The Authority of Reflection

Carla Bagnoli

ABSTRACT: This paper examines Moran’s argument for the special authority of the first-person, which revolves around the Self/Other asymmetry and grounds dichotomies such as the practical vs. theoretical, activity vs. passivity, and justificatory vs. explanatory reasons. These dichotomies qualify the self-reflective person as an agent, interested in justifying her actions from a deliberative stance. The Other is pictured as a spectator interested in explaining action from a theoretical stance. The self-reflective knower has authority over her own mental states, while the Spectator does not. I highlight the implications of this construal for a theory of action, and call attention onto some other interesting normative relations between the self-reflective agent and the Other that escape both the first-person and the third-person approach. My contention is that the authority of self-reflection (and of reason) is best understood as a relation of mutual recognition between self and others, hence from a second-person stance.

Keywords: Richard Moran, reflexivity, first-person authority, second-person standpoint, mutual recognition.

One of the central claims of Richard Moran’s Authority and Estrangement is that avowability is fundamental to self-knowledge, and its special importance is to be understood in practical terms, that is, as a kind of authority. To refine this claim, in Chapter IV, Moran invokes a specific account of self-consciousness according to which the self-reflective agent exerts a special kind of authority on her own mental life. This kind of authority is fundamentally first-personal and, under this reading, it is conceptually linked to (or even identified with) rational freedom.

The purpose of this paper is to assess the merits and limitations of Moran’s first-personal approach to the authority of the self-reflective agent. Its uttermost merit is that it relates some modes of “knowing oneself” to the modes of “making up oneself”, hence treats self-knowledge as a form of self-constitution. In this light, the nature of self-knowledge becomes a practical rather than a theoretical matter, and more precisely a deliberative matter, the domain of choice and decision. In section §1, I will examine Moran’s argument for the special authority of the first-person, which I will refer to as “the argument of double suspension”. When I reflect about my own mental states I can suspend both their normative and their psychological force. The aim of the argument from double suspension is to prove that there are irreducible asymmetries between claims in the first and in the third person. It is not simply that we are aware of our own mental states in a special way (for instance, directly and immediately) or that we exert a special kind of control over our own mental states that cannot be extended to the mental states of others (except perhaps indirectly). Rather, claims in the first person have a special authority and bear a special relation to the agent. In section §2, I will show that the Self/Other asymmetry grounds other important dichotomies such as the practical vs. theoretical, activity vs. passivity, and justificatory vs. explanatory reasons. These dichotomies qualify the self-reflective person as an
agent, interested in justifying her actions from a deliberative stance. The Other is pictured as a spectator interested in explaining the actions from a theoretical or contemplative stance. The self-reflective knower has authority over her own mental states, while the Spectator does not. I will highlight the implications of this construal for a theory of action, and call attention onto some other interesting normative relations between the self-reflective agent and the Other that escape both the first-person and the third-person approach. In section §3, I pursue further the idea, and suggest that the authority of self-reflection (and of reason) is best understood as a relation of mutual recognition between self and others, hence from a second-person stance. I argue that normative relations of mutual recognition between an agent and another are central modes of self-constitution. My argument is that the authority of reflection (and of reason) is fundamentally second-personal, based on mutual recognition. This is not to deny that there are interesting and important asymmetries between the perspective of the agent and the perspective of a spectator. Such asymmetries, however, should be investigated against the background of a deeper and more fundamental symmetrical relation of recognition of authority.

1. The argument of double suspension and the Self/Other Asymmetry

Claims to knowledge in the first-person have a distinct and special kind of authority, Moran holds. This special relation one entertains to her own mental states is apparent when we consider the effect of self-reflection. To illustrate reflexivity, Moran deploys the metaphor of stepping back. Distancing implies bringing into view, separating from, confronting and facing (Moran 2001: 143). Suppose I feel angry about a social disservice, and reflect about my own attitude. Reflection requires that I consider my own attitude at a distance. As I step back from it, I halt it. This is something that I can do only insofar as anger is “mine”. The normative capacity to attribute or withdraw recognition to a given mental state is where the mind resides.

Moran takes the suspension to be both psychological and normative, hence concerning the legitimacy and authority of the mental state. But he insists on the psychological aspect of suspension as most clearly distinctive of the first-person. One can be aware of the states of another, and judge them inappropriate or appropriate. This normative assessment may not have any, and certainly does not have any direct effect on such a mental state. I cannot directly and immediately alter or stop another’s mental states simply by reflecting on it. I may exert some contrasting authority on another’s mental state. For example, I may question the legitimacy of your anger by questioning its status or by offering you contrasting reasons not to be angry. This exercise of authority, however, does not precisely amount to suspending the authority of your mental state. I simply confront your claim by making a claim myself. This is a clash between distinct sources of authority. It could be that this confrontation has a special effect on your mind, such that you are persuaded to change your mind. But this effect cannot be equated to the suspension of the normative force of a mental state that the self-reflective agent reaches by self-reflection.
The argument thus highlights an important asymmetry between claims in the first-person (the agent's perspective) and claims in the third-person (the spectator's perspective). Reflecting on my own mental state allows me not only to hold it in judgment, but also to decide whether to have it or not. Self-knowledge distinctively involves avowal, decision, and choice. Hence, the authority of the first person cannot be explained in terms of theoretical awareness. It is not the claim that one knows oneself better, through a privileged epistemic access point, or with an immediacy that our knowledge of others' mental states lacks. The claim is that the agent who reflects about herself sees her mental states as options that she can choose: they are in an important sense “up to her”. The reflective self is thus inescapably a self-constituting agent; she cannot be a mere bystander (Moran 2001: 142, Korsgaard 1996a: 200-201).

2. Self-Reflection and the Deliberative Stance

The argument from suspension throws some light on the relation between self-consciousness and rational freedom (Moran 2001: 144). When the self calls into question and halts its own mental state, it proves to be free. When I reflect on my anger, I am free to choose whether to act on it or not. Reflection shows that the occurring mental state does not dominate me. More precisely, reflection establishes that there is a perspective we may call I from which mental states can be not only known and assessed, but also and more importantly, avowed. Self-reflection thus introduces a gap between my occurring mental states and my self. This gap shows that I am not simply determined by the desires and beliefs I happen to have. I can choose whether to have them.

The two aspects of suspension introduced by self-reflection may come apart. Disavowal may not entail halting. I can disavow my anger as I recover some overriding considerations not to be angry; and yet such reasons may not alter my anger. The normative suspension may not determine a psychological suspension. If this happens, however, my action is not “fully mine”. I am merely moved by a feeling that I do not control. I contemplate the forces that operate on my mind and cause my movements, without having any authority over them. These are forms of alienation and estrange-

---

1 "In so far as it no longer dominates me I am not simply free to appraise it, but also free to choose whether this shall be something I act upon or not. For, of course, I remain free to approve or disapprove of an impulse that does not nonetheless continue to dominate me, just as I may with respect to anyone else's inclinations or attitudes. There is nothing peculiarly reflective or first-personal about the ability to make normative assessment of some perception or impulse. But when I “suspend” the force of some impulse in this latter effective sense, my stance toward it makes a difference to what happens, and I can actually refrain from proceeding with it", Moran 2001: 144. Compare Frankfurt: "A person's will may be overpowered and violated by forces, such as those of anxiety or addiction, that are generated within him but are nonetheless not in the fullest sense his own. They are forces with which he does not identify, whose influence he struggles to resist. When they dominate, he is dominated by them; he is not in control of himself", Frankfurt 1988: 183.

2 To the extent that self-reflection involves this gap, Moran's account seems to me very close to Frankfurt and Korsgaard, and vulnerable to similar problems. I have examined this issue concerning the modes of self-reflection and deliberation in Bagnoli 2007, chapters 4 and 5.
ment, cases in which the self has lost its first-person authority. When my reflection is psychologically inert, I lose my privilege as an agent and become as passive as a spectator toward my mental life.

A spectator can account for the agent’s action performance third-personally, that is, by reviewing the reasons that the agent might have endorsed and that might have caused the action. But only the agent can account for her action as performed under a given description and because of such and such reasons. In this respect, the position of the agent regarding her action, and thus regarding the knowledge of the action’s true description, is fundamentally different from that of any other person (Moran 2001: 128); she is the only one who can endorse a reason for action as a reason for action, that is, as a reason that guides her in considering what to do and that will be part of the true description of the action. This means also that the person’s relation to her actions and attitudes “must express the priority of justifying reasons over purely explanatory ones” (Moran 2001: 128).

To privilege a deliberative perspective and identify it with the domain of reason and freedom has been a common strategy among philosophers of Kantian provenance. Self-reflection makes the divide between the practical and the theoretical stance. On a Kantian view, this means that self-reflection opens up the space of reasons (Kant 1785: 431-432, Korsgaard 1996a: 113). Kantians insist on the first-person or deliberative perspective to define the domain of the normative as opposed to the descriptive. But for Moran “there is nothing peculiarly reflective or first-personal about the ability to make normative assessment of some perception or impulse” (Moran 2001: 144); it’s the psychological suspension that makes the difference. The mechanisms by which the authority of the first person is exercised such as avowal and disavowal are not to be confounded with disapproval, approval or normative assessment, which are not distinctively first-personal attitudes. Disavowal is a way of disowning some mental state as external, and thus distancing and dissociating oneself

---

3 The idea is that authority is first-personal “not in the sense that various forms of alienation are not possible, but that they are possibilities as forms of alienation”, Moran 2001: 144. Compare Frankfurt: the upshot of deliberation is a “radical separation of the competing desires, one of which is not merely assigned a relatively less favored position, but extruded entirely as an outlaw” (Frankfurt 1988: 170); see also Frankfurt 1988: 63, 66-67; Frankfurt 2001: 11; Frankfurt 1988: 172; Frankfurt 1999: 136. The aim of this strategy is not so much to resolve the conflict by annulling one desire as to produce a “well-ordered self” by removing the internal obstacle. This aim is achieved by altering the nature of the conflict: once one of the conflicting desires is disavowed, there would be no inner division. It is an alien force not because it is an irresistible force, but because it has been disowned by deliberation. On the contrast between power (psychological pull) and authority (normative force), see Frankfurt 1999: 139.

4 I must confess I do not understand why Moran talks about the “priority” of justifying reasons. I think the argument does not establish the priority of the justifying reasons over the explanatory reasons. It establishes the distinction between these two sorts of reasons, hence the distinction between the deliberative and the theoretical stance. In fact, this is all we need to show the self/other asymmetry.

5 “Shall I?” is a deliberative question; it’s different from a theoretical-predictive “Will I?”, but it’s also different from the practical-normative “Should-Ought I?”. 
from it. Hence, disavowal is not simply a disclaimer; it’s an act of choice determining withdrawal of ownership and authorship.6

What do we gain in identifying the peculiarity of the first-person with the deliberative stance? First of all, we gain access to a more specific domain of the normative, that is, the domain of justifying reasons. Justifying reasons are reasons that I endorse as I deliberate about what to do, what to feel or what to believe. This kind of reasons is not available in the third-person perspective, when I consider the matter as a spectator. In the third-personal perspective, Moran argues, reasons can only be explanatory: they amount to rationalizations of why the agent acted in a certain manner. They trace back the causes of the agent’s action, rather than giving the agent considerations upon which to act.

The stance of rational agency is deliberative or practical rather than explanatory or theoretical. The agent who reflects “is asking not what must best explain the movement that constitutes the agent’s action, but instead he is asking for the reasons he takes to justify his action” (Moran 2001: 127). If the agent did not undertake this practical stance, and conduct reflection in the manner described, there would be no point for him in deliberating about what to do or what to believe (Moran 2001: 127). This ultimately is the capacity to make up one’s mind (Moran 2001: 134).

Granted that the first-person perspective is deliberative, its construal in terms of justifying reasons involves a mischaracterization of the third-person perspective. The third-person perspective is not always cast so as to renounce justification. On the contrary, it is often proposed as a theory of justifying reasons for action (Nagel 1970). As a perspective on reasons for action, the third-person approach takes reasons to be irresistible evidences irrespective of any relation they entertain with the agent; call this the Impersonality claim. In virtue of this claim, anyone, being in the agent’s position or not, ought to recognize the same kinds of reasons.

I call attention onto this alternative construal of the third-person approach to show that the third-person perspective does not necessarily map onto the theoretical or contemplative perspective, nor does it coincide with the domain of explanatory reasons, as Moran’s discussion assumes. The issue at stake in contrasting the first-person and the third-person approach to reasons (for action) is whether they bear an intrinsic and irreducible connection to the agent. According to Thomas Nagel, for example, they do not: whether reasons are produced or discovered by the agent, they can always fundamentally be translated into third personal judgments without any loss in meaning (Nagel 1970: 100-102). Reasons for action, as well as reasons for belief, can be presented as evidences. The agent is not in any special position in regard to them; in fact, in order to perceive them as reasons, she has to lose any sense of her privilege; she has to consider herself in no relevant way different from an impartial spectator. On this reading, however, representing oneself as an impartial spectator does not coincide with playing the role of a “by-stander”; rather, it is the representational device that rules out special pleading and warrants impartiality in the content of the reasons.

---

This alternative characterization does not leave the third-personal account in any safer place regarding the issue of authority and authorship on action. It remains a question how the reasons so conceived are authoritative for the agent, and how one can conceive of action if not at the same time by conceiving of oneself as its author (Kant 1785: 431). However, by presenting the third-person perspective as a theory for action, we are in a better position to identify its shortcomings. We can reject the third-person perspective not because it is the bystander’s view, disinterested in justifying action, but because it treats justification as a matter of empirical discovery, intellectual insight or epistemic recognition, rather than as a matter of choice.

Philosophers dissatisfied with the third-personal account of reasons for action have turned to the deliberative perspective primarily to make sense of the bindingness of reasons. They are interested in making a conceptual connection between “There is a reason $X$ to $\phi$” to “I have a reason $X$ to $\phi$”. The claim of impersonality of the third-person approach gets in the way as we try to account for this conceptual link: it makes reasons either inert or mysteriously powerful. But this counts as a failure only on the assumption that reasons for action should be conceptually linked to agency in this manner.

In warning us against the risks of overlooking the asymmetry between the first-person and the third-person stance, Moran is not so much preoccupied with the bindingness of reasons as with responsibility. The third-personal claim of impersonality is an evasion of responsibility. I welcome this shift as a profound insight: the core issue here is not how to make reasons efficacious, but how to make responsible choices. The deliberative stance must have priority if self-knowledge is a matter of taking responsibility for oneself (for one’s actions, beliefs, and attitudes). The phenomenology of choice supports this reading. The pain of failure in carrying out my commitment to an action I avowed cannot be easily relieved by thinking (third-personally) that the action could be otherwise carried out or that the desired state of affairs could be brought about by another agent who is supposed to have the very same reason to act. It matters that we behave according to how we judge. It matters to us. By focusing on the relation between agent and action, the deliberative view accounts in a distinctive way for the importance of acting according to our best judgment. Judgments are not simply expressions of our view of the matter, they are also and more importantly carriers of our identity. We care about acting according to our judgment because we care about ourselves: we want to make a difference in the world and we want to do it in a way expressive of the kind of persons we are. Identity and autonomy are thereby strongly connected, if not coincident. We can account for categories such as freedom and choice on the basis of the self-constitutive capacity of the reflective self, which makes choice unavoidable and inescapable.7

---

7 “Choice and consciousness are one and the same thing… to be conscious of ourselves and to choose ourselves are one and the same”, Moran 2001: 140.
3. The Authority of Reflection and the Second-Personal Approach to Deliberation

The claim has been that first-person reflection entails special responsibilities because it is the expression of a capacity that is exclusively first-personal, that of making up oneself. In this final section, I reconsider the adequacy of the first-person stance to fully express this claim. My proposal is that we conceive the deliberative perspective as second-personal, rather than first-personal. This is not in order to deny or downplay the asymmetry that Moran highlights. I do not dispute that the reflective-self has a special authority on her own mental states which differs from an external authority. Such a special authority remains in place even when it is challenged by another source of authority. For example, I retain a first-person authority on my belief that I love my brother even when I am faced with definitive evidence that I have no brother or that I do and feel nothing that is ordinarily associated to loving one’s brother. Likewise, I retain authority on my feeling pain because my left arm hurts even when the doctor proves me that I haven’t got a left arm. In those cases, doctors “know better” and their knowledge is better grounded than mine, but this does not affect my epistemic state: I still have the privilege to avow my own mental states. I still retain the option of avowing them. This is a privilege I continue to have insofar as I reason in the first-person mode. I may become alienated and thus lose authority on my self, but typically the healthy self has a relation of authority with itself.8

To fully appreciate this claim, I contend, we have to move toward the second-person stance. We typically acquire justificatory reasons via reflection. The structure of reflection is second-personal. The practice of self-reflection is also fundamentally second-personal, the activity of a self addressing itself.9 Even when performed in solitude, as a monologue, this activity is always done in front of an audience. The double suspension celebrated by Moran as a first-person privilege is an achievement that is possible when the self addresses itself. It is the product of a second-personal address.

What does this self-to-self address concern? As Moran suggests in accounting for the normative aspect of the double suspension, the address is an exchange of reasons.10 Of course, the self may be deaf to such reasons, but this means to be self-estranged and alienated. The healthy self claims ownership on his mental states and authorship on action because it follows its own reasons. Through reflection we gain

8 I say “typically” because I think there are some (perhaps paradoxical) cases where alienation is a healthy response to a tragic choice.

9 This is first noted by Kant, although he is generally invoked to support the first-person approach to deliberation. When describing the working of “conscience”, which is one paradigmatic exercise of self-reflection, Kant notices that the “A man who accuses and judges himself in conscience must think of a dual personality in himself, a doubled (doppelte) self…”, Kant 1797: 438. This raises the problem as to how the agent can bide herself. The dialogical model is an attempt to respond to this question by invoking the notion of mutual recognition. I develop this view, and uncover its Kantian roots in Baghelli 2007.

10 “For an agent to conceive himself as capable of forming an intention and implement it (...) he must take his intentional actions to be determined by reasons, and thus he is in the position to know a true description of his action in knowing his reasons”, Moran 2001: 127.
not only authorship on action and ownership on states of mind, but also and most importantly justification based on reasons. This is a crucial aspect of reflection as a key mode of self-constitution in that it differentiates reflection guided by reasons from other forms of rationalization which are typical of alienation.

This remark invites us to take another step toward a second-personal account, and notice that the structure of justification is also second-personal. Reasons are considerations offered to another to justify a course of action or a mental state. My proposal is that we understand the second-personal structure according to a dialogical model. “I shall” is the conclusion of a dialogue that emerges from the recognition that I should account for my actions to others, and that I should demand justifications from them. I put myself under the rational scrutiny of others and demand the same. What I offer as a reason must count as a reason also for others, and what others offer as a reason must be something intelligible to me as a reason.

On this approach, first-person authority is not a solitary achievement. It is an achievement that we are capable of insofar as we relate to others and to ourselves as self-originating sources of legitimate claims. I have authorship on myself insofar as I have a relation of respect toward others. This kind of dialogical deliberation presupposes that there is a community of equals that I recognize and of which I am a member. Authority is the result of mutual binding. What makes something authoritative and binding is not the endorsement by individuals in the first-person mode, nor is it the attribution of authority in the third-person mode (e.g. the recognition on behalf of the concrete community which administers norms and sanctions). What grounds the community as well as its members as normative agencies, and consequently what grounds their claims, is simultaneous and mutual recognition.

This is not because, the communal practices are prior to the constitution of the self, as communitarians hold. The communitarian thesis opposes the deliberative model because of the priority it accords to the self, but it is equally partial and misleading. The alternative I am sketching consists in elucidating the implications of taking self-reflection to be guided by reasons. While I argue that self-reflection and self-constitution are fundamentally second-personal, I do not mean that they presuppose or depend on actual practices of recognition. To stress the second-personal character of this achievement does not commit one to the priority of the community; and to argue that the modes of self-knowing are the modes of self-constitution does not commit one to the priority of the self. Rather, it encourages us to investigate further how the self claims authorship on action. My contention is that in claiming authorship on one’s self, one represents oneself as a member of a community of equals.

---

11 For a full-fledged dialogical account, see Bagnoli 2007. There are alternative models of the second-personality of moral judgments, compare Brandom 1999, Honneth 1992, and Darwall 2006. Korsgaard’s view on reciprocity can also be seen as a theory of recognition, but with some important qualifications, see Korsgaard 1996b, compare Bagnoli 2007, chapter II.

12 This is generally presented as Hegel’s thesis, see Brandom 1999, p. 219. I believe that this part of Hegel’s conception of recognition is already present in Kant’s conception of respect; I argue for this claim in Bagnoli 2007, chapter I.
There is much to gain in treating the phenomenon of first-person authority according to the dialogical model. Reference to an ideal community of equals gives us a diagnostic test for our social vulnerability and mutual dependence, which would be a significant complement to Moran's account of alienation and estrangement. We can thus appreciate that our first-person authority can be shaken and fundamentally undermined from the outside, through practices of discrimination and withdrawal of membership. Our first-person authority is vulnerable to Others exactly because they are not mere bystanders. They are our interlocutors. Relating to oneself authoritatively is a form of achievement that consists in healthy relations with others.

The claim that authority depends on mutual recognition bears important consequences, which I can only mention here. First, it shows the social nature of the self. This means not only that the individual self is formed through practices of social recognition, but also and most importantly that the self's authority on itself cannot be understood independently of how it relates to other selves: they are instituted simultaneously and reciprocally limit their authority. Second, if self-reflection is understood in second-personal terms, it is not reducible to a sequence of individual, distinct and discrete acts of avowal: it needs an audience and displays a history. Finally, and perhaps most importantly, the second-personal stance understands rational freedom not as a metaphysical property of the self, but as a practical relation structured by mutual recognition.

This emphasis on the relationality of freedom and the historicity of self-reflection does not shift the burden of choice onto others (or onto social institutions and practices). This is an important remark because Moran's main rationale for undertaking a first-personal approach to the deliberative stance is to present claims of self-knowledge as modes of undertaking responsibility for ourselves. Likewise, in the second-personal model, we cannot elude the responsibilities associated with being a reflective self; we cannot but choose where we stand. Freedom remains inescapable and burdensome. But it is a burden we are bound to share with others.

As Brandom writes: “Both selves and communities are normative structures instituted by reciprocal recognition. (...) Merely biological beings, subjects and objects of desires, become spiritual beings, undertakers (and attributors) of commitments, by being at once the subjects and the objects of recognition attitudes. At the same time and by the same means that selves, in this normative sense, are synthesized, so are communities, as structured wholes of selves all of whom recognize and are recognized by each other”, Brandom 1999: p. 217.

This is a revised version of a paper presented at the XVI Conference on Philosophy and Cognitive Sciences at the University of Valencia, 27th March 2006. I would like to thank Richard Moran for his response to this comment, which prompted me to revise the paper. I am also very grateful to Josep Corbí, Luca Ferrero, Gianfranco Soldati and most of all to Angus Ross, for their illuminating comments and criticisms.
REFERENCES


Carla Bagnoli is Associate Professor of Philosophy at the University of Wisconsin. She does research on ethical theory and practical reasoning, and particularly on issues related to the authority and objectivity of ethics. Among her recent publications: *Dilemmi morali*, "Breaking Ties" (*Dialectica*, 2006); "Humanitarian Intervention as a Perfect Duty" (*Nemos*, 2004); "Il costruttivismo kantiano" (in *Ragioni morali*, ETS, 2004); "Respect and Loving Attention" (*Canadian Journal of Philosophy*, 2003); "Moral Constructivism" (*Topoi*, 2002); "Value in the guise of regret" (*Philosophical Explorations*, 2000). She edited Robert Nozick (*Croatian Journal of Philosophy*, 2004); she co-edited *Meaning, Justification, and Reasons* (*Topoi*, 2002), and John Rawls (*Croatian Journal of Philosophy*, 2001). She is Editor in Chief of the *European Journal of Analytic Philosophy*.

Address: Department of Philosophy, 3243 N Downer Ave., Milwaukee, 53211 Wisconsin, USA. E-mail: cbagnoli@uwm.edu.