionally boarding a plane headed to 'London' but unintentionally boarding a plane headed to London, Ontario. Actions can be intentional under one description but unintentional under a different one: mistakes are a paradigmatic example. But that's also the problem with this interpretation: it is not as if those cinema-goers can say that they meant to eat fresh popcorn instead of stale popcorn the way in which Davidson's character can say that she meant to board a plane headed to

[4] The idea behind nudges is for political institutions to appropriate old and well-known marketing tools. The following experiment may provide a good illustration of how the private sector has been using 'nudges' all along: Iyengar & Lepper (2000) presented shoppers with two tasting booths with either six or 24 different varieties of jam. While more shoppers stopped at the booth with more varieties of jam, more shoppers bought from the booth with less varieties of jam (see also Gigerenzer 2007: 31). This experiment too appears to confirm an often heard commonplace according to which too many alternatives make choosing difficult.

[5] Note that Davidson offers only necessary conditions to avoid the problem of deviant causal chains.

[6] This interpretation is available whichever account of the individuation of action one endorses.

London, England instead of a plane headed to London, Ontario. Why not? Davidson's character can say the following: had I known that this plane was headed to London, Ontario I would not have boarded it. Can't the cinema-goers similarly say that had they known that the popcorn were stale they would not have eaten it?

There is an important parallel between the two cases, but also an important difference. The parallel is that, in both cases, the agent could stop at any time boarding/eating. She is neither being carried forcibly onto the plane in the one scenario nor is she being force-fed in the other scenario. But while there is a bit of information that would give Davidson's character reason to stop boarding the plane (the fact that the plane is headed to London, Ontario) there is no symmetrical bit of information that should lead the cinema-goers to stop eating. The fact that the popcorn is a week old is not the relevant bit of information; it is not as if the popcorn had been secretly poisoned. That would be a parallel case, in which the cinema-goers have an external reason to stop eating (that the popcorn has been poisoned) but no internal reason to stop eating because they have no access to the relevant secret, just as Davidson's character has no access to the secret that the plane is headed to London, Ontario. Cinema-goers have access to all the information they need in that they can *taste* their popcorn. Their situation is more like the one in which someone would point to Davidson's character that her plane is actually headed to London, Ontario for her to reply that any London will do and board the plane. London, England would have been better, but London, Ontario will do too. Fresh popcorn would have been better, but stale popcorn will do too.

Both interrupting the experimenter and eating stale popcorn do not count as unintentional actions on traditional action theories. Does that mean that these actions are intentional? The short answer is yes; the long answer is that there are action-descriptions of what the experimentees do in these experiments that would not count as intentional, but that does not mean anything because there are always alternative action-descriptions which are not intentional. So we could say that on traditional accounts 'eating one-week-old' popcorn is not intentional because the agent does not know that the popcorn are one week old; but that's the same sense in which we can say that, in going for a walk, I did not intentionally take 8743 steps, nor did I intentionally walk down 'Forest Road', just because I didn't know about these properties of my actions: that does not make my action of taking a walk unintentional, just as not knowing that the popcorn are one week old does not make my eating them unintentional.

Indeed, with the priming cases it is even more difficult to argue that they are unintentional, as there aren't any easily available unintentional descriptions: one can say, at the most, that the experimentees in the rude group did not intentionally 'interrupt more often than those in the polite group', or that the experimentees in the 'old stereotype' group did not intentionally 'walk slower than those in the control group'; this, again, does not say anything about the intentionality of walking or interrupting.

Traditional action theories hold that agents' behaviour in these experiments is intentional; that, by the way, seems to also be the observer's intuition and the intuitions that agents themselves also have when in the debriefing they show no awareness of the role of the prime.[7] In the next section I argue that the empirical data provides a challenge to traditional action theory.

[7] Here I should mention that one may take evidence on so-called 'explanatory vacuum' to speak against the idea that the experimentees take themselves to be acting intentionally. Experimenters have recorded some

3. The experiments' challenge

We must distinguish between the claim that a combination of psychological states such as desires, beliefs, and intentions (with a specific content) is necessary for intentional action and the claim that such a combination of psychological states is sufficient for intentional action. Normally, this distinction is important with respect to the problem of deviant causal chains, where performances that are intuitively unintentional and accidental meet the conditions for intentional action. A large literature on the problem of deviance tries therefore to deal with these counterexamples in order for traditional action theories to be able to offer psychological states as sufficient conditions.[8]

But this is not our concern here: our concern in distinguishing between the necessity of psychological states and their sufficiency is that if psychological states are sufficient for intentional action, and if the examples in question are indeed intentional actions as we have argued in the previous section, then influences such as priming should either not manifest themselves in the agent's behaviour or they should manifest themselves in the agent's psychological states. This disjunction is obviously rhetorical in so far as the first disjunct is contradicted by the experimental evidence under discussion.

We cannot ask action theory to provide an empirical explanation of priming effects; still, these effects nevertheless present action theory with a challenge. According to traditional action theory some combination of desire, belief, or intention is sufficient for intentional action. Which part of this combination of psychological states accounts for the effects of priming? We can't ask an account of intentional action to explain priming effects, but we can ask an account of intentional action to rationalize those intentional actions that subjects perform after having been primed. The latter is a philosophically legitimate request to traditional accounts of intentional action, such that if these accounts cannot satisfy this request, then priming speaks against the accounts in question.

The following action-description offers a true description of what agents do after having been primed: "S interrupted rather than not interrupting". This action-description, as I have argued in the previous section, is an intentional action-description or, for short, intentional action (as intentionality, as we have also shown in the previous section, is a property of action-descriptions). What this means is that traditional accounts of intentional action can be legitimately expected to rationalize "S interrupted rather than not interrupting". If they fail to do so, then "S interrupted rather than not interrupting" will constitute a counterexample to those accounts of intentional action that fail to rationalize it.

---

agitation in subjects that have been primed which they have not recorded in subjects that they have consciously induced to behave in a comparable manner (Oettingen et al. 2006, Parks-Stamm et al. 2010). To stay with one of the examples I discuss here, the idea would be that those who have been primed to walk slower are afterwards agitated while those who have been explicitly warned to walk slower because, say, the floor in the corridor is wet, do not show similar levels of agitation. I won't discuss this issue in any detail here, but one could pursue this evidence in an attempt to argue that subjects do not experience themselves as acting intentionally in priming cases. But the explanatory vacuum data also has a much more important role to play for my argument: that agents are agitated speaks against the attribution of the relevant psychological states: why would people experience negative affect about what they do after having been primed if they were motivated to do it? To put it another way: if they had the relevant psychological states motivating them to act as they do, then what would explain the experienced negative affect?

[8] For a good introduction to this debate see Stout 2010.

Let us put this point in another way, which I take to be equivalent but helps illustrate my argument further: if some factor X can make a difference as to whether the agent does A or does not do A, and traditional action theory explains the agent's doing A instead of not doing A in terms of the agent's psychological states and only her psychological states (sufficiency), then factor X must be found somewhere among the agent's psychological states – obviously this applies only to those Xs which *can* be found among psychological states (Xs which we can psychologically represent). And if factor X cannot be found anywhere amongst the agent's psychological states, then that shows that the agent's psychological states are not sufficient to explain why the agent does A instead of not doing A.

Here we should probably distinguish between different cases amongst the experimental evidence: what we just said, for example, seems to apply to the interrupting case but not to the walking slower case. Whether or not one interrupts seems to be different from whether one walks at 4,8km/h or at 4,5km/h. The latter seems equivalent to changing the acclivity of the floor under which agents routinely walk and record how the agent's walking speed will adapt to the new acclivity. If you are subtle enough, the change will go unnoticed.

Still, even for the walking case we can say something similar to what we said for the interrupting case: if something the agent is unaware of can change the agent's behaviour, then the agent's behaviour cannot be purely conscious behaviour. If it were purely conscious behaviour, it would not be changed by factors the agent is not aware of.[9] The difference with the interrupting case and the popcorn case is, though, that while we may demand that a reasons explanation explain us why the agent interrupted instead of not interrupting and why the agent eat popcorn instead of not eating popcorn, it may be too much to demand of a reasons explanation to tell us why the agent walked at 4,8km/h instead of 4,5km/h.

Taking the experimental evidence seriously means, I think, accepting that the only difference between those experimentees that interrupted and those that did not interrupt was the priming and its effects; similarly, it means accepting that the only difference between those who ate stale popcorn and those who didn't was their habits and their effects. Now, obviously to accept that the only differences are the priming and its effects in the one case and the habit and its effects in the other case is not the same as to accept that there is no difference in psychological states, otherwise the argument would already be settled from the start. The question which remains to be answered, then, is whether the differences constituted by priming and habits can be captured in terms of psychological states.

Let us start from the difference between those who ate stale popcorn and those who did not eat stale popcorn. An explanation in terms of psychological states would appeal to things like the belief, in one case, that the popcorn was stale against the lack of such belief in the other case. The idea would be that both agents or groups of agents had a pro attitude towards eating popcorn but one agent or group of agents also had the belief that the popcorn was stale which explains why this agent or group did not eat the stale popcorn. As the other agent or group did not have the relevant belief, they ate the stale popcorn. There are two problems with this *belief* approach: firstly, it flies in the face of what participants said afterwards about the popcorn. Namely, participants across conditions in both studies reported liking the fresh popcorn and disliking the stale popcorn (2011:5 for study 1 and 2011: 7 for study 2). So it is not as if they did not notice the difference. The cognitive relation of all participants across conditions to the popcorn did not vary with whether or not they had the habit. Secondly, if what explains why an agent or group ate the stale

[9] At the end of this section I discuss the possibility that the psychological states which explain those behaviours are unconscious states of the agent.

popcorn while the other agent or group didn't was the presence or absence of a belief about the popcorn being stale, then the eating patterns should not be sensitive to conditions such as location or method of eating – but the experiments show that they are.

The same argument would apply to trying to explain the difference between those who eat stale popcorn and those who don't in terms of their desires or other pro attitudes; if the desires of the agents would explain why they eat stale popcorn rather than not, then their eating patterns should be sensitive to a change in pro attitude rather than being sensitive, as the experiments show, to changes in environmental cues.

It could be objected to the arguments above that changes in the environmental cues can be themselves explained in terms of the agent's psychological states, in so far as a change of location is likely to result in a change of the agent's relevant psychological states. One could just say that, for example, some environments trigger (maybe even automatically) in some people a desire or an intention to eat popcorn (the cinema) while some other environments do not trigger in the same people a desire or an intention to eat popcorn (the meeting room). This is plausible: places are associated to smells, they are associated to memories, and these things may contribute to generate the relevant psychological states. But this, again, does not explain the experimental data across all conditions: it does not explain, for example, why an agent or group would eat stale popcorn when eating with her dominant hand but not eat stale popcorn when eating with her non-dominant hand. What is the difference in terms of psychological states between the habitual cinema-goer who eats with her dominant hand and the habitual cinema-goer who eats with her non-dominant hand? The explanation in terms of psychological states, here, looks to come to a quick end.

There is also another aspect of the empirical evidence which speaks against explaining what agents do in terms of their desire to eat popcorn: namely, on average and across conditions, agents didn't seem to like the popcorn, whether stale or fresh. On a liking scale 1 to 7, subjects did like the fresh popcorn more than the stale popcorn, but both fresh popcorn and stale popcorn ranked below the middle point of the scale. Obviously, liking something and desiring something are two different attitudes; but, as desire in the action theory literature is normally a place-holder for a wide range of attitudes that Davidson classically referred to as 'pro-attitudes', it seems fair to look at whether subjects liked the popcorn, as liking does seem to be a pro-attitude. What we can extrapolate from the data, then, is that the average attitude to popcorn of subjects across conditions was rather negative – which in turn speaks against explaining what subjects do in terms of a *pro*-attitude towards the popcorn.

We have seen that the explanation of habitual behaviour in terms of the agent's psychological states of belief, desire and intention fails. What about the cases of priming? I think that here there may be some important differences to habits: take the case of the 'rudeness' priming and consequent interrupting of the experimenter. Here it does not seem implausible to try to say that, without their realising it, the 'rudeness' scrambled sentence task may change subjects' mood, make them more aggressive or less patient, and these attitudes would in turn generate the relevant psychological states which explain why subjects tend to interrupt the experimenter.

The idea, then, is that the content of the scrambled sentence task would, directly or indirectly, provoke the relevant psychological states in the agent. Take the words of the rudeness priming again: "*aggressively, bold, rude, bother, disturb, intrude, annoyingly, interrupt, audaciously, brazen, impolitely, infringe, obnoxious, aggravating,* and *bluntly*" (234). Is it plausible to say that working for a while with these words may make one more aggressive? Maybe: but is being more aggressive a matter of the agent's intentional psychological states of desire,

belief, and intention? Same goes for moods in general and patience in particular. Here we should say that the claim does not need to be that the priming words provoke the psychological states directly nor does it need to be that growing more aggressive is a matter of change in one's psychological states. The story could rather be that those words make one more aggressive (or less patient) and then, having grown more aggressive, the agent would develop psychological states that fit her new mood: the agent would, for example, be disposed to develop a desire to interrupt if made to wait.

I think this is a plausible story, with three important caveats: (1) that such a story can only be told by appealing to unconscious psychological states; (2) that the story may work for psychological states such as some affective pro attitudes but not work for the cognitive psychological state of belief; (3) the explanatory vacuum evidence speaks against the attribution of the relevant desires. Let us look at these three points in reverse order.

(3) As I mentioned already in the previous section, there is some evidence that agents who have been primed experience negative affect or agitation after the action while agents who have been consciously and explicitly given the goal and then perform the same action do not experience negative affect or agitation (Oettingen et al. 2006, Parks-Stamm et al. 2010). The hypothesis in the studies on the explanatory vacuum is that agents experience negative affect or agitation because they are at a loss as to why they behaved as they did: they are uncomfortable with what they just did because they don't know where it came from – as opposed to when they are given the same goal consciously and explicitly and can then point to that goal to fill the 'vacuum'. Here the idea is that if subjects did indeed have the relevant desires and beliefs which represented their performance, then they would not find themselves in an explanatory vacuum; and that therefore the explanatory vacuum speaks against the attribution of the relevant psychological states. Here two things could be objected: firstly, that the explanatory vacuum data speaks against the intentional character of the performances in question, but as I have dealt with the issue of intentionality in the previous section I won't repeat those arguments here. Secondly, one could argue that what explains the 'vacuum' is not the absence of the relevant psychological states but rather the fact that the agent is not aware of those psychological states. I discuss this possibility below under caveat number (1).

(2) If the strategy described in this section to explain action affected by priming effects in terms of the agent's psychological states needs to be vindicated, then it must be a plausible story not just in the one case of the rudeness stereotype but rather across the myriads of cases of priming; and this means that the proposal according to which the priming would, directly or indirectly, provoke unconscious psychological states needs to be applicable not just to desires but also to beliefs. Already if you take the case of walking slower out of the experiment's room it looks harder to cash it out in terms of an unconscious desire: can we suppose that agents, as a result of having being subjected to the elderly priming, feel tired? And even if we said that, would being tired provoke a desire to walk slowly? Walking slowly as a result of tiredness seems, whether or not it is related to priming, a difficult case for the belief-desire model. But, again, it is not about the individual cases, given the diversity of priming cases. Another difficult one to account for in terms of desire is the one where agents lower the tone of their voice upon being shown the picture of a library (Aarts & Dijksterhuis 2003). The general idea, then, is that the defender of the belief-desire model needs to be able to extend the proposal we discussed for desires to the psychological state of belief too.

What's the problem with extending this approach to belief too? The problem is that it does not look plausible to think that a truth-insensitive process such as priming could possibly

generate beliefs, albeit unconscious ones, if we take beliefs to be cognitive states aimed at the truth who commit an agent to the truth of a proposition (for a good discussion of this issue see Szabo Gendler 2008: 657-658, especially footnote 47 & 48). What kind of information about the way the world actually is could one get from the priming? Yes, one could get a bunch of false beliefs, but that's not really the point (on false beliefs, see the end of this section). The point is rather whether they could qualify as beliefs at all. Say we supposed that the rudeness stereotype somehow initiated a sort of ego-trip where subjects became convinced that other people ought to treat them with utmost respect and deference. The problem here is not the plausibility of this hypothesis or whether it could explain the data: the problem is rather than going on such an ego-trip would hardly count as subjects having changed their beliefs or formed some new beliefs. Take, alternatively, the library case. One may attribute subjects an unconscious belief that they are in a library. And, again, that such a belief would be false isn't important: what is important is how the belief would have been generated, by a mechanism which is insensitive to the truth. Also, what role would the belief play with respect to other psychological states of the agent? The agent, it seems, does not really think she is in a library rather than in an experiment's room, otherwise she would hardly comply further with the experimenter's requests. The problem here seems to be that the attribution of the relevant belief would fit one of its functional roles by, for example, explaining one of the agent's actions but the attribution would also conflict with other functional roles: other actions and other psychological states of that agent.[10]

A good example of the above problem with the attribution of the relevant beliefs to primed agents is provided by another famous priming experiment (also in Bargh et al. 1996): experimentees were subliminally primed by being shown either African-American faces or Caucasian faces. Those who had been primed with African-American faces then reacted with more hostility to a presumed technical failure after which they would have had to repeat the whole computerized visual task they had been asked to work at. One could try to explain the behaviour of those primed with African-American faces by attributing to them stereotypical racial attitudes that had been awakened or otherwise provoked by the priming. But should those attitudes count as belief? Consider what these subjects would say and do before and after the experiment: that would no longer distinguish them from those in the other group to which we do not attribute racial attitudes. Again, it looks as though those attitudes, also because they would not resist the test of time, should not count as beliefs (see Schwitzgebel 2010a for a discussion of this issue: his position is that those are 'in-between' cases, where we should neither ascribe the relevant belief nor deny that the agent has the relevant belief).

(1) Let us now go back to the first caveat, namely that the relevant psychological states would have to be unconscious. As across the priming cases experimentees are not aware of the priming or of its effects, psychological states would have to be unconscious. Defenders of the belief-desire model traditionally allow for unconscious psychological states of sorts. Here's Davidson:

> We cannot suppose that whenever an agent acts intentionally he goes through a process of deliberation or reasoning, marshals evidence and principles, and draws conclusions. Nevertheless, if someone acts with an intention, he must have attitudes and beliefs from which, had he been aware of them and had the time, he could have

[10] There is a growing literature on whether the attitudes which are the result of priming could possibly be considered to be *beliefs*. Some examples: Wilson, Lindsey, and Schooler 2000; Rowbottom 2007; Zimmerman 2007; Gendler 2008a & 2008b; Hunter, (2009); Sommers (2009); Schwitzgebel 2010a & 2010b; Gertler 2011.

> reasoned that his action was desirable (or had some other positive attribute). (1978: 85)

Here are Mele and Moser:

> A state of having a plan can, however, be dispositional in the way belief is. It can exist even while unmanifested in an event of planning, and even while absent from awareness or consciousness. Laura need not be constantly aware of her plan or continually engaging in acts of planning to flip the switch. (1994: 227).

I don't think it is clear that either Davidson or Mele and Moser are talking about unconscious psychological states in the sense required by the priming cases. To begin with, Davidson seems to be talking only about the fact that a belief-desire model does not need to suppose that agents actually go through the deliberative process of reasoning that the belief-desire model reconstructs and attributes to them. But that is not the same as saying that agents need not be aware of their reasons for action. Still, Davidson does go on to imply exactly that: "he must have attitudes and beliefs from which, had he been aware of them and had the time, he could have reasoned that his action was desirable". So agents must have the relevant attitudes and beliefs but they need not be aware of those.

But what does it mean that agents need not be aware of their beliefs and desires? It could mean, minimally, that those beliefs and desires need not be occurrent. That is indeed what Mele and Moser suggest in saying that plans and beliefs can be dispositional and that agents need not be *constantly* aware of them. The distinction between occurrent and dispositional psychological states is a well-established one and therefore, whatever its value (see below), I think it is legitimate for defenders of the belief-desire model to appeal to this distinction.[11] I say 'whatever its value' because I think there is something strange in appealing to the occurrent-dispositional distinction in the case of action explanation: namely, the idea of dispositional beliefs and dispositional desires is at its most plausible in avoiding overclogging the conscious mind with psychological states that are, most of the time, at best in the background as they are, most of the time, irrelevant. Most of our beliefs are, at any given time, irrelevant to the current situation – but that, plausibly, does not show that, at that time, we do not believe the irrelevant proposition in question. Having said that, the case of action is different: at the time of action, the beliefs that are supposed to motivate us to act are, by definition, not irrelevant: they are motivating us! (Let alone causing our movements!)

Still, I think there is a more serious problem with the 'dispositional' suggestion above: the sense in which the psychological states in the priming cases are *unconscious* seems to be different from the sense in which *dispositional* psychological states can motivate us and cause our actions. The difference, in short, is *access*: Ned Block (1995) famously distinguishes between *access* consciousness and *phenomenal* consciousness, where a state can be conscious without having any quality or phenomenal feel to it. Accordingly, we need to distinguish between access-unconscious psychological states and phenomenally unconscious psychological states. It seems to me that Davidson and Mele and Moser are talking about phenomenally unconscious psychological states; and that indeed, in general, dispositional psychological states are characterised by not having a quality or phenomenal feel because they are not occurrent, but dispositional psychological states must be access conscious – in the sense in which I have access to my belief that my son was born in April without thinking about it all the time. The belief that my son was born in April has a quality to it when it is

[11] Some influential discussions of this distinction: Price (1969), Armstrong (1973), Lycan (1986), Searle (1992), and Audi (1994).

occurrent, but it is mostly dispositional: still, I can at any time recall it at will just as many environmental cues will activate it. There is nothing mysterious about it.

The problem, in short, is that the putative psychological states which we would attribute in priming cases would not be access conscious, and would therefore be unconscious psychological states in the strongest sense of unconscious: psychological states which are not available to the agent – indeed, as in the case of racist attitudes, psychological states that the agent would explicitly dispute.

3.1 Further problems with traditional rationalization

In this sub-section I present two further arguments against rationalizing what agents do after having been primed in terms of their psychological states of belief, desire, or intention: (a) the first argument has to do with some further experimental evidence from the priming literature while (b) the second argument is about false beliefs.

(a) There is further reason why I think that priming effects cannot be properly captured in terms of psychological states: agents' responsiveness to priming does not appear to be as strong as their responsiveness to reasons. A replication of the priming experiments illustrates this point: in a priming 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.

If one were to explain both subject's helpful behaviour and their reaction to the leaking pen in terms of psychological states, then it would not be apparent why the reason provided by the leaking pen would completely silence priming effects: there should rather be a dilemma-like tension between competing reasons which one would also expect reflected in the data. What this experiment rather suggests is that agents are considerably less responsive to priming effects than they are to reasons, suggesting that if indeed priming effects must be cashed out in terms of psychological attitudes, those attitudes must be different from the psychological states of belief, desire and intention.[12]

As I have at times pointed to relevant differences between the various priming cases, let me say that the effect measured by Macrae and Johnston does not depend on any particular features of the scenario they set up but can rather be applied across the different scenarios: we can, for example, imagine that the polite concept priming would not have had any effect on someone who was in a rush to get home to relieve the babysitter.

(b) Here is one final problem with explaining the evidence in terms of beliefs. Let us put to the side the previous arguments and suppose that one could explain the evidence in terms of the agent's conscious or unconscious beliefs. I think that would still be problematic, because

[12] Here one could suggest that conscious reasons override unconscious reasons. But, firstly, that's not at all obvious and indeed a lot of the psychological literature speaks against it; secondly, the evidence seems to point to something stronger than 'overriding': the priming effect *disappears*.

it would lead to a verdict of irrationality that does not fit the cases. If we were going to explain the interrupting case with some unconscious belief of the agent that she is rude, or the walking slower case with some unconscious belief of the agent that she is old, then we would be attributing those agents false beliefs about being rude or being old – at least false across the group of experimentees even though not necessarily false in each individual case. Now, in the presence of false beliefs we normally have the kind of case that we described in the previous section with London, Ontario. An observer would be puzzled as to why someone who wants to go to London, England is boarding a plane to London, Ontario – taking that behaviour to be prima facie irrational, given the agent's preferences. Upon finding out that the agent believes that the planed headed to London, Ontario is actually headed to London, England, the observer would then realise that the behaviour in question is not irrational. This is what normally happens with false beliefs. They would yield, at least from the external perspective, a verdict of irrationality. But that's not the case with our habitual and priming cases: in those cases there is no irrationality, not even from the external perspective. An observer would not judge the agents' walking slower or interrupting or eating stale popcorn as irrational.[13] And this speaks against the attribution of false beliefs independently of the other problems with explanation in terms of psychological states that we have already discussed.

Summing up, I have presented the challenge that experiments on habits and priming constitute for belief-desire theories of action. I have shown that, despite the possibility of attributing the relevant psychological states in some cases, the belief-desire model cannot be plausibly taken to explain all cases of habitual behaviour and primed behaviour. In the next section I present an alternative solution, so-called *aliefs*, that has been recently been taken to be able to account at least for priming cases: I find aliefs, at least in their application to the priming literature, unsatisfactory. In the final section, then, I sketch my own approach.

4. Aliefs and priming

In two influential articles, Tamar Szabo Gendler has recently argued for a new psychological state, *alief* (2008a & 2008b). Gendler thinks that *aliefs* can explain a lot of difficult cases that cannot be explained simply by appeal to beliefs, including priming cases: indeed, she makes explicit reference to some of the cases that I discuss in this paper, such as for example the interrupting case. Let me clarify that the success of my argument in the previous two sections does not depend on my discussion Gendler's aliefs here.

How can aliefs explain priming effects such as interrupting, walking slowly, or being hostile? Here is Gendler's answer to the interrupting case:

> ...what Bargh and his colleagues have done, I want to argue, is to induce in their different sets of subjects different sorts of occurrent alief. As the result of the pre- or quasi-conscious activation of the cluster of affective tendencies and behavioral repertoires associated with the notion of rudeness, subjects in the third condition find themselves more likely to act in ways that they would act in the presence of rudeness; as the result of the pre- or quasi-conscious activation of the cluster of affective tendencies and behavioral repertoires associated with the notion of politeness, subjects in the second condition find themselves more likely to act in ways that they would act in the presence of politeness. (2008: 658)

[13] In this respect the library case may constitute an exception.

Gendler gives a similar answer to the elderly case, and this time she formulates the content of the alief:

> Bargh's elderly-primed subjects occurently alieved below the level of conscious awareness something like: "Old. Tired. Be careful walking to that elevator…"—and the activation of this behavioral repertoire made them more likely to act in accord with it. (2008: 659)

The suggestion, then, is that subjects in priming experiments identify with the content of the priming, but only at the alief level: so that they *alieve* that they are old, they *alieve* that they are rude or polite, they alieve that they are go-getters, and they presumably *alieve* that they are African-American; even though subjects presumably do not *alieve* that they are a library, they rather alieve that they are *in* a library.

This is a promising notion and one that one could try to apply to the explanation of the habits experiments too. "Popcorn. Cinema. Tasty! Eat as much as possible!" The idea would be that habits are associated to particular aliefs which activate automatic reactions; when one is not acting habitually (as in the non-dominant hand case), then in the absence of the relevant alief the corresponding automatic reaction is not activated and the agent realises that the popcorn isn't actually that nice, this time. So aliefs, differently from the other psychological states that we have looked at, have the advantage of being more generally applicable to the different kinds of experimental data. Aliefs would also sort out the problem with which we concluded the previous section, where agents appear to be less responsive to priming effects than to reasons: one may think that agents are less responsive to aliefs than they are to beliefs, which would also fit Gendler's paradigmatic cases such as the Skywalk: I may hesitate, sweat, trip and behave immaturely, but I will get to the other hand, because my beliefs trump my aliefs, supposedly.

I now present three problems with the notion of alief: (A) a problem with the possible application of the concept of alief to the experimental data that we have been discussing in this paper; (B) a more general problem with the nature of alief; and (C) a third problem which has to do with the status of alief as arational.

(A) The first problem is that it is not clear to me that the paradigmatic cases of alief such as Skywalk are comparable to priming cases such as the leaking pen: in the Skywalk case the effect of the alief is maybe trumped, but not completely silenced – while in the leaking pen case there is no priming effect. The vertigo-related behaviour in the Skywalk case is exactly what is belief-discordant and therefore it is exactly that part of the agent's behaviour (rather than making it to the other end, say) which motivates the notion of alief. But in the leaking pen case there is no parallel behaviour: there is no recorded hesitation or doubt as a result of the 'helpful' conscience which has just been primed (as in, the data do not show any effect of the priming).

(B) The second problem with alief is that once we give it enough content so that it can explain the data, then it becomes a propositional attitude like belief. Here is a case in point presented by Mandelbaum (2012), where it looks as though alief is a propositional attitude with inferential content. Mandelbaum (2012: 10-11) presents the experimental case of George Clooney fans who are willing to pay money (lots of money) for a bandana worn by Clooney. Now, it turns out those fans are not willing to pay as much if they are told that the bandana has been washed since Clooney wore it (Newman et al. 2011). The idea would be that those fans alieve that the bandana contains Clooney's essence, but this alief interacts, inferentially, with beliefs of the agent (such as the belief that the bandana has been washed),

so that agents conclude that the bandana, after having been washed, will no longer contain Clooney's essence and is therefore no longer worth as much. On the one hand, then, it seems that attitudes towards celebrities and their belongings are exactly the kind of attitudes that alief accounts for – the subjects in the above study won't admit to believe any such nonsense; but on the other hand it looks as though alief is propositional in that it is available to interact inferentially with other propositional states of the agent – as illustrated by this case. But then what difference is left between aliefs and beliefs?

(C) Finally, the third problem is related to my remarks about false beliefs in the previous section. Aliefs, according to Gendler, are arational: but what aliefs would be meant to explain, in cases such as habitual behaviour and priming, are perfectly rational actions of the agent: walking slower is perfectly rational, interrupting is perfectly rational, and eating the popcorn is also perfectly rational. So the explanation in terms of aliefs would give the wrong verdict. Let me clarify that here the problem would not be an arational state such alief giving rise to a perfectly rational action; there may be no tension there; still, if the alief is arational, in virtue of what are we going to rationalize the perfectly rational actions in question? Not in virtue of aliefs, anyway.

In this respect, the cases of habits and priming that we have been looking at are different from the Skywalk case and similar cases that motivated Gendler's introduction of alief.[14] In the next and final section I sketch my own proposal as to how to deal with the experimental data, which builds on the old medieval story of Buridan's ass.

5. Buridan, Indifference, and Rationality

In this final section I make a very preliminary start at putting forward my own view on mindless phenomena (for more, please see Di Nucci 2013). It's not just that I think that automaticity and habitual behaviour cannot be plausibly and comprehensively explained in terms of the agent's psychological states of belief, desire or intention – as I have been arguing throughout this paper. I also think that it is to our evolutionary advantage that there are mechanisms other than our propositional attitudes controlling action in the automatic cases of the priming literature or the cases of habitual and skilled actions.

We are neither bothered enough nor knowledgeable enough for our psychological states to explain our automatic, habitual and skilled behaviour. My idea is quite simple: the priming data shows that our psychological states have control over things such as whether to walk towards the elevator or in the other direction; but, normally, our psychological states don't control things such as whether to walk at 4,2km/h (the speed at which the primed group walks) or at 4,8km/h (the speed of the control group). Similarly, our psychological states control things such as whether to go to the cinema or to the Opera, but not, normally, which popcorn to eat or how much of it. Normally, there are a lot of options that we are either not bothered about or not in a position to differentiate between – and then it is a blessing that our sub-personal automatic mechanisms and processes help us out.

This is, by the way, the point of nudges: they turn potentially damaging features of persons such as indifference and ignorance to the advantage of the persons themselves and society at large. That people are indifferent and ignorant about vegetables and their dietary properties is potentially a serious problem for the subject's health at the personal level and also a social and political problem. But people's indifference and ignorance can be turned to their

[14] There may be exceptions here, such as, again, the library case.

advantage and to the advantage of society, it turns out, by simply *moving* the vegetables. And, importantly for the arguments in this paper, this is only possible because of the influence of automaticity and habits over a person's agency and life: otherwise people wouldn't start to suddenly eat more vegetables just because they have been moved to the front.

I think going back to the old story of Buridan's ass helps illustrate my point here: 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. To be clear: I am no interested in the actual story of two identical alternatives; but in the idea behind the story that the two alternatives, whatever their actual properties, equally satisfy the chooser so that they are rationally identical (in terms of the chooser's preferences) if not factually so. This clarification is important because it shows that potentially this can be applied to very different alternatives including those, for example, which are relevantly different in terms of the agent's welfare (healthy vs. unhealthy, say).[15]

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 one particular tomato can rather than any other one. So while you are under rational pressure to buy t1 v t2 v t3 v … tn (that is: the extremely long exclusive disjunctive that includes all the tomato cans), you have no reason to buy t1 rather than t2 v t3 v … tn. The puzzle is: how do you choose?

Now we can see how 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 Macrae and Johnston's (1998) results: when you have a reason not to help – namely avoiding the leaking pen, then the priming won't have any effect. 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 not to do 4.2 km/h, as much as 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, and have no reason to do any particular one of the different speeds that accord with their plans. If 4.8 km/h would have been more beneficial to their plans, primed subjects would not have been slowed down by the 'elderly' stereotype which, therefore, has an effect only between different options that all satisfy the agent's plans – that is, Buridan cases.

This, then, suggests that far from being a threat, 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 or the sub-personal mechanisms developed through habituation will choose for me.

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 does not look like a Buridan alternative: those are two very different things. But 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

[15] Thanks to a referee for pressing me on this point.

concerned. So the different tomato cans might be factually different in some important respects, 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.

This is the point about indifference and knowledge with which I started: many two aspects of the world are not different enough from one another for us to notice, either because of our cognitive limitations or because of our relative affective indifference. 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 very 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.

This is also why I don't think that this sort of picking rather than choosing is irrational. Indeed, I don't even think that it is arational or non-rational, as it is sometimes suggested (see Gendler 2008, for example). Bernard Gert seems to go in a similar direction, when he says that for an action to be rational, it suffices that it is not irrational. "Defining a rational action simply as an action that is not irrational does not impose a fictitious and misleading uniformity on all rational actions" (1998, p. 61). Sometimes picking is, indeed, perfectly rational: as Buridan's ass teaches us, it is insisting on choosing that would be irrational.

References

Aarts H. & Dijksterhuis A. (2003). The Silence of the Library. Journal of Personality and Social Psychology. Vol. 84, No. 1, 18–28.
Armstrong D.M. (1973). Belief, truth, and knowledge (Cambridge: Cambridge UP).
Audi R. (1994). "Dispositional beliefs and dispositions to believe", Nous, 28, 419–434.
Bargh J.A., Chen M. & Burrows L. (1996). Automaticity of Social Behavior. Journal of Personality and Social Psychology Vol. 71, No. 2, 230-244.
Bargh, J.A., Gollwitzer P.M., Lee-Chai A., Barndollar K., & Trötschel R. (2001). The Automated Will. Journal of Personality and Social Psychology. Vol. 81. No. 6. 1014-1027.
Beilock S.L., Carr T.H., MacMahon C., & Starkes J.L. (2002). When paying attention becomes counterproductive. Journal of Experimental Psychology: Applied 8: 6-16.
Beilock S.L., Bertenthal A.M., McCoy A.M., & Carr T.H. (2004). Haste does not always make waste. Psychonomic Bulletin and Review 11: 373-79.
Block N. (1995). On a Confusion About the Function of Consciousness. Behavioral and Brain Sciences 18:227--47.
Bratman, M. (1987), Intention, Plans, and Practical Reason. Cambridge, Mass.:Harvard University Press.
Davidson, D. (1980), Essays on Actions and Events (2nd ed: 2000). Oxford UP.
Davidson, D. (2005). Truth, Language, and History. Oxford UP.
Di Nucci, E. (2009), 'Simply, false', Analysis 69/1: 69-78.
Di Nucci, E. (2010), ´Rational Constraints and the Simple View', Analysis 70/1: 481-486.
Di Nucci, E. (2013), *Mindlessness*. Cambridge Scholars.
Gendler, Tamar Szabó (2008a), "Alief and belief", Journal of Philosophy, 105: 634–663.
Gendler, Tamar Szabó (2008b), "Alief in action, and reaction", Mind and Language, 23: 552–585.
Gert, B. (1998), Morality: its nature and justification. Oxford UP.
Gertler, Brie (forthcoming). 'Self-Knowledge and the Transparency of Belief', in A. Hatzimoysis (ed.) Self-Knowledge. Oxford: Oxford.
Gigerenzer G. (2007). Gut Feelings. Penguin.
Hunter, David (2009). 'Belief, Alienation, and Intention', unpublished manuscript.
Iyengar S.S. & Lepper M.R. (2000). When choice is demotivating. Journal of Personality and Social Psychology 79: 995-1006.
Lycan W.G. (1986) "Tacit belief", in R.J. Bogdan, ed., Belief: Form, content, and function (Oxford: Clarendon), 61–82.
Macrae, C.N. & Johnston, L. (1998), 'Help, I Need Somebody: Automatic Action and Inaction', Social Cognition 16: 400-417.
Mandelbaum E. (2012). Against Belief. Philosophical Studies (forthcoming).
McCann, H. 1991. Settled objectives and rational constraints. American Philosophical Quarterly 28: 25–36.
McCann, H. 2010. Di Nucci on the Simple View. Analysis 70: 53–59.
McCann, H. 2011. The Simple View again: a brief rejoinder. Analysis 71(2): 293-95.
Mele, A. and Moser, P. K. (1994), 'Intentional Action', Nous 28: 39-68.
Montero, B.G. (2016), *Thought in Action*. Oxford University Press.
Neal D.T., Wood W., Wu M., Kurlander D. (2011). The Pull of the Past. Personality and Social Psychology Bulletin XX(X): 1-10.

Newman G.E., Diesendruck G., Bloom P. (2011). Celebrity Contagion and the Value of Objects. The Journal of Consumer Research (forthcoming).
Oettingen G., Grant H., Smith P.K., Skinner M., Gollwitzer P.M. (2006). Nonconscious goal pursuit: Acting in an explanatory vacuum. Journal of Experimental Social Psychology 42: 668-75.
Parks-Stamm E.J., Oettingen G., Gollwitzer P.M. (2010). Making sense of one's actions in an explanatory vacuum. Journal of Experimental Social Psychology 46: 531-42.
Price, H. H. (1969). Belief. London: George Allen & Unwin.
Rowbottom, Darrell P. (2007). ' "In-Between Believing" and Degrees of Belief', Teorema 26,pp. 131–7.
Schwitzgebel E. 2010a "Acting Contrary to Our Professed Beliefs, or the Gulf Between Occurrent Judgment and Dispositional Belief" Pacific Philosophical Quarterly, 91, 531-553.
Schwitzgebel E. 2010b Belief. Stanford Encyclopedia of Philosophy.
Searle J.R. (1992) The rediscovery of the mind (Cambridge, MA: MIT).
Sommers, Fred (2009). 'Dissonant Beliefs', Analysis 69, pp. 267–74.
Stout, R. (2010), 'Deviant Causal Chains', in O'Connor, T. & Sandis, C. (eds.), A Companion to the Philosophy of Action. Blackwell: 159-165.
Thaler R.H. & Sunstein C.R. (2008). Nudge. Caravan.
Wilson, Timothy D., Samuel Lindsey, and Tonya T. Schooler (2000), "A model of dual attitudes", Psychological Review, 107: 101–126.
Wood W. (2012). On Ruts and Getting Out of Them. Science 336 (6084): 980-81.
Zimmerman A. (2007), "The nature of belief", Journal of Consciousness Studies, 14(11): 61–82.