
Are There Any Nonmotivating Reasons for Action?

When performing an action of a certain kind, an agent typically has several reasons for doing so. I shall borrow Davidson's term and call these rationalising reasons (Davidson 1963, 3). These are reasons that allow us to understand what the agent regarded as favourable features of such an action. (There will also be reasons against acting, expressing unfavourable features of such an action, from the agent's point of view.) I shall say that R is a rationalising reason of agent X 's for K -ing iff R consists of (i) a desire of X 's to L and (ii) a belief of X 's that K -ing promotes L -ing (to be discussed shortly). It is frequently said that when an agent X is K -ing and has several rationalising reasons for K -ing, not all of those reasons are reasons for which X is K -ing, that motivate X 's K -ing, or that explain X 's K -ing. In this paper I challenge this view.

The notions of *a* reason for which X is K -ing, *a* reason that motivates X 's K -ing, and *a* reason why X is K -ing incorporated in this popular view I assume to be nonpragmatic notions that do not depend on a context of enquiry. It should also be noted that there are pragmatic reason notions that are especially likely to be intended when the definite article is used – when one talks of *the* reason why X is K -ing, etc. Such notions are context-dependent, taking into consideration the knowledge and interests of an enquirer. The same holds for the notions of *a* cause and *the* cause. I am not denying in this paper that there are reasons that might be regarded as nonmotivating, etc., in the pragmatic sense that they are not of interest in the context in question.¹

I have formulated my thesis in terms of reasons for which X is K -ing, reasons that motivate X 's K -ing and reasons why X is K -ing (i.e.

¹ When a rationalising reason R is held not to be a reason which motivates X 's K -ing, it might also be said that X does not K in order to L , and that X does not K with the intention of L -ing. I have not included these formulations among my targets in this paper, as I take them generally to be understood in a pragmatic sense. But I do take the arguments of this paper to show that there are no nonpragmatic senses in which these statements are true.

reasons that explain X 's K -ing).² I take these to be equivalent, but I shall not argue for this here. For the challenge I am presenting in this paper would apply to all three, should anyone draw distinctions among them. For the sake of simplicity I shall cast my discussion in terms of motivating reasons, except when referring to authors who have chosen a different formulation.

A principal assumption of this paper is that the notion of a non-motivating reason is not fundamental, but supervenes on other notions without necessarily being analysable in terms of them. It will accordingly have criteria, by which I mean conceptually sufficient conditions, given in terms of these other notions. It is also sharp rather than vague – either a reason motivates or it does not. Accordingly, it must have sharp criteria. So if there are no such criteria there can be no nonmotivating reasons. My strategy is to consider a number of seemingly plausible criteria of nonmotivating reasons and to argue that none would be ultimately acceptable to those who believe in such reasons.

Part I introduces some qualifications and clarifications in the formulation of rationalising reason. In part II I discuss the causal theory of acting for a reason and its potential role in distinguishing motivating from nonmotivating reasons, commenting on the views of Donald Davidson and George Wilson. Part III examines some uses that have been made of the claim that not all reasons are motivating. I begin with a paper by Ramsey, Stich, and Garon on connectionism, turn to Harry Frankfurt on freedom and responsibility, and finally mention the numerous uses of nonmotivating reasons in morality and the law. Part IV discusses a number of proposed criteria of a nonmotivating reason, under the headings of neurophysiology, counterfactuals, timing, conscious thought and method of deliberation, and explanatory salience, and finds none to plausibly show that there are nonmotivating reasons. Part V is a brief conclusion.

I

K -ing might be doing something with a certain character, for example telling the truth or keeping with tradition, or it could be bringing about

² I restrict this to *rationalising* reasons why X is K -ing. There is a wider sense in which 'reason why X is K -ing' can be used to denote any causal explanation of X 's K -ing. A reason that explains X 's K -ing I take to be any satisfactory answer to the question "Why is X K -ing?" Many hold that although a rationalising reason shows us something the agent sees in K -ing, not all such reasons satisfactorily explain why she is K -ing.

a certain kind of outcome, such as sending someone a letter or poisoning a colony of ants.

I shall stipulate that ‘desire to *K*’ is to be understood in a broad sense to include such attitudes as thinking it good to *K*, thinking it right to *K*, and thinking one ought to *K*. I could have adopted Davidson’s term ‘pro-attitude’ for this broad notion (Davidson 1963, 4). But either choice of term involves some stipulation. If I used ‘pro-attitude’ I would need to stipulate that these pro-attitude terms are not to be understood in the sense in which one may believe that a prospective action of *K*-ing is right and good, and that one ought to *K*, without being in the slightest bit moved to *K*.

The notion of a rationalising reason should be distinguished from that of a normative, justificatory, or good reason for *K*-ing. I am not denying that there can be nonmotivating normative reasons for action.

As I am not trying to analyse the notion of intentional action in this paper, I shall stipulate that by *K*-ing I mean *intentionally* doing some *action* of kind *K*, rather than the broader notion of doing some action of kind *K* or the still broader notion of doing some event of kind *K*. This allows the above belief and desire conditions to serve as a definition of rationalising reason. Without such a stipulation the belief and desire conditions would be necessary but not sufficient for being a rationalising reason. For it makes no sense to talk of a belief/desire pair as a reason of *X*’s for *K*-ing if *K*-ing is not *intentionally* doing some *action* of kind *K*. Suppose that I want to inherit my uncle’s fortune and believe that I would achieve this by accidentally killing him but not by deliberately killing him. We could not say that this belief and desire constitute a reason that rationalises my accidentally killing him when it is a truism that I cannot do this intentionally. And suppose I want to be rid of my insomnia, and believe that this would be promoted by my falling asleep naturally without adopting any ploy. We could not say this desire and belief rationalises the event of my falling asleep naturally when it is a truism that I cannot do this intentionally. These belief/desire pairs would constitute reasons why I might be pleased if I *K*’d, but are not what is meant by reasons for *K*-ing.

The most frequent formulation of the belief component of a reason is that *K*-ing is, or will turn out to be, *L*-ing. Having this belief is best understood as taking it to be more probable than not that *K*-ing will turn out to be *L*-ing on the occasion in question. But this seems too strong in cases in which someone bets on long odds because of a large possible gain. In such cases one does not believe that one’s betting will turn out to be one’s placing a winning bet, yet the prospect of winning can clearly play a role in rationalising the action of betting. To replace this formulation with the weaker belief that *K*-ing *might* turn out to be

L-ing would, I think, be too weak. For if *X* holds that the alternative(s) to *K*-ing have a greater subjective probability of turning out to be *L*-ing than *K*-ing does, one should not count the belief that *K*-ing might turn out to be *L*-ing as featuring in a reason to *K*. Thus I think the belief that should feature in the definition of rationalising reason is the regarding of *K*-ing as having a greater subjective probability of turning out to be *L*-ing than some of the alternatives to *K*-ing do, and this is what I have termed the belief that *K*-ing *promotes L*-ing.³

II

According to a fairly typical formulation of the causal theory of motivating reasons (and similarly for explanatory reasons, acting for a reason), *X*'s *K*-ing is motivated by reason *R* iff *R* causes *X*'s *K*-ing in the right way, or nondeviantly. If it is not stipulated that *K*-ing is intentional, then, as indicated above, I think this formulation of the causal theory should be revised to talk of *R* as a belief/desire pair instead of as a reason. Alternatively, if one is not trying also to analyse intentional action, then one may adopt my proposed stipulation that *K*-ing is intentional. This would allow to avoid the problem of deviant causal chains, as the problem arises precisely when an instance of *K*-ing fails to be an instance of intentional action. The causal theory may then be stated simply as maintaining that *X*'s *K*-ing is motivated by reason *R* iff *R* causes *X*'s *K*-ing. 'Causes' should be understood here in a nonpragmatic sense to match that of the reasons notions. The formulation I prefer replaces 'causes' by 'is causally relevant to' as it is perhaps least likely to be interpreted in a pragmatic sense.

One might think that such a causal theory straightforwardly supplies a criterion for a reason to be nonmotivating, namely the following: a rationalising reason of *X*'s for *K*-ing is nonmotivating if it is causally irrelevant to *X*'s *K*-ing. I accept this causal theory and hence this causal criterion. But such a criterion does not allay scepticism about nonmotivating reasons, as causal notions too are not primitive but supervene on further notions. There must accordingly be criteria for a rationalising reason to be causally irrelevant. Thus the ensuing search for criteria of nonmotivating reasons can also be taken as a search for criteria of causally irrelevant reasons.

It is commonly thought that the causal theory plays the substantial role of ruling certain rationalising reasons to be nonmotivating because they are causally irrelevant to the action. And it is frequently assumed

³ I shall allow this to include the special case in which $K = L$, when *X* is *K*-ing for the simple reason that she feels like *K*-ing (has an intrinsic desire to *K*).

that this view is explicitly expressed by the foremost advocate of the causal theory, Donald Davidson, in his 'Actions, Reasons, and Causes' (Davidson 1963). The following two passages give some support to this interpretation, but they are not conclusive.

In one passage, Davidson claims that "a person can have a reason for an action, and perform the action, and yet this reason not be the reason why he did it" (Davidson 1963, 9). This first passage can be interpreted as affirming the claim I am challenging in this paper by recasting it to read "a person can have a reason for an action, and perform the action, and yet this reason not be *a* reason why he did it." But perhaps all Davidson meant is that not all of a person's reasons for an action would be singled out in the context as *the* reason that explains his action, i.e. the most explanatorily salient reason. And this does not commit him to the view that some of one's reasons for an action one performed are not explanatory.

In a second passage Davidson says that you may be mistaken about the claim that you are *K*-ing for reason *R*, for you may have two reasons for *K*-ing one of which pleases you and one which does not. You may be wrong about which motive made you do the action (Davidson 1963, 18). This can be interpreted as meaning that you may be mistaken about *R* being *a* reason that made you *K*. And on such an interpretation Davidson would be affirming the claim I am challenging – that you could *K* with *R* being a reason of yours for *K*-ing yet *R* not be a reason which made you *K*. But again it is also possible that Davidson means you may be mistaken about *R* being *the explanatorily salient* reason which made you *K*. And this does not commit him to the view that some of one's reasons for an action one performed are not reasons for which one acted or reasons that made one act.

If the arguments of this paper succeed in showing that there are no nonmotivating or causally irrelevant reasons, it follows that the causal theory does not eliminate any rationalising reasons from the class of motivating reasons. But the causal theory will not then be vacuous, for it still asserts a conceptual connection between motivating reasons (reasons for which, reasons why) and causally relevant reasons that was widely rejected prior to the publication of Davidson's seminal paper.

George Wilson offers two examples that he takes to undermine the claim that a causal theory provides sufficient conditions for acting for a reason (Wilson 1997, 72). They can also be taken as offering intuitively compelling cases of rationalising reasons that are not reasons for which the agent acts. In the first example, Norbert has a desire to phone his mistress and believes that he can attain this end by walking to the phone. But he also has a competing second-order desire to be rid of that first-order desire, and decides that the best way of satisfying this higher-order

desire is to phone his psychiatrist. He also believes he can attain this end by walking to the phone. Now when he does walk to the phone, Wilson claims that the first reason may not be a reason for which he acted even though it played a nondeviant role in causing the action.

I shall pass over Wilson's claim that the causal role is nondeviant in this example. But I do wish to challenge the claim that Norbert has both reasons to walk to the phone. To make his example seem plausible we must suppose that Norbert genuinely has an occurrent desire to phone his mistress simultaneously with an occurrent higher-order desire to extinguish that first-order desire. And we must suppose that there is no doubt in his mind as to whom he will phone when he reaches the phone. For otherwise it becomes plausible that Norbert is genuinely motivated by two reasons in crossing the room to the phone. Wilson rightly notes that Norbert believes that by walking to the phone he can phone his mistress. But this connection is too weak to provide a reason for Norbert to walk to the phone on this occasion. The belief that by *K*-ing *X* can *L* does not feature in a reason of *X*'s to *K* as it is compatible with the belief that by *K*-ing *X* is less likely to *L*. In Wilson's example, Norbert does not believe that walking to the phone promotes phoning his mistress, because the circumstances as described are such that he believes he will decrease his chance of phoning his mistress by walking to the phone. So the putative example of a rationalising reason that is not a reason for which the agent acts turns out not to be a rationalising reason at all.

The same applies to Wilson's second example: "Nancy's desire to get a copy of *Newsweek*, combined with her belief that by heading for Boston on flight 1 she would get a copy, constitute (at least in Davidson's sense) a primary reason for her to head for Boston on flight 1" (Wilson 1997, 73). This belief/desire pair would indeed constitute a primary reason in Davidson's sense, though not in what I have argued to be the important sense of rationalising reason. For in Wilson's construction of this example, Nancy does not believe that taking flight 1 raises her chances of getting a copy of *Newsweek*, as she believes she will be equally able to get a copy if she takes flight 2 (or decides not to take either flight).

Someone might think that in focusing on the case in which *X* has rationalising reasons for *K*-ing and *X* does *K*, I am artificially excluding the huge class of rationalising reasons that are nonmotivating when *X* fails to *K*. And one might further suspect that if it is possible for rationalising reasons to be nonmotivating when *X* fails to *K*, then it ought to be possible too in the case I'm discussing. However, I do not wish to ignore these cases. Nor do I think they offer a significant departure from my thesis, or suggest a problem for it. In the case in which *X* fails to *K*, *X*'s rationalising reasons for *K*-ing trivially fail to motivate or be causally relevant to *X*'s *K*-ing. But no trivial semantic feature precludes

them from counting as causally relevant to X 's refraining from K -ing. If they are causally relevant, they will be negatively causally relevant in this case. As 'motivating' is necessarily positive we would need another term to correspond to negative causal relevance – 'inhibiting' perhaps. If we now interchange ' K -ing' and 'refraining from K -ing,' we are effectively looking at X 's rationalising reasons *against* K -ing when X is K -ing, and considering whether some or all are negatively causally relevant to X 's K -ing, or inhibit X 's K -ing. An examination of the thesis that some but not all of an agent's reasons against K -ing are inhibiting reasons could be undertaken *via* a search for criteria of noninhibiting reasons, parallel to that offered here for nonmotivating reasons. I expect that such an enquiry would mirror the one I am offering, and allow the result to be generalised to claim that there are no nonmotivating, noninhibiting, or causally irrelevant reasons for or against acting.

III

I turn now to examine some more of the uses that have been made of the claim that there are nonmotivating reasons. This will allow us to see what is at stake in accepting this claim. And it will introduce some further considerations relevant to its assessment.

First I want to consider an argument presented by Ramsey, Stich and Garon in their 1990 paper 'Connectionism, Eliminativism, and the Future of Folk Psychology.' They argue that Folk Psychology is committed to propositional modularity, that such modularity cannot be accommodated on connectionist models of the mind, and hence that Folk Psychology must be eliminated if connectionist models turn out to offer the best accounts of human mentality.⁴ Propositional modularity is characterised as the view that there are (i) functionally discrete, (ii) semantically interpretable states that (iii) play a causal role in some cognitive episodes but not in others.⁵ To illustrate how Folk Psychology takes propositional attitudes to be functionally discrete causally active states, they offer two examples. In the first they consider Alice's action of going to her office and offer two belief/desire pairs that rationalise her action. They say: "In such cases, commonsense psychology assumes that Alice's going

⁴ I'm much more inclined to adopt the option they recognise in Ramsey, Stich & Garon 1990, 352 of running *modus tollens* with this conditional, and inferring the falsity of any connectionist model that is incompatible with Folk Psychology.

⁵ The definition of propositional modularity is given in Ramsey, Stich & Garon 1990, 355 without the statement that the causal role is to feature in some cognitive episodes but not in others. I have inserted this clause as it plays a crucial role in their argument and is stated in their conclusion (Ramsey, Stich & Garon 1990, 374).

to her office might have been caused by either one of the belief/desire pairs, or by both, and that determining which of these options obtains is an empirical matter” (Ramsey, Stich & Garon 1990, 357). In the second example they say: “On the commonsense view, it may sometimes happen that a person has a number of belief clusters, any one of which might lead him to infer some further belief. When he actually does draw the inference, folk psychology assumes that it is an empirical question what he inferred it from, and that this question typically has a determinate answer” (Ramsey, Stich & Garon 1990, 357-358).

The second example provides an interesting analogue in the case of inference to the problem I am discussing about action. The claim that an inference to a new belief may have been caused by one but not another coexisting reason for the belief appears plausible in the circumstances in which the reasoner has two beliefs before the mind, q and r , and then deduces p from q while failing to see that p could also be deduced from r . This involves a failure of rationality on the part of the reasoner in not drawing the logical consequence of a belief. Such failures occasionally occur, especially if the deduction of p from r is at all complex. But I think it would be widely agreed that it would make no sense to say that someone notices that he can deduce p from either q or r and then goes ahead and infers p from q but not from r . These two ways of interpreting the claim that one can have a belief that plays no role in inference are analogous to two interpretations of the claim that one can have a non-motivating reason for action, except that it is the second, not the first, of these interpretations that is at issue in the case of action. Someone might not desire to K despite her belief that K -ing promotes L -ing and her desire to L . But such failures of rationality would be rare as the inference involved is extremely simple.⁶ The interpretation of interest in the case of action, rather, is that in which the agent sees perfectly clearly how her belief and desire rationalise her action, but just (allegedly) fails to be motivated by that reason in acting.

I also wish to note in passing the claim made in the first passage quoted, that common sense psychology views it as a matter for empirical investigation which reason caused the action. The view may very well be a feature of common sense psychology, but it will be mistaken if there are no clear criteria that can be called upon to judge that a reason is nonmotivating. For if nonmotivating reasons are to be empirically discernable, they must be linked by criteria to features that are empirically discernable.

⁶ One might prefer to augment the definition of rationalising reason so as to redescribe these failures of rationality as cases in which the agent lacks the reason for K -ing, although she has both belief and desire.

These two examples are the only ones Ramsey, Stich and Garon use to illustrate the claim, to which they take Folk Psychology to be committed, that there are functionally discrete states that play a causal role in some cognitive episodes but not in others. One might infer from this that they take the causal claims that there are reasons for an action that are causally irrelevant to that action, and reasons for a belief that play no role in inferring the belief, to be essential features of Folk Psychology. It would indeed be a striking claim if they took the causal thesis that there are nonmotivating reasons for action to be such a central feature of Folk Psychology that if it could not be sustained we would have to abandon Folk Psychology and declare that there are no beliefs or desires. It seems much more likely, however, that they would regard the examples they use to illustrate (iii) – how states may play a causal role in some episodes but not in others – as inessential to their position. Take the first example in which the agent's action of *K*-ing is said to be caused by one of its rationalising reasons, R_1 , but not by another, R_2 . As a full illustration of (iii), the example would need to be supplemented by another one showing that R_2 has effects that R_1 does not have. But an example involving reasons that rationalise different actions would seem to provide a much more obvious illustration: My desire to eat protein and my belief that tuna is rich in protein cause me to eat tuna but not to fetch my umbrella. My belief that it's raining and my desire not to get wet cause me to fetch my umbrella but not to eat tuna.

So the claim that there are nonmotivating reasons is not essential to component (iii) of their definition of propositional modularity. Nor does it appear to be essential to either of the other components. For the central characteristic of functional discreteness as they state it is its typically making sense to say an agent may acquire or lose a single propositional attitude (together with those conceptually entailed by it), and it would typically make sense to say an agent may acquire or lose a single rationalising reason, whether or not there are nonmotivating reasons.⁷

⁷ Nor does the thesis that there are nonmotivating reasons provide a special case of propositional modularity that is essential to their argument that all three components of propositional modularity are inconsistent with connectionism. For concerning (ii) they argue that Folk Psychology treats the class of believers that dogs have fur as a natural kind whereas this class would not be a natural kind under connectionism. Concerning (i) they argue that connectionist networks appear not to encode propositions in fully discrete ways. And concerning (iii) they argue that it makes no sense under connectionism to ask whether or not the representation of a particular proposition plays a causal role in the network's computation. This they attribute entirely to its not making sense to treat the representation of any given proposition as a discrete entity. None of these arguments requires the existence of nonmotivating reasons.

Next I turn to Harry Frankfurt, who argues that judgements of responsibility for actions depend on the reasons that motivated the action (Frankfurt 1969). Suppose one has a reason R_1 for K -ing and a coercive reason R_2 for K -ing. According to Frankfurt, one is intuitively excused for K -ing, i.e. one lacks responsibility for K -ing, if one K 's solely from the coercive reason R_2 . However, if one K 's solely from R_1 then one may be responsible for K -ing even though one could not have done otherwise, because of the causally inoperative coercive reason R_2 . (In this example 'could not have done otherwise' is to be interpreted in the sense of 'could not reasonably be expected to have done otherwise.')

This Frankfurt takes as undermining the *Principle of Alternate Possibilities*: that one is morally responsible for an action one performs only if one could have done otherwise. Frankfurt says that it will be very difficult to tell in a situation like this whether X is acting from R_1 , from R_2 , or from both reasons (though he adds that "it is not impossible, however, that the situation should be clearer than situations of this kind usually are"). But he does not say how one would tell, or offer a metaphysics of reasons and causes that allows for these distinctions to be made.

If I am right that R_1 and R_2 cannot be nonmotivating because there are no nonmotivating reasons, then this counterexample of Frankfurt's to the *Principle of Alternate Possibilities* disappears. But it may be that the situation he does not discuss in which the agent K 's from both R_1 and R_2 still serves as an adequate counterexample. For as before, the presence of the coercive reason R_2 ensured that X could not (reasonably) have done otherwise than to comply with the threat. And it could plausibly be argued that X K 's with diminished responsibility when she K 's from both R_1 and R_2 , but with responsibility nonetheless. So Frankfurt could still use the presence of coercive reasons in constructing a counterexample to the *Principle of Alternate Possibilities*. Thus it turns out that the main point of Frankfurt's paper, like that of Ramsey, Stich and Garon, can be argued without using the claim that there are non-motivating reasons.

Another frequent use of the claim that there are nonmotivating reasons is in arguing that judgements of moral worth or praiseworthiness depend on the reasons that motivated the action. This time it is supposed that X has two reasons R_1 and R_2 for K -ing, both of which are noncoercive and so allow responsibility. But one of these reasons is applauded while the other is decried. A popular illustration is the case of pulling the plug on the life support system of an elderly parent. One reason for doing so is that one wants one's parent to die in dignity. Another reason is that one wants to avoid being bankrupted by hospital bills. The idea is that one may have both reasons but so long as one acts solely from the reason of dignity one has not performed a blameworthy act.

Something very similar is argued by a number of Kant scholars who take Kant's thesis that morally worthy actions must be performed solely from the motive of duty not to preclude the presence of reasons for doing the dutiful act that are based on inclinations such as sympathy. They interpret Kant as claiming that the mere presence of such inclination-based reasons does not prevent an act from having moral worth so long as those reasons are nonmotivating.⁸

And there are numerous statutes enshrined in law claiming that someone is liable when acting (e.g. hiring or firing) on the basis of race, gender, etc. Merely having such a race- or gender-based reason for an action one performs is insufficient for incurring liability.⁹

In all these moral and legal cases I think that the work that the notion of nonmotivating reason is required to do can be, indeed must be, taken up by other notions, such as those shortly to be examined as candidates for criteria of nonmotivating reasons. This will involve clarification and necessarily some revision. I turn now to those proposals.

IV

Neurophysiology

In the course of arguing that a causal condition is needed in an account of acting for a reason, Alfred Mele claims that if a person has two reasons R_1 and R_2 for K -ing and R_2 is rendered impotent by a neurosurgeon, one wouldn't say the person K 's for reason R_2 .¹⁰ And some have suggested that in a technologically advanced future state, one could look inside the brain and see what reasons are motivating and what aren't from their neural properties. What underlies both these suggestions is the idea that there are neural criteria of nonmotivating reasons.

A first response to this idea is that it is assumed in morality and the law, and in commonsense psychology according to Ramsey, Stich and Garon, and in the paper by Frankfurt, that we can at least sometimes determine that a reason is nonmotivating. And this requires that non-motivating reasons have macroscopic criteria that are accessible to the layperson who is not equipped with a brain-scanner. If there are such macroscopic criteria and they in turn have neural criteria, it follows that

⁸ In Latham 1994 I discuss the implausibility of attributing this view to Kant, as well as the difficulty in making sense of nonmotivating inclinations. Many of the arguments of the present paper are based on that discussion.

⁹ I offer some examples in Latham 1994.

¹⁰ In his introduction to Mele 1997, 11-13.

there will be neural criteria of nonmotivating reasons. But it would conflict with common intuitions if the *only* criteria of nonmotivating reasons were direct neural ones that are not derivable from macroscopic criteria.

If we fail to find any macroscopic criteria we might become persuaded that there aren't any to be found, and this would lead us to give up the intuition that nonmotivating reasons are macroscopically accessible. We might then give up the idea that there are any nonmotivating reasons. But before coming to this conclusion we should look further into the possibility that there are neural criteria of nonmotivating reasons but no macroscopic criteria. We have already noted that this view faces the problem that people believe there are macroscopic criteria.

How then would we decide what is a correct neural criterion of the presence of a nonmotivating reason, either occurring naturally, or after the neurosurgery? Mele suggests the condition in which the neural realiser of a reason remains intact while its access to action is blocked. This suggests that realisation may provide the required link between talk of reasons and talk of neural entities, so let us explore this idea. Strictly speaking, a neural realiser of X 's reason R at t (or of X 's K -ing at $t+\epsilon$) must be an instance of a complex neural property of X 's brain and physical property of its surroundings that guarantees the presence of R at t (or of X 's K -ing at $t+\epsilon$). Presumably, parts of the brain must be specified in great neural detail, but the rest of the brain can be specified more roughly, perhaps in terms of an acceptable range of parameters. It follows that any instance of a mental property will have many neural realisers corresponding to the many ways in which detail can be left out. Furthermore, it is unlikely that there will be a unique minimally detailed neural realiser of a given mental property instance. For it is plausible that there is some overdetermination in the brain, and hence that different regions of the brain may serve as those to be specified in greatest detail. Let us suppose we can make some sense of the location in the brain of an instance of a mental property M , such as perhaps the union of all regions that must be given in greatest detail in a minimal neural realiser of the instance of M . Now suppose we refer to the location of R_1 at t as L_1 , of R_2 at t as L_2 , and of X 's K -ing at $t+\epsilon$ as L_3 . One proposal we might consider is that a reason R_1 is *nonmotivating* if there is a chain of neuronal firings between t and $t+\epsilon$ leading to L_3 from a part of L_2 that doesn't overlap with L_1 , and no such chain leading from a part of L_1 that doesn't overlap with L_2 . This proposal tries to exploit the idea of realisation as a way of linking mental talk to neural talk, and it adds an account of causation at the neural level in terms of chains of neural firings.

A second problem for the idea that there are neural but not macroscopic criteria of nonmotivating reasons is that whether any such pro-

posal is ever satisfied will depend on empirical claims about the brain. And this clashes with another commonly held intuition about nonmotivating reasons, namely that there is no question that they occur. In the above proposal, for example, it is a substantial empirical assumption that L_1 and L_2 will not fully overlap. So this proposal requires certain empirical facts to hold about the locality of the reasons in the brain in order for there to be any nonmotivating reasons. Yet, it is commonly supposed that we do often have nonmotivating reasons, and that even if it is not possible to tell when we have them, it could not turn out that our brains are wired in such a way that we never have any.

I also think there is a third problem for the idea that there are neural but not macroscopic criteria of nonmotivating reasons. Suppose some such criterion of a nonmotivating reason is proposed. Why should it be accepted that the relevant notion of causation involves a chain of neuronal firings? We often find absence of movement to be positively causally relevant at the macroscopic level, so it might turn out that we find the nonfiring of some neurons to have positive causal relevance. There may therefore be a variety of causal notions at the neuronal level to choose from, and thus the choice that features in any neural criterion needs to be defended. How would one do so? Suppose one discovers that actions performed habitually do not involve any neuronal activity in regions of the brain where the reasons are located, and that the chain of neuronal firings corresponds to what is on the person's mind immediately prior to acting. One would then need to decide whether reasons are necessarily nonmotivating once they become habitual, and whether motivating reasons must be on the agent's mind while acting. So in order to test out any neural criterion it appears we need an understanding of when a reason is a nonmotivating reason at the macroscopic level, and this seems to require a macroscopic criterion.

Counterfactuals

I turn now to a search for macroscopic criteria of nonmotivating reasons, beginning with proposals that use counterfactuals. Let us seek a criterion of a reason R 's being nonmotivating, and let us label the remaining independent reasons for K -ing P and the reasons against K -ing N . Assuming that X does K , it follows that R and P together are sufficient to counteract the negative reasons N . Now let us define a reason R as sufficient for K -ing iff X would have K 'd even if X had not had P . And R is necessary for K -ing iff X would not have K 'd if X had not had R . The counterfactual circumstances in which a reason (say R) does not exist should be understood either as those in which X has a neutral attitude to L -ing, or as those in which X believes that K -ing is just as

likely as any of its alternatives to turn out to be *L*-ing. (Later I shall be considering a different way of understanding the counterfactual circumstances.)

There are three cases to consider

- (1) neither *R* nor *P* is sufficient for *K*-ing, i.e. both *R* and *P* are necessary for *K*-ing.
- (2) both *R* and *P* are sufficient for *K*-ing, i.e. neither *R* nor *P* is necessary for *K*-ing.
- (3) *R* is neither necessary nor sufficient for *K*-ing, i.e. *P* is both necessary and sufficient for *K*-ing.

As we are interested in the case in which *R* is a nonmotivating reason we can ignore case (1) in which both *R* and *P* are necessary, and hence both are motivating. Case (2) provides the most interesting alleged instances of nonmotivating reasons since they involve strong reasons that somehow fail to motivate. But a plausible counterfactual criterion is hard to find. We cannot say *R* is nonmotivating if it is not necessary, since we would have to say the same for *P*, and clearly we cannot say both *R* and *P* are nonmotivating. Suppose we break the symmetry by proposing that *R* is nonmotivating if it is not necessary and is weaker than *P*, where reason strength is a counterfactual notion constructed in terms of how big a disincentive to *K* the reason would overcome.¹¹ Two scenarios make this proposal implausible, however. Scenario 1: Initially *R* is *X*'s only reason for *K*-ing. It is sufficient, but just barely, and *X* is hovering on the brink of *K*-ing. Then a slightly stronger sufficient reason *P* arrives, whereupon *X* immediately *K*'s, thinking of this as making the decision an easy one. Scenario 2: Initially *R* is *X*'s only reason for *K*-ing. It is sufficient, and *X* has decided to *K* at a certain moment in a certain manner. Then sometime before the allotted time she acquires a new stronger sufficient reason *P* for *K*-ing, yet doesn't change her plan at all but *K*'s exactly as originally planned. In both these scenarios *R* would be declared nonmotivating according to the proposed criterion, but I doubt a believer in nonmotivating reasons would want to say that *R* is nonmotivating on either scenario.

Case (3) seems more promising. But we cannot simply say that *R* is nonmotivating if it is neither necessary nor sufficient. For if we replace *P* by the individual reasons that it comprises, we could have a set of

¹¹ Such a notion of reason strength will not necessarily coincide with any phenomenological notion. It is generally hard if not impossible to compare reason strengths phenomenologically.

reasons for *K*-ing all of which are neither necessary nor sufficient for *K*-ing. And we cannot say that all *X*'s reasons for *K*-ing are nonmotivating when *X* does *K*. Can we then at least say that *R* is nonmotivating if it is neither necessary nor sufficient and another single reason of *X*'s is sufficient? The trouble with this proposal is that in some cases the question whether we have a single reason or a number of separate reasons seems purely a matter of convention, as for example when we consider the reason of not wanting to upset one's parents. Is this one reason or two? Perhaps, however, there are at least some reasons that cannot be broken up in this way, such as the desire to save one's life and the belief that *K*-ing promotes that end. Let us consider then the proposal that *R* is nonmotivating if it is neither necessary nor sufficient and there is a sufficient reason for *K*-ing that cannot be broken down into smaller reasons.

But when we focus on a single reason and wonder whether it was nonmotivating, do we think this really depends on how the remaining reasons are divided up? Suppose I have two reasons for sending a donation of 500 euros to the local zoo. I've been told it will be used to buy medicine to save the life of an elephant and one of their many lizards. I care for the lizard but not enough to donate the money just to save its life. However, I do care enough about the elephant to donate the money to save its life alone, and this reason cannot be broken down into smaller reasons. So on the criterion under consideration, my reason of wanting to save the lizard would be counted nonmotivating. Yet, it would not be counted nonmotivating according to this criterion, if I am told that instead of saving an elephant the money will be used to save the lives of 100 sick reptiles, and I am happy to donate the money for this purpose. Saving the life of that lizard is now a reason of comparable strength to those of saving the lives of each of the 100 other reptiles. The proposed criterion of being a nonmotivating reason thus appears implausibly arbitrary. Intuitively it will be thought that if wanting to save the life of the lizard provides a nonmotivating reason in the 100 reptiles case it will do so in the elephant case too, and that whatever criterion is at work in the former case will apply also to the latter case. So it is implausible to think that the current proposal is what is at work in the elephant case, and the search for a criterion is still on.

The proposal can be seen as implausible quite apart from this arbitrariness by reflecting on a variant of scenario 1 considered earlier. Suppose *X* has two reasons, one of which, R_1 , is almost sufficient while the other, R_2 , is unbreakable and just barely sufficient. Someone believing in nonmotivating reasons would probably not think R_1 nonmotivating if *X* was hovering on the brink of *K*-ing before the arrival of R_2 , but then *K*'d once R_2 came along, thinking of this as making the decision an easy one.

It is probably the fact that R_1 and R_2 are almost equal in strength that makes it seem that both are motivating in this last scenario. But many would hold that very weak reasons must sometimes be nonmotivating. Suppose I'm out walking in the forest and come to a plant with some bright juicy-looking berries. I think it would be worth picking them to eat, although I see that this would involve the minor inconvenience of stepping in some mud. My companion then informs me that the berries are extremely poisonous, and I refrain from picking them. Clearly a reason stemming from my desire not to be poisoned is one motivating reason for leaving the berries alone. But it would seem absurd to say that my desire not to step in the mud was another motivating reason. The example seems compelling, and countless others like it are easy to think up.

The fact that many people would claim that examples like these constitute nonmotivating reasons is not in dispute. The task before us is to find a criterion that would justify such claims. What could it be? It cannot be that *a reason is nonmotivating if it is weak*, because the notion of a nonmotivating reason is not vague – either a reason motivates or it does not. Nor can it be that *a reason is nonmotivating if it falls below a threshold of strength*, because any choice of threshold would be implausibly arbitrary.¹² And in any case, both these proposals suffer from the fatal problem that they would absurdly declare all of X 's reasons for an action she is performing to be nonmotivating in a scenario in which X has many reasons, all weak and all below whatever level one sets the threshold at, that yet together are sufficient for her acting.

So far I have said that R is unnecessary for K -ing if X would K anyway even if X did not have R , where we understand the counterfactual circumstances in which X does not have R as circumstances in which R is reduced to neutrality. Let us now consider a final counterfactual proposal involving a stronger sense in which R is unnecessary for K -ing, namely that *R is nonmotivating if X would K anyway even if R is replaced by a reason against K -ing*. Such counterfactual circumstances would obtain either if X has an aversion to L -ing and believes K -ing promotes L -ing, or if X desires to L and believes that K -ing promotes the nonoccurrence of L -ing. However, this proposal suffers from problems similar to those that beset the last pair of proposals. First, some level of strength of the reason replacing R in the counterfactual circumstances would have to be designated. This could not be infinite, as the proposal would not then be satisfied by any reasons. And any finite choice of strength would be implausibly arbitrary. Second, the proposal would also face the fatal objection that for any designated strength, a scenario is possible in which

¹² Except for the high threshold discussed earlier equivalent to the strength of P .

all X 's reasons satisfy the criterion and hence all would absurdly be declared nonmotivating.

I conclude that no purely counterfactual criterion can support the intuition that there are nonmotivating reasons.

Timing

Reflection on scenarios 1 and 2 above suggests that the timing of a reason might be relevant to whether it is nonmotivating. An initially appealing proposal suggesting that the earlier of two reasons is nonmotivating is that a reason R_1 is nonmotivating if X has R_1 but has not yet $K'd$, and then K 's immediately upon the arrival of a sufficient reason R_2 . But this has already been shown to be implausible in discussing scenario 1 when R_1 and R_2 are almost equal in strength and X is having a hard time deciding whether to K before the arrival of R_2 .

From scenario 2 we get the following suggestion of circumstances in which the later of two reasons appears nonmotivating: R is nonmotivating if X plans to K on the basis of reasons she has for and against K -ing, and doesn't change her plan concerning when and how to K when she acquires a new reason R for K -ing. However, intentions are not commitments to act no matter what. Between the time the intention is formed and the time of action, people are sensitive to relevant features that might lead them to change their mind. And this suggests that whether or not a reason is motivating should be assessed purely in terms of how the reasons appear to the agent at the time of action. You've planned to attend a reception because you've promised to help the host with some minor details, but you would be happy to opt out if a good excuse arose and you are wondering whether you made the right decision. Then you learn that someone you need to talk with is going to be there, and this reason puts an end to your wondering whether it is really worth your while to attend the reception, even though it may not have been sufficient on its own for you to attend. You head off to the reception thinking of what you want to say to the person, not thinking at all of the help you will be providing the host. Should we say that your desire to talk with this person is a nonmotivating reason because it arrived late on the scene? I think that after reflecting on examples like this, a believer in nonmotivating reasons is more likely to hold that whether or not a reason is motivating does not depend on the timing of the reasons.

Conscious Thought and Method of Deliberation

The last example points to the relevance of the agent's conscious thought in determining whether a reason is nonmotivating. One idea to consider

is that R_1 is nonmotivating if X does not consciously consider it, but does consciously consider a sufficient reason R_2 . The idea here is that if the only reason you think of while deliberating and acting is sufficient, then that is your motivating reason and all other reasons are nonmotivating. However, it is widely held that actions can be motivated by unconscious reasons, as well as by reasons that would be available to consciousness but are not consciously reflected upon because the agent is habituated to acting from those reasons or is acting with a large degree of spontaneity, or because the reasons are too obvious. But the proposal may seem more plausible when strengthened by combining it with a counterfactual condition considered earlier: R_1 is nonmotivating if it is neither necessary nor sufficient and X does not consciously consider it, but does consciously consider a sufficient reason R_2 .

Still, I doubt that what is on the agent's mind will be thought to make the crucial difference in responding to the kind of case based on scenario 1 considered earlier in which R_1 is neither necessary nor sufficient although it is almost sufficient and the other reason R_2 is just barely sufficient. When X was hovering on the brink of K -ing prior to the arrival of R_2 , and then K 'd immediately upon the arrival of R_2 , it did not seem plausible to say that R_1 was nonmotivating. Let us now construct a case of this sort in which R_1 is unconscious. Imagine that Greta was told under hypnosis that she would leave the waiter a \$20 tip next time she is in a bar. Later that evening she is leaving a bar and the post-hypnotic suggestion is not powerful enough for her to comply. But she finds herself surprised at her anxiety as she is about to leave a \$5 tip, thinking how stingy she is being even though the drinks bill came to just \$25. Then her companion tells her that the secret to getting great service at this bar is to give the waiter a \$20 tip on your first visit. Let's suppose that without the hypnotic suggestion she would feel a little foolish and skeptical in going along with this idea but would give it a shot anyway. But on the occasion in question she throws down the \$20 bill without hesitation and with a surprising sense of relief. I doubt it will be thought that the unconscious reason she acquired under hypnosis is nonmotivating in this case. I think it will seem more plausible to say that both reasons are motivating, given that her state of mind contrasts markedly from her uncertainty and inner tension when just one of the reasons for acting is present.

Similar examples can be constructed in which R_1 is habitual or unreflective instead of unconscious, where the psychological impact of R_1 shows itself in the agent's emotional state, suggesting that R_1 is partly motivating the action. As an example of a putative nonmotivating reason involving what I am calling unreflective reasons, imagine a composer, Noriko, working on a score. Usually she adds phrases with nothing more

than a vague idea that it will sound right. Yet, upon reflection she is often able to give an articulate analysis of some of the considerations that led to her choice, and occasionally one such consideration is on her mind at the very moment she makes her choice. Today she is wondering whether a certain addition to her work would be a good idea. It sort of feels right to her, but she would have refrained were it not for her sudden realisation that the addition would provide a kind of reference to a favourite work of hers, upon which she cheerfully puts it down without reflecting on the other good features it might have. Suppose that if those other positive features had been absent, Noriko would have used the artistic allusion in any case, but with far less conviction. The difference it makes to her state of mind tells in favour of the view that those unarticulated reasons are partly motivating, contrary to the proposed criterion. I doubt it would be thought plausible to say that under such circumstances when one reason is before the mind, all other unarticulated reasons are nonmotivating.

I turn now to a proposal based on a method of deliberation found in Kant's ethics. Suppose that *X* intends always to follow the principle she has adopted of considering first reasons of a certain sort, call them *A*-reasons, and acting on the basis of them, except when the circumstances reveal that *A*-reasons are neutral with regard to what to do, in which case she may pursue what is favoured by the balance of the remaining reasons, call them *B*-reasons. The most plausible candidates for *A*-reasons are moral, aesthetic, and religious reasons. And now consider the suggestion that *R* is nonmotivating if *R* is neither necessary nor sufficient and is a *B*-reason, and *X* has overall *A*-reason to *K* and has adopted a principle of giving priority to *A*-reasons.

Again I think the proposal will be found implausible. For it is clear that *X* will suffer lapses from her principle when reasons to depart from it get sufficiently strong, and in these cases, in which *X* may be said to act with weak will, she finds herself intentionally going against her principle. On these occasions she will clearly be motivated by *B*-reasons. If *B*-reasons are motivating on such occasions it seems plausible that when *X* manages to act continently by following her *A*-reasons we should say that she acts despite the calling of her *B*-reasons. The *B*-reasons are negatively causally relevant to, or inhibit, her action. Why then can't *B*-reasons be positively causally relevant, or motivating, when they agree with her *A*-reasons? Here's an example that suggests that they can be. Suppose you dislike the taste of venison, but politely eat it whenever it is served at a dinner party at which you are a guest. You also intend as a first priority always to follow the guidance of your rabbi, and have recently heard him say that deer are unclean animals. The next time you are at a dinner party you hear from one of the guests that you

are going to be served pork, which you have occasionally illicitly eaten and loved. You entertain but reject the thought of keeping quiet about your principles, and are feeling a good deal of turmoil about making your announcement. Then when another guest says she thinks venison will be served, all hesitation and awkwardness vanishes as you proudly proclaim that your religious principles won't allow you to eat it. The reason for refusing that stems from your distaste for venison is postulated as neither necessary nor sufficient, and as a *B*-reason, and you do have an *A*-reason on this occasion. Yet contrary to the proposal, it is implausible to suppose that your *B*-reason is nonmotivating in this case.

Explanatory Salience

Consider finally the proposal that *R is not a reason why X K'd if there is no possible enquirer for whom R should be offered as the appropriate reason to explain why X K'd.*

This proposal interestingly makes use of a pragmatic reason notion to provide a criterion for a state of affairs that is given in terms of a nonpragmatic notion. Unlike the previous proposals, this one I do find plausible. However, it fails to eliminate any rationalising reasons in my view, because any rationalising reason can be regarded as *the* reason why the agent acted, given a suitable enquirer. Consider the case of Tom, who finds the smell of garbage mildly appealing. To an enquirer who knows this, we might offer the high salary as the reason why he took a job collecting garbage. But to a typical enquirer who doesn't know this, and who would imagine the smell to be an overriding disincentive, Tom's liking for the smell would be the most natural choice of the reason why he took the job. This explanatorily salient reason need not be necessary since Tom might well have taken the job if he'd been merely indifferent to the smell, and it need not be sufficient, since without the salary he wouldn't have taken the job. Nor is it Tom's strongest reason for taking the job.

One might think this reason misleading, and hence not genuinely explanatory for a typical enquirer, as it would lead her to suppose that Tom had a strong perverse attraction to the smell that led him to seek out a job where he could satisfy his unusual taste. According to this thought, the non-misleading, genuine, explanation would be to describe Tom's attitude to the smell as close to indifference, or to say something like "He actually doesn't mind the smell." This amounts to offering a richer notion of reason which includes mention of the strength. However, although a simple reason explanation might seem an unnatural response to "Why did Tom take the job?" it wouldn't be misleading. For to be told the agent likes the smell of garbage is sufficiently surprising for the

enquirer to find it a satisfactory explanation without thinking the reason has to be strong. And in general, I do not think it would be misleading to offer a single reason as explanation when there are other types of explanation which may be more appropriate in the context, such as explanations that state important facts concerning the agent's past. In a situation in which the enquirer is clearly requesting the explanatory rationalising reason, the reason that should be given is that of the agent's liking the smell of garbage.

From this example it can be seen how structurally similar examples can be constructed to show that any weak insufficient reason can be the reason (which explains) why X K 'd, relative to a suitable enquirer. Just take an enquirer who believes X has a strong disliking for the action kind (L -ing) that features in the reason in question, but is knowledgeable about X 's other reasons. Thus I take it that the criterion under consideration would not be satisfied by any reasons, and hence would not support the view that there are nonmotivating reasons.

V

Conclusion

I have looked at a number of approaches to providing criteria for a reason to be nonmotivating and argued that none that would be satisfied by any reasons would be regarded by proponents of nonmotivating reasons as plausible. Furthermore I think it is clear from those arguments that they would apply also to any proposed criterion formed by combining any of the individual criteria. Indeed some of the criteria considered were precisely such combinations. I am thus drawn to the view that there are no nonmotivating reasons. And from this I conclude that all rationalising reasons are reasons for which the agent acts, reasons that are causally relevant to, motivate, and explain the agent's action.¹³

Thus the terms 'motivating reason,' 'explanatory reason,' 'reason for which,' and 'causally relevant reason' turn out to be blunt instruments that do not distinguish among rationalising reasons, unless they are being used in a pragmatic sense to refer to the rationalising reasons of most

¹³ If a similar search for criteria of *motivating* reasons were equally unsuccessful, we would have to conclude that there are no motivating reasons either. This would force us to take the notions of nonmotivating reason and motivating reason not as exhausting the class of rationalising reasons, but as two incoherent categories. However, a search for criteria of motivating reasons would not lead us to the conclusion that there are no motivating reasons as there is at least one obvious criterion: *a reason is motivating if it is necessary.*

interest in the context. For those who have formulated claims purporting to distinguish nonpragmatically among rationalising reasons using these terms, my conclusion is revisionary. The ideas expressed in such claims were inscrutable at best. They might be clarified, however, by seeking to capture what was important about them in terms of some of the notions I have been considering as criteria of nonmotivating reasons, such as the strength, necessity and sufficiency of a reason, the time of its occurrence, whether it was consciously entertained, and how it entered into the agent's deliberation.¹⁴

REFERENCES

- Davidson, D. (1963). Actions, Reasons, and Causes. *Journal of Philosophy*, **60**, 685-700. Reprinted in *Essays on Actions and Events*, 3-19. Oxford: Clarendon Press 1980.
- Frankfurt, H. (1969). Alternate Possibilities and Moral Responsibility. *Journal of Philosophy*, **66**, 829-839. Reprinted in *Free Will*, ed. D. Pereboom, 156-166. Indianapolis: Hackett 1997.
- Latham, N. (1994). Causally Irrelevant Reasons and Action Solely from the Motive of Duty. *Journal of Philosophy*, **61**, 599-618.
- Mele, A. (1997). *Philosophy of Action*. Oxford: Oxford University Press.
- Ramsey, W., Stich, S. & Garon, J. (1990). Connectionism, Eliminativism, and the Future of Folk Psychology. *Philosophical Perspectives*, **4**, 499-533. Reprinted in *Mind Design II*, ed. J. Haugeland, 351-376. Cambridge, MA: MIT Press 1997.
- Wilson, G. (1997). Reasons as Causes for Action, *Contemporary Action Theory*, ed. G. Holmstrom-Hintikka & R. Tuomela. Dordrecht: Kluwer.

¹⁴ Thanks to Al Mele, Scott Sehon, Ralf Stoecker, and Sven Walter for helpful comments.