

**SENTIR-SE RESPONSÁVEL E INTEGRADO:
UM DESAFIO AO INTELLECTUALISMO ACERCA DE AÇÕES**

**SENTIRSE RESPONSABLE E INTEGRADO:
UN DESAFÍO AL INTELLECTUALISMO ACERCA DE LAS ACCIONES**

**FEELING RESPONSIBLE AND INTEGRATED:
A CHALLENGE TO INTELLECTUALISM ABOUT AGENCY**

Leonardo de Mello Ribeiro

Professor da Universidade Federal de Minas Gerais

Resumo: De acordo com uma tradição, que chamaremos de *intelectualista*, confere-se à reflexão um lugar proeminente na caracterização do agir racional (no sentido de agir por uma razão). Para essa tradição, existe uma distinção *qualitativa* a ser feita entre ações que são o resultado de processos reflexivos e ações que são simplesmente motivadas por pró-attitudes do tipo de desejos. Uma dificuldade recorrente com tal distinção qualitativa diz respeito às ações de respostas automáticas e imediatas que realizamos frequentemente na vida cotidiana. Como podem estas ser acomodadas no modelo intelectualista? A partir de uma discussão deste tipo de ações e de uma resposta intelectualista possível ao problema, nosso objetivo aqui é enfatizar uma dificuldade adicional associada àquela distinção qualitativa: uma dificuldade relacionada aos nossos *sentimentos* naturais de responsabilidade por nossas ações. Como resultado, sugeriremos que apenas uma caracterização do agir racional que sustente uma relação direta, “interna”, entre desejos e ações seria capaz de conferir sentido a tais sentimentos.

Palavras-chave: racionalidade prática; intelectualismo; ações automáticas; responsabilidade prática.

Resumen: De acuerdo con una tradición, que llamaremos intelectualista, se confiere a la reflexión un lugar prominente en la caracterización de la acción racional (en el sentido de actuar por una razón). Para esa tradición, existe una distinción cualitativa que debe ser hecha entre acciones que son el resultado de procesos reflexivos y acciones que son simplemente motivadas por pro-actitudes del tipo de los deseos. Una dificultad recurrente con tal distinción cualitativa se refiere a las acciones de respuestas automáticas inmediatas que realizamos frecuentemente en la vida cotidiana. ¿Cómo pueden ser estas acomodadas en el modelo intelectualista? A partir de una discusión de este tipo de acciones y de una respuesta intelectualista posible al problema, nuestro objetivo aquí es enfatizar una dificultad adicional asociada a aquella distinción cualitativa: una dificultad relacionada a nuestros sentimientos naturales de responsabilidad por nuestras acciones. Como resultado, sugerimos que apenas una caracterización de la

acción racional que sostenga una relación directa, “interna”, entre deseos y acciones sería capaz de conferir sentido a tales sentimientos.

Palabras clave: racionalidad práctica; intelectualismo; acciones automáticas; responsabilidad práctica.

Abstract: Reflection is put at the forefront of rational agency (in the sense of acting for a reason) by what we will call here ‘the intellectualist tradition’. According to this tradition, there is a sort of *qualitative* distinction between actions which are the result of reflection and those which are simply motivated by pro-attitudes like desires. One long-standing problem with such qualitative distinction is concerned with the so many swift, *quasi*-automatic actions we perform in everyday life. How can these be accommodated within the intellectualist framework? In the context of these actions and of providing an answer to that question, our aim here is to highlight another problem with that qualitative distinction: one which is related to our natural *feelings* of responsibility for our actions. In the end, we shall suggest that only an account of rational agency that holds an “internal”, straightforward relation between desires and actions would be able to make sense of those feelings.

Key-words: practical rationality; intellectualism; automatic actions; practical responsibility.

The Background. The story we are told about human agency by philosophers pertaining to what we will call here ‘the intellectualist tradition’ goes as follows:¹ the phenomenon of human agency evinces a problem posed by the simple fact that we, humans, are capable of reflection. As soon as we realise and exercise our reflective capacity, we *must* put (or cannot help putting) our pro-attitudes such as desires into question, endorsing or rejecting them.² Accordingly, this is the only way of our being the genuine authors of our actions, being rational (in the sense of acting for a reason, no matter what) and being responsible for them.³ At the heart of this story lies a *qualitative* distinction between the results of reflection (over desires) and the workings of unreflective desires. This distinction is exactly the distinction that is supposed to account for rational action or full-blooded human agency. Following a way of characterizing the issue usually associated with the Kantian tradition, we will hold that such a qualitative distinction makes the workings of unreflective desires *external* to the agent’s practical rationality.⁴ Unless such desires are subject to reflection and get

¹ Contemporary philosophers like Korsgaard and Scanlon would fall into the intellectualist tradition as we take it here.

² This terminology is borrowed from Davidson (1980). But while Davidson seems to prefer to use ‘pro-attitude’ as the general term for the whole category of motivating attitudes, we will prefer to follow the most recent trend in the contemporary philosophical literature of using ‘desire’ as the umbrella term for that same category.

³ Thus, ‘rational’ here, unless otherwise noted, should be understood as meaning ‘acting for a reason’ (which may refer either to a motivational or to a justificatory sense).

⁴ This is so because Kant said things such as this: “the inclinations themselves, as sources of needs, are so far from having an absolute worth, so as to make one wish

endorsed, they are all external to one's practical rationality. It is *not* that they might be taken as *irrational*. Rather, as they stand, they are not even candidates for practical rationality: they are *a-rational*.

Christine Korsgaard, one of the main contemporary representatives of this sort of view, summarizes those commitments in the following series of passages.⁵ She says that

The capacity for self-conscious reflection about our own actions confers on us a kind of authority over ourselves (...). (Korsgaard, 1996, p. 19-20)

And we have normative [i.e., rationality] problems because we are self-conscious rational animals, capable of reflection about what we ought to believe and to do. That is why the normative question can be raised in the first place: because even when we are inclined to believe that something is right and to some extent feel ourselves moved to do it we can still always ask: "But is this really true?" and "Must I really do this?" (...) To raise the normative question is to ask whether our more unreflective (...) beliefs and motives can withstand the test of reflection. (Ibid., p. 46-47)

If the problem is that our perceptions and desires might not withstand reflective scrutiny, then the solution is that they might. We need reasons because our impulses must be able to withstand reflective scrutiny. We have reasons if they do. (Ibid., p.93)

From this Korsgaard concludes:

"Reason" then means reflective success. So if I decide that my desire is a reason to act, I must decide that on reflection I endorse that desire. (Ibid., p.97)

to have them, that it must instead be the universal wish of every rational being to be altogether free from them." (*Groundwork*, 4:428, p.37) A (Kantian) philosopher like Korsgaard takes it to mean that, for Kant: "Anything outside of the will counts as an alien cause, including the desires and inclinations of the person. The free will must be entirely self-determining." (Korsgaard, 1996, p. 97) This captures nicely the sense of 'external' that we have in mind here, despite the fact that Kant himself is talking about morality. However, since morality is, for him, reduced to rationality, his claims would perhaps not be very far from the characterization of intellectualism we have put forward here, even though his detour is definitely different from ours. But this is a point that we cannot pursue here.

⁵ We will focus throughout our discussion mostly on Korsgaard's work because she has a fully developed account of the topic in question here as well as because of the complexity and sophistication with which she aims to tackle the issue.

And this is a verdict that she claims to borrow from a point by Kant:

He [i.e., Kant] says, “We cannot conceive of a reason which consciously responds to a bidding from the outside with respect to its judgments.” If the bidding from outside is desire, then his point is that the reflective mind must endorse the desire before it can act on it — it must say to itself that the desire is a reason. (Korsgaard, 1996, p. 80)

Reason and rational action are then reflective success, according to this picture. Fair enough. But we do not know how strong and precise such a criterion is until we understand more clearly what reflection amounts to and how it is exercised to produce action. To begin with, the reflective success criterion invites the following naïve and challenging question: must we *always* put our desires into question? A negative answer to this would seem, at first sight, to lead us to challenge that traditional intellectualist conception of human agency, according to which, as we have seen, reflection is at the forefront of rational action. It is the main aim of this article to provide reasons for challenging exactly this commitment of intellectualist accounts of agency, namely, that reflection is constitutive of human agency and rationality. As we will see, sometimes thinking about our actions as springing from the workings of unreflective desires in an integrated fashion with our practical capacities can be the only way of making sense of some important features of our evaluative life. And this can only be fully appreciated if we dismiss that *qualitative* distinction between the results of reflection and the workings of unreflective desire—or so we claim. Let us now see why.

The Problem. We can begin our discussion by asking: What about the so many swift, *quasi*-automatic actions we perform in everyday life? If reflection is necessary for practical rationality, how could we make room for those actions?

Here we should notice, first of all, that there certainly is a difference between those swift, *quasi*-automatic actions, which are intentional (although perhaps not fully consciously or deliberately undertaken), perfectly intelligible to us, and mere bodily movements (which we would not deem intentional at all and refuse

to include among our list of proper actions). But if so, there is *a fortiori* a distinction to be made between actions that at least *look* entirely rational to us as we act (at least in the sense that we do not interfere with their performance) and mere bodily movements. How can we identify the rational agent on the side of those swift, *quasi*-automatic actions? Would we be happy to say that the rational agent exists only insofar as she genuinely reflects? This would be, we are assuming here, an uninteresting answer. After all, we need a model of guidance of actions (that's what, after all, practical reasons are about); and a model that would fail to do exactly this thing most of the time—in our everyday engagements, which very often involve a dynamic, fast, heuristic dimension of our lives—cannot be a good one.⁶

This is the kind of problem that any intellectualist picture of rational agency like Korsgaard's has to face. By putting too much emphasis on the role of reflection in the characterization of rational action, they need to provide answers to those naïve, but equally challenging questions.

One way of trying to make an intellectualist view like Korsgaard's compatible with the kind of concern expressed in our question about those swift, *quasi*-automatic actions would be to provide a broader understanding of the nature of reflection at the same time that the criteria for conferring rationality on actions are weakened. One solution along these lines (the one that Korsgaard herself seems to endorse)⁷ would be to hold that someone does not need to endorse all her actions (and underlying desires) but just needs to be *disposed* to do so, i.e. to be able to avow statements like "I would have endorsed it", *retrospectively*.

This may well be true as far as it goes. But that dispositional or retrospective form of justification cannot be enough. The problem is that it seems to be at odds with something empirically evident already mentioned, viz., even when proper reflection does

⁶ See Railton (2006) for a series of lively examples which provide strong support for this claim about our practical lives.

⁷ Korsgaard says: "We do not always do what upon reflection we would do or even what upon reflection we have already decided to do. Reflection does not have irresistible power over us. But when we do reflect we cannot but think that we ought to do what on reflection we conclude we have reason to do. (Op. Cit., p.104) See also Scanlon (2000, p.47) for a similar point.

not take place our behaviour very often *looks* entirely rational: we look rational agents *as we act*—not only retrospectively. That is to say, Korsgaard still seems to need a distinction between retrospectively endorsing actions (and corresponding desires) and currently endorsing actions (and corresponding desires) *as we act*, even though the latter does not need to involve a fully conscious and deliberative standpoint.

At this moment, let us grant something to Korsgaard and the intellectualist tradition. She could say, very plausibly, that any human action encompasses some minimal level of reflection. Indeed, this may be taken as constitutive of human rational action. Korsgaard seems to come exactly to this when she says that:

An animal, whose desire is its will, is a wanton [i.e., a being that acts unreflectively and simply follows the strongest inclinations of the moment]. I am arguing here that a person cannot be like that, because of the reflective structure of human consciousness. A person must act on a reason, and so the person who acts like a wanton must be treating the desire of the moment as a reason. (Korsgaard, 1996, p.99, footnote)⁸

So, it seems, a human agent (or a person), unlike animals and wantons, is necessarily reflective whenever she acts properly. Even if she acts on a desire of the moment, she must be reflectively treating such a desire as reason to act. Now, if reflection is a pervasive feature of human agency, then, taking into consideration all the swift, *quasi*-automatic actions we perform in everyday life, the suggestion here could be that a form of *monitoring* of our actions is taking place in such dynamic contexts. Indeed, we seem to hold *tacitly* many “norms of conduct”, deep values, personal commitments, emotional attachments, etc., on which we act, and they seem to be activated as we act even though we are not fully conscious of them, deliberate about or clearly entertain them in our practical engagements.⁹ Monitoring, thus, seems to indicate some minimal level of reflection taking place, since monitoring involves our being able to step back immediately whenever something seems

⁸ Korsgaard makes it clear in this passage that she is using ‘wanton’ in the same specific sense as it appears in the work of Harry Frankfurt.

⁹ Again, Railton (2006) offers a detailed presentation of a bunch of interesting cases.

to us to take the wrong course. So, monitored actions do not seem to be totally unreflective (or at least not unqualifiedly unreflective).

This looks like a big hand given to Korsgaard and the intellectualist tradition, since now there is room for them to try to accommodate the problem we started with: they apparently could now say that reflection in the sense they require for rational action to occur could take place, at least in a minimal sense, in cases of monitoring. Given that our mechanisms for processing information coming from the external world and for performing complex cognitive tasks can be very fast and not clearly noticeable to our conscious perspective, it would seem that Korsgaard could try to remove the pressure coming from empirical reality on her theory by accepting that there can be a sort of higher-order monitoring perspective giving endorsement to the actions one is *currently* performing.

However, even granting that monitoring may involve reflection in a minimal sense, we can still doubt whether Korsgaard can really accommodate some aspects typical of monitoring into her view on practical rationality. For Korsgaard wants reflection to be *necessary* for rational action in a (robust) way that seems incompatible with some manifestations of that sense of monitoring we have just seen (and as it will become even clearer in a moment). Let us explain.

Reflection for Korsgaard, as we have seen from the previously quoted passages, is necessary in the sense that there must be endorsement of action through a process of filtering one's desires from a normative conception of oneself that takes place at a higher-order level. If desires pass the test, they get endorsement and the actions they bring about are rational. Thus, endorsement in this intellectualist sense favoured by Korsgaard is something that is supposed to reveal agential *authority*.¹⁰ As such, one's endorsements and decisions are, as it were, "epistemically clear" to the Korsgaardian agent: their *choices* must be "luminous" to them.¹¹ So,

¹⁰ Cf. Korsgaard, 1996, p.104: "We might say that the acting self concedes to the thinking self its right to government. (...) So the reflective structure of human consciousness establishes a relation here, a relation that we have to ourselves. And it is a relation not of mere power but rather of authority."

¹¹ To do justice to Korsgaard, she actually considers one aspect of this claim (Korsgaard, 1996, p.92; p.100; p.144) and, in one sense, denies that our minds are

a *mistake* occurring here seems *impossible*. Even if the Korsgaardian agent happens to have some false beliefs that lead her to make a choice to do something that she does not *really* endorse—that is, a wrong choice at the end of the day—, her *choice* itself, that is, its *content*, as it appears to the agent, must be infallibly accessed. This is something that cannot be alienated from a rational agent because it is a *necessary*, constitutive condition of rational action, for the intellectualist picture at stake. So, once the Korsgaardian agent has chosen a certain course of action, her choice is authoritative for herself because she cannot be mistaken about what choice it is.¹²

But now the problem is that not all manifestations of monitoring seem to be amenable to this *robust* sense of reflection and endorsement underlying the Korsgaardian intellectualist picture. After all, one interesting thing that the story about monitoring tells us and invites us to consider is that we may sometimes *not* be entirely sure why we acted in the way we did. In severe cases, we can even act for a consideration that was *not* the one we *thought* we did. Sometimes this is something that may be discovered (if it really can) only through some sort of retrospective

luminous or transparent to ourselves. She says that “some philosophers have supposed that this means that [the contents of] our minds are completely accessible to us—that we can always be certain what we are thinking and feeling and wanting—and so that introspection yields certain knowledge of the self.” And she thinks that this is not true. However, although she is willing to grant that there may be contents related to the workings of our minds which are not accessible to our self-conscious perspective, she does not seem to concede that such workings could *speak for the agent*, regardless of her being aware of and endorsing them. For Korsgaard, it seems, the denial that our minds are completely luminous reveals something about our psychology, but this does not seem have any relevant implication for her from a *normative* point of view and, as a result, for her understanding of agency. Whenever normative and agential questions are raised, it seems, the agent must be certain about her endorsements. And this is our point here about the intellectualist agent’s epistemic clearness and infallible access to her choices.

¹² Interestingly, Kant says about moral conscience that “an *erring* conscience is an absurdity. For while I can indeed be mistaken at times in my objective judgment as to whether something is a duty or not, I cannot be mistaken in my subjective judgment as to whether I have submitted it to my practical reason (here in its role as judge) for such a judgment; for if I could be mistaken in that, I would have made no practical judgment at all, and in that case there would be neither truth nor error.” (Kant, 1996, 6:401, p. 161)

story that we tell to ourselves or others (very often, that we tell to ourselves with the help of others). And here there is always the possibility that we may end up discovering that we pursued something for some reason that we were not aware of at the moment of action.¹³

So, as far as the monitoring story goes, the agent in such cases could have given some sort of endorsement (if at all) to a certain action, but not for the reason she thought she did. The “endorsement” of the action, if any, would not have been given, in that case, by the consideration that the agent thought she was entertaining at the moment of action. So, the agent was mistaken about her reasons. But this seems *impossible* for the Korsgaardian agent to make sense of—or at least if we really want to continue calling her deed ‘rational’. The Korsgaardian intellectualist agent can, of course, be mistaken about the reasons she thought she was following. However, the result seems to be that, in the end, it is not something that the agent *did*. Since the behaviour was not reflected or endorsed (even in the weakened monitoring sense), it fails to qualify altogether as *rational*. Therefore, the intellectualist picture does not seem to be able to avoid saying that endorsement must come from a robust understanding of the reflective, self-conscious access (and endorsement) of the rational self. If it does not come from this source, the action cannot be rational at all.¹⁴

¹³ This sort of retrospective justifying story is championed by Williams (1994, p.44ff.). Recent psychological research makes use of other (more “scientific”) methods which are not based on self-report (such as eye-tracking devices, semantic tasks, timing of behavioural and physiological responses, etc.). See, for example, the experiments described by Barh & Chartrand (1999); Bargh et al (2001) Fazio et al (1986); Nisbett & Wilson (1977); Ross & Nisbett (1991).

¹⁴ The following passage by Korsgaard seems to illustrate the point nicely: “willing is *self-conscious* causality, causality that operates in the light of reflection. To will is not just to be a cause, or even to allow an impulse in me to operate as a cause, but, so to speak, to consciously pick up the reins, and make *myself* the cause of what I do. And if I am to constitute *myself* as the cause of an action, then I must be able to distinguish between *my* causing the action and some desire or impulse that is ‘in me’ causing my body to act. I must be able to see *myself as* something that is distinct from any of my particular, first-order, impulses and motives, as the reflective standpoint in any case requires. Minimally, then, I am not the mere location of a causally effective desire but rather am the *agent* who acts *on* the desire.” (Korsgaard, 1996, p.227-228)

To give an example, let us suppose that, after having saved the life of an important political leader, Bill discovers through some sort of story (retrospective, cognitive psychological, social psychological, etc.) that he did it because of some benefit that he was expecting to gain from the politician for having saved his life, and *not* (as he thought it to be the case at the very moment of action, whether or not through some form of monitoring) that he was saving the political leader's life because it was his "moral duty". Now, after discovering this, Bill may very plausibly be shocked, reproach his attitude and take himself blameable for such egoistic behaviour.¹⁵

Now, supposing that the behaviour performed by Bill was motivated by the workings of unreflective desire, it would have failed to pass the test of rationality required by Korsgaard (since there would have been no endorsement from the agent himself) and, as such, it would not even qualify as an (rational) action of Bill at all. And it is easy to see how big this problem is for Korsgaard: her story is supposed to show *not* that endorsement is something that may happen or not in order to rationalize actions. Rather, it is supposed to show that it is something *necessary* for us to qualify as rational agents or to individuate rational actions. So, Korsgaard's intellectualist picture may start looking a bit arbitrary. For her, Bill's

¹⁵ Although the story here is only a fiction, the general pattern on which it is based is hardly a fiction. Recent psychological literature on the unconscious, automaticity, priming effects and situational psychology seems to confirm cases like Bill's. Basically, the recent literature on non-conscious processes attempts to confirm that such processes (which may range from simple representations and subliminal information processing to more complex tasks such as concept activation, goal activation, ordering of preferences, resolution of conflicts among attitudes, persistence in performing behavioural tasks in the face of obstacles, etc.) may be much more sophisticated than the traditional psychological literature supposed (with its main focus on the conscious and the interpretation of the non-conscious as merely disruptive of the conscious). See, for example, Barh & Chartrand (1999); Bargh et al (2001) Fazio et al (1986); Nisbett & Wilson (1977); Ross & Nisbett (1991). For an overview of the results of experiments and developments of this area of research in psychology, see Ferguson (2006). For the theoretical commitments and implications of this psychological literature, see Bargh & Morsella (2008). See also Evans & Over (1996) for a defence of the sort of dual-process theory of (rational) thinking that seems to reinforce and to be amenable to many claims in that psychological literature.

action *cannot* be rational, since *he*, the rational self, did *not* endorse the action.

The result is that the following empirical fact is still pressing Korsgaard: at the moment of action it made perfect sense to Bill the way he acted. He may learn later, for the reasons presented above, that the action in question was one that he would not approve of or endorse, but it simply seems to prove too much to conclude that it was not a (rational) action; after all, it made sense to him as *his* action, as he acted.

An alternative account and its Implications. In fact, we can do better than the Korsgaardian intellectualist picture and make sense of monitoring as a full-blown rational engagement, even if some mismatches like Bill's may happen. But in order to achieve this result, we have to dramatically weaken the intellectualist requirements of practical rationality. This is so because we have to understand rational action in a different way, leaving its contours, so to speak, "more elastic".

Let us call one such an alternative picture, which broadens (if not revises) the intellectualist concept of practical rationality, 'Humean'—since Humeans are said to propose a straightforward connection between the workings of desires and practical rationality.¹⁶ Humeans, as we take them here, can confer practical rationality in a minimal sense. In a Humean scenario, an action may qualify as being done for a reason insofar as it is motivated by a desire, even if this is not clearly consciously entertained or endorsed by the agent. Thus, given this very minimal criterion for conferring rationality to actions, that kind of mismatch that occurs in Bill (between what he entertains as his motivating and justifying reasons in the sort of minimal awareness characteristic of monitoring and what *really* motivates him) would not license us to call his action *non-rational* at all. His action, at the very moment in which it is performed, *looks* perfectly rational for him. And the fact

¹⁶ There is no claim here that this is Hume's view. We say 'Humean' given the more or less agreement in the contemporary philosophical literature that Humeans hold that connection. But the view might as well be called, it seems, 'Williamsonian', 'Davidsonian', 'Situationist' (given the tenets of situationism in social psychology), or perhaps even 'Archaic Greek' (as Bernard Williams might suggest).

that it stems from a desire (even if unreflective) is enough for the Humean to explain rational action even if only in a minimal sense.

Thus, the agent may, of course, be wrong about the desire on which he acted, about his motivations at the very moment of action, and about what justifies them. As soon as he discovers this, he may profoundly reproach himself, if he is against the course of action or the motivations for the action performed. But, for our Humean, there is no problem for making sense of rationality or acting for a reason here. The occurrence of mismatches in one's motivational profile is something very amenable to the sort of Humean story we have just told.

But now some may point out that there is now an air of paradox in such Humean scenario and that this is exactly what the intellectualist picture was trying to avoid: accordingly, we can be acting for a reason that we somehow did not consciously entertain or endorse, and did not even know about. This has now some costs. One of them is that the agent may now be blamed and be taken responsible for (causing) something for reasons that he was not aware of or endorsed. However, as we will try to show, this seems at the end of the day to be in favour of the Humean at least with respect to the sort of problem we have posed here. Let us see why.

We said that the Humean may profoundly regret having acted in the way he did, feel blameable, reproach himself, and so on, if he is against the action and the motivations for the action performed. So, even lacking full awareness of the motivations for the performance of the action, he may say retrospectively that his action was based on something he now thinks he had *no* (or *less than optimal*) reason (in a robustly normative or evaluative sense) to do. And by saying this he means two things: he condemns his own action and takes *full responsibility for having performed it*. Our point here now is that, if he wholeheartedly reacts in this way, he *could not* be an intellectualist Korsgaardian agent; he could only be a Humean (in the terms of our debate). The proper rational Korsgaardian agent could not *make sense* of having *reacted* in that way since her reasons must be *epistemically clear* to her, as we have seen, and she can only ultimately respond for these. Thus, as a result, it is not clear that Bill, from a Korsgaardian perspective, can now reproach himself for "his" performance, for behaving on the

basis of those (egoistic) motivations. He may, perhaps, somehow *regret* what happened. But in a sense, Bill, as a Korsgaardian agent, wants to say boldly that *he didn't do it*; that it was *not him*; that it was a non-rational behaviour motivated by the workings of *desire alone*. The Korsgaardian agent can only speak from the perspective of the *rational, fully endorsing self*, and give the following response for having behaved that way: "It was not me!"¹⁷

The reason why the Korsgaardian agent cannot make sense of *feeling* responsible for having acted in the way she did, in cases like Bill's, should now be clear and has to do with the criterion for rationality that the intellectualist picture of agency holds: the agent's action *fails* to qualify *as rational* (even in a minimal sense). For Korsgaard and the intellectualist tradition, the difference between acting rationally and acting motivated by the workings of unreflective desire is a *qualitative* one, as we have put it. This means that, in the end, the Korsgaardian agent can at most regret for not, say, following Reason (or her rational abilities) and for *what resulted from this failure*; she cannot really make sense of feeling sorry for *her* having acted on an egoistic basis.

Now we can see a decisive difference between the Humean and Korsgaardian pictures: the Humean takes our acting motivated on the basis of desires (even when there is not full awareness of their operation) as sufficient to make sense of action (as being done for a reason) or confer rationality (at least in a minimal sense) to the ensuing behaviour. In this way, we might say that, according to

¹⁷ In a similar fashion, Velleman (2000, 126-7) considers a case in which an agent's later reflection on a given (past) action of his (and its respective motivations) leads him to realize that "desires of mine caused a decision, which in turn caused the corresponding behaviour; and I may acknowledge that these mental states were thereby exerting their normal motivational force, unabatted by any strange perturbation or compulsion. But do I necessarily think that I made the decision or that I executed it? Surely, I can believe that the decision, though genuinely motivated by my desires, was thereby induced in me but not formed by me; and I can believe that it was genuinely executed in my behaviour but executed, again, without my help. Indeed, viewing the decision as directly motivated by my desires, and my behaviour as directly governed by the decision, is precisely what leads to the thought that as my words became more shrill, it was my resentment speaking, not I." Velleman's main point here is to suggest that a decision might take place without it being endorsed by the agent; and this has, for him, in a similar way to our characterization of intellectualism here, the result that the behaviour is not recognized as being the agent's own action.

the Humean picture of agency and rationality, pro-attitudes like desires are, in a sense, all *internal* to the agent. Not so for Korsgaardian agents: for these our mental life with respect to practical issues is, so to speak, “fractured”: there is the rational self and the blind workings of (unreflective) desire. The former must rule the second to guarantee rationality even in a minimal sense. So, there is no unqualified *integration* in the mind and actions of a Korsgaardian intellectualist agent. The workings of unreflective desires are, as a matter of principle, *external* to the Korsgaardian agent’s practical rationality. Desires get integrated into one’s practical rationality only if they get endorsed.¹⁸

Not so for the Humean. And now we can see in which sense this favours the Humean picture when it comes to offering a response to cases like Bill’s so as to make sense of an important aspect of our evaluative lives. The Humean agent, by taking his mental life filled up with desires whose workings might make sense of acting for a reason regardless of the full awareness of the “thinking” or endorsing self, feels integrated and sees no *qualitative* distancing from the desires he may happen (or be motivated) to act on, *even if they are unbeknownst to him*. So, he can truly feel sorry, disapprove, reproach, blame himself, etc. After considering more vividly the situation and realizing that he acted on a desire that he does not (deeply) favour, he still recognizes that *it was him who did it*. The action looked perfectly rational to him, since he *understands* all his actions as coming from the workings of his desires. That’s why the Humean can wholeheartedly say: “Yes, I did it; I’m terribly sorry for that.”

¹⁸ A description of such “fractured” nature of our agency can be found in the following passage by Korsgaard: “Although I have just been suggesting that we do make an active contribution to our practical identities and the impulses that arise from them, it remains true that at the moment of action these impulses are the incentives, the passively confronted material upon which the active will operates, and not the agent or active will itself. (...) The reason I must identify with my principle of choice when I act really has nothing to do with whether my first-order impulses seem totally alien to me or I regard them as my own productions. It is rather that at the moment of action I must identify with my principle of choice if I am to regard myself as the *agent* of the action at all.” (Korsgaard, 1996, p. 240-241)

For our Humean agent, an integrated view of his agency can encompass the workings of his unreflective desires, which he recognizes as internal to him. Thus, feeling responsible, integrated and blameable are not things that depend on the agent's full awareness and reflection on everything that motivates him,¹⁹ but only on his integrated view of his practical rationality. But for the Korsgaardian agent there is no such integration, no internality—the agent can only take responsibility and recognize that she deserves to be blamed (or not) for what *she* did on *full awareness* or by reflectively endorsing (or rejecting) desires: her criterion for rationality is completely *external* to the workings of unreflective desires. These are, as a matter of principle, external to her. The intellectualist agent is thus fractured, and will apparently continue so.²⁰

Referências

- BARH, J. A. & CHARTRAND, T. L. *The Unbearable Automaticity of Being*, *American Psychologist*, 54(7): 462-479. 1999.
- BARGH, J. A. , LEE-CHAI, A., BARNDOLLAR, K., GOLLWITZER, P. M., TRÖTSCHER, R. *The Automated Will: Nonconscious Activation*

¹⁹ Perhaps not even on his intentionally doing something—no matter how we characterize this; if either consciously or non-consciously doing—as Bernard Williams would probably like to say. See, for example, his ‘Moral Luck’ and his considerations about responsibility for doing something unintentionally or under “unusual” states of the self in chapter three of *Shame and Necessity*. He says, for example, that it is “a mistake to think that the idea of the voluntary can itself be refined beyond a certain point. (...) If we push beyond a certain point questions of what outcome, exactly, was intended, whether a state of mind was normal or whether the agent could at a certain moment have controlled himself, we sink into the sands of an everyday, entirely justified, skepticism.” (*Shame and Necessity*, p.67) And this advice may now make it intelligible to claim about certain cases that “(...) you may have deliberated as well as you could, but you still deeply regret that that was how the deliberation went, and that this was what you did. This is not just a regret about what happened, such as a spectator might have. It is an agent's regret, and it is in the nature of action that such regrets cannot be eliminated, that one's life could not be partitioned into some things that one does intentionally and other things that merely happen to one.” (Idem, p.70)

²⁰ I thank George Botterill for many discussions on the topic of this paper.

- and Pursuit of Behavioral Goals*, *Journal of Personality and Social Psychology*, 81(6): 1014–1027. 2001.
- BARGH, J. A. & MORSELLA, E. ‘*The Unconscious Mind*’, *Perspectives on Psychological Science*, 3(1): 73–79. 2008.
- DAVIDSON, D. *Actions, Reasons, and Causes*, In *Essays on Actions and Events*, Oxford: Clarendon Press. 1980.
- EVANS J. ST.B. T. & OVER, D. E. *Rationality and Reasoning*, Hove: Psychology Press. 1996.
- FAZIO, R. H., SANBONMATSU, D. M., POWELL, M. C., KARDES, F. R. ‘On the Automatic Activation of Attitudes’, *Journal of Personality and Social Psychology*, 50 (2): 229-238. 1986.
- FERGUSON, M. J. *The Automaticity of Evaluation*, In J. A. BARGH (Ed.), *Social Psychology and the Unconscious: The Automaticity of Higher Mental Processes*, New York: Psychology Press. 2006.
- KANT, I. *Groundwork of the Metaphysics of Morals*, translated and edited by Mary Gregor, Cambridge: Cambridge University Press. 1998.
- KANT, I. *The Metaphysics of Morals*, translated and edited by Mary Gregor, Cambridge: Cambridge University Press. 1996.
- KORSGAARD, C. M. *The Sources of Normativity*, Cambridge: Cambridge University Press. 1996.
- NISBETT, R. E. & WILSON, T. D. *Telling More Than We Can Know: Verbal Reports on Mental Processes*, *Psychological Review*, 84(3): 231-259. 1977.
- RAILTON, P. ‘Normative Guidance’, In R. Shafer-Landau (ed.), *Oxford Studies in Metaethics, Volume 1*, Oxford: Oxford University Press. 2006.
- ROSS L. & NISBETT R. E. *The Person and The Situation: Perspectives of Social Psychology*, London: Pinter & Martin Ltd. 1991/2011.
- SCANLON, T. *What We Owe To Each Other*, Cambridge, Mass.: Harvard University Press. 2000.
- VELLEMAN, J. D. *What Happens When Someone Acts?* In *The Possibility of Practical Reason*, Oxford: Clarendon Press. 2000.
- WILLIAMS, B. *Moral Luck*, In *Moral Luck: Philosophical Papers 1973-1980*. Cambridge: Cambridge University Press. 1981.
- WILLIAMS, B. *Shame and Necessity*, Berkeley: University of California Press. 1994.