J. Indian Counc. Philos. Res.
https://doi.org/10.1007/s40961-018-0142-1

REVIEW ARTICLE

A Critique of Alfred R Mele's Work on *Autonomous Agents: From Self-Control to Autonomy*

Oxford University Press, New York, 1995

Pujarini Das[1]

Received: 26 March 2016 / Revised: 7 October 2016 / Accepted: 22 March 2018
© ICPR 2018

Introduction

The book, *Autonomous Agents: From Self-Control to Autonomy* (1995), by Alfred R. Mele, deals primarily with two main concepts, "self-control" and "individual autonomy," and the relationship between them. The book is divided into two parts: (1) a view of self-control, the self-controlled person, and behaviour manifesting self-control, and (2) a view of personal autonomy, the autonomous person, and autonomous behaviour. Mele (Ibid.) defines self-control as the opposite of the Aristotelian concept of *akrasia*, or the contrary of *akrasia*, which implies weakness of will, incontinence, or lack of self-control—the state of mind in which one acts against one's better judgement. According to Mele, the concept of self-control can be approached from two perspectives: (a) how self-control affects human behaviour, and (b) how self-control-associated behavior can enhance our understanding of 'personal autonomy' and 'autonomous behaviour'—personal autonomy requires self-control, and autonomous persons and autonomous behaviour are naturally found together. Therefore, I might say that self-control is essential to enhancing one's autonomy.

In part I, we find an account of self-control where Mele argues that even an ideally self-controlled person might lack autonomy. In part II, Mele gives an explicit account of autonomy and explains what must be added to self-control to achieve autonomy. This is the pivotal claim made by Mele (dismantling the intuitively connected notions of self-control and autonomy).

✉ Pujarini Das
pujarini.hcu@gmail.com; pujarini@iitk.ac.in

1 Department of Humanities and Social Sciences, Indian Institute of Technology Kanpur, Kanpur, U.P. 208016, India

Published online: 24 April 2018

Part I

The concepts of 'self-control' and 'autonomy' are used in a variety of senses. The term "autonomy" is derived from the two words: *autos*, which means 'self,' and *nomos*, which means 'rule' or 'law.' Thus, etymologically *autonomy* entails 'self-rule' or 'self-government,' which seems to inherently imply self-control. From a metaphysical perspective, autonomy might be understood in terms of free will, free action, or free choice. Joel (1986, 28) states that 'the meaning of autonomy,' when applied to individual persons can be understood in a variety of ways, including (1) the capacity to govern oneself, (2) the actual condition of self-government, (3) an ideal character trait (a virtue) derived from that conception, or (4) the sovereign authority to govern oneself. All of these conceptions seem to imply self-control in some sense, either as a capacity, a condition, or a right.

A self-controlled person has some sort of personal autonomy and can be referred to an autonomous agent. Mele discusses what it means to be an autonomous person and to act autonomously, and the issue of whether the autonomous agency is open to us. In this context, he considers both compatibilist and incompatibilist approaches. Mele does not reconcile these viewpoints to understand the importance of personal autonomy.

By stipulative definition, 'autonomy' is a property possessed solely by 'moral agents' who engage in moral conduct. In the *Nicomachean Ethics* (Ross 1915, hereafter NE, 10.7), Aristotle (350 B.C.E) refers to hypothetical gods as moral agents whose moral virtues are *otiose*—serving no practical purpose. Mele applies the term 'autonomy' interchangeably with the concept of 'freedom'. He has no theoretical need for an account of 'autonomy' that stipulatively limits the possession of autonomy to 'moral agents.' Just as there is no need to assert the existence of the will in arguing for the reality of 'weakness of the will,' it is similarly unnecessary to presuppose the existence of self-control in order to establish the reality of autonomous agency. He argues that even the person with ideal self-control can fall short of achieving of personal autonomy. The common notion that individual autonomy implies self-control may be more nominal than substantive, and depends on the extent to which its proponents share a conception of self-control.

For Mele, the account of self-control includes *continent* and *incontinent* behavior. With regard to the question of the continent, in chapter 7 of *De Motu Animalium* (NE, 701a7–8), Aristotle asks that "How does it happen that thinking is sometimes followed by action and sometimes not, sometimes by motion, sometimes not?" More specifically, psychiatrist Ainslie (1975, 463) contemplates, "why we sometimes pursue the poorer, smaller, or more disastrous of two alternative rewards even when we seem to be completely familiar with the alternatives, and also why we seek the better or larger reward when we act." Even in *Delay of Gratification*, psychologist Walter Mischel (1973) asks the same question.

The recent psychological literature on 'self-control' has been extensive and motivated both by *theoretical concerns*, e.g., delayed gratification, impulsive conduct, the motivational springs of action, and the place of cognition in the explanation of motivated behavior, and by the *practical problems*, e.g., drug abuse,

alcoholism, eating disorders, smoking, spending, phobias. We have seen that for example, how 'theory informs practice' and is informed by observations of the results of practice. It seems to me that we need to create a balance between self-monitoring (the act of observing or supervising something) and self-regulating (the act of self-activating or self-limiting) with respect to any act.

For Aristotle, *akrasia* is uncompelled, intentional behavior that acts against the agent's *best* or *better judgment* (judgment to the effect that it would be best or better to do A than to do B). *Enkrateia* means self-control, continence, and strength of will—behavior that conforms to one's best or better judgment in the face of temptation to act to the contrary. Aristotle (NE, 1150 a11–13) states that "the *akratic* person is in such a state as to be overcome even by those pleasures which most people master," while a self-controlled person is able "to master even those by which most people are defeated." Mele, like Aristotle, believes that self-control and *akrasia* are two sides of the same coin. He views self-controlled individuals as agents who possess both significant motivation to conduct themselves as they judge best, and a strong capacity to act in the face of actual or anticipated competing motivation. *Akratic* individuals suffer from a deficiency in one or both of these connections (1995, 5). Mele considers the example of someone who is relatively self-controlled in a certain sphere of life. In a particularly taxing situation, such a person might succumb to temptation in that sphere, (*A-ing*) intentionally and in the absence of compulsion, although he judges it best not to do so. Such behavior is considered *akratic* or incontinent, even if it manifests, not weakness of will or possessing subnormal powers of self-control (only an associated imperfection) (1987, 4). Likewise, an agent who suffers from *akrasia* may successfully exercise his modest powers of self-control and act in a continent manner.

Mele, unlike Aristotle, contends that agents may exhibit self-control not only over overt actions that correspond to their evaluative judgments—which include judgments having nothing to do with bodily appetites—but also in their acquisition or retention of beliefs. "Just as agents can act *akratically*, they can believe *akratically*, as in some cases of self-deception" (Ibid., Ch. 8). Aristotle views the continent or self-controlled agent as a person whose "desiring element" is 'obedient' to his 'reason or rational principle,' though less obedient than in virtuous persons (Ibid., NE, 1102 b26–28). He writes that "a person is said to have or not to have self-control, according to whether his reason has or has not the control (*kratein*), on the assumption that this (reason) is the man himself" (Ibid., NE, 1168 b34–35). Given his contention that 'reason more than anything else is man,' Aristotle's identification of self-control with control by one's "reason" is predictable. Mele's view is that human being is more holistic in nature, and so, "the self of self-control is not properly identified with reason, it is, rather, to be identified with the *person*" (1995, 6). Mele expresses a conception of self-controlled individuals who are characteristically guided by their better or best judgment even in the face of strong competing motivation, but his view does not imply that emotion, passion, and the like have no place in the self of self-control. Mele, therefore, in contrast to Aristotle, claims that self-control can be exercised in support of better judgment based in part on a person's appetites or emotional commitments.

According to Mele, self-controlled individuals need not be stoic or unemotional (Ibid.). Their feelings, emotions, and appetites can influence their conceptions of the good life or their systems of values. The better judgments that self-control serves in a particular self-controlled person may rest on a principle of practical reasoning that measures the importance of others on the basis of his emotional bonds with them. Like an Epicurean, the self-controlled person makes judgments that depend significantly on their desires, values, loves, pleasures; therefore, a better judgment (a more autonomous judgement) derives from an agent's own desires or passions, which motivate them. In this context, I would claim that a self-controlled person should be a mixture of both reason and passion—'*without reason, passion is senseless, and without passion, reason is worthless*.'

Another of Mele's books, *Irrationality* (1987), provides a partial understanding of 'how self-control can function at a certain crucial location in the generation of continent action,' and it is based on the following two theses (Ibid., 7):

1. The motivational force of our desires is not always in line with our evaluation or assessment of the "objects" of our desires.
2. Decisive better judgments are formed on the basis of our evaluation or assessment of the objects of our desires.

On the one hand, thesis 1 has considerable empirical support confirmed by common experience and by various thought experiments. The strength of desire is influenced not only by our evaluation or assessment of objects of desire, but also by such things as the perceived proximity and prospects for desire satisfaction, the perceived or imagined importance of desired objects, and the manner in which we attend to desired objects. On the other hand, thesis 2 is a standard conception of practical reasoning. When we reason about what to do, we inquire not as to what we are most motivated to do, but rather what it would be "*best*" to do. Mele points out that "an agent can… refuse, at the time of action, to focus his attention on the attractive aspects of the envisioned *akratic* action and concentrate instead on what is to be accomplished by acting as he judges best. He can attempt to augment his motivation for performing the action judged best by promising himself a reward for doing so" (1987, 23). Thus, Mele asserts that self-control can be exercised either through *present motivation* or through *anticipated motivation*.

Rorty (1980) makes the distinction that self-control may be either *regional* or *global*, and exists in degrees. An agent may exhibit remarkable self-control in one area of his life (e.g., eating), and be weak-willed in another (e.g., watching movies). Agents possessed of global self-control (self-control in all areas of their lives) would be particularly remarkable: in every area, their self-control would significantly exceed that of most people. Apparently, some self-controlled agents are more self-controlled than others.

Mele claims that being a self-controlled person is not enough for being an autonomous person. Autonomous agents possess and exercise some degree of control over their lives. In *Springs of Action: Understanding Intentional*

Behavior, Mele develops a causal view of the explanation of intentional actions and argue that the most detailed completing anticausal theories fail. He claims that those who hold that autonomous decisions are uncaused (no causal/explanatory grounds) should offer an account of deciding for reasons that they do not treat the reasons for which we decide the causal production of the decision, that is, decision cannot play its explanatory role vis-à-vis intentional action unless it incorporates preponderant motivation, or one might advert to alleged meaning conventions about 'decision.' The term 'decision' has multiple referents. It refers (1) to act of deciding; (2) to the immediate issue of the act, a decision *state*, a state of being decided upon something; and (3) to *what* we decide, as in 'his/her decision was to A' (1992, 158). He suggests that it is difficult to consider how reasons can help to account for an agent's decision to do A if those reasons do not play a causal role in the output of the decision. As we know that control is a causal phenomenon and that autonomous agents are viewed as being (at least to some degree) in control of what they decide to do, control is thus the capacity to intend to act as one judges it best to act, or the ability to intend on the basis of one's relevant evaluative judgments. Consider that an agent judges that his doing A would be morally better than his doing B, and yet, in the absence of coercion, he intentionally does B rather than A. In doing B, he need not be acting *akratically*, as he may judge that all things considered, his doing B would be better than his doing A. Even in doing B, he may be acting *continently*. Therefore, Mele believes that *akratic* action is that which violates a decisive better judgment (1987, 5–6)—an evaluation of what is best to do according to the subject's self-defined grounds. For Mele, a judgment made by an agent is a decisive better judgment if and only if, in the agent's mind (e.g., an agent's own desires, beliefs), it is the better or best action given his circumstances, and not just in some respect or other (e.g., financially), but unconditionally.

Mele explains that the kind of *akratic* action (uncompelled intentional action) which is contrary to the agent's decisive better judgment is a *strict akratic* action (Ibid., 7). Strict akratic actions seem to be an unfortunate fact of life, a matter complicated by our having at least two perspectives on the justification for intentional actions: motivational and intellectual (1995, 16). Central to the *motivational perspective* is the idea that what agents do when they act intentionally depends on what they are most strongly motivated to do. Unlike motivational and intentional action, the intellectual perspective applies only to intellectual beings. Central to the *intellectual perspective* is the belief that better judgments play a substantial role in explaining the intentional actions of rational or intellectual beings. The *motivational perspective* is suited to *akratic* action, because when acting *akratically*, one most likely does what one is most strongly motivated to do at the moment. However, this threatens the intellectual perspective. We would prefer to say that the motivational and intellectual perspectives on intentional human action includes choice, decision, evaluative judgment, motivational strength and intention at a particular time and situation. It depends on a subjective perspective of an individual.

In "How is Weakness of the Will Possible," Donald Davidson (1969) sets out three principles concerning the problem of the possibility of *akratic* action:

P1: If an agent wants to do x more than he wants to do y, and he believes himself free to do either x or y, then he will intentionally do x if he does either x or y intentionally.
P2: If an agent judges that it would be better to do x than to do y, then he wants to do x more than he wants to do y.
P3: There are incontinent actions.

P1 and P2 involve a close connection between better judgment and intentional action, but a connection that P3 seems to contravene. Davidson's aim is to demonstrate that the principles form a coherent relation. In this context, Pettit and Smith (1993, 53–54) address how the motivational and intellectual perspectives regarding intentional action can be squared with each other. The motivational perspective links motivation to intentional action, i.e., P1, and the intellectual perspective links practical reasoning and evaluative judgments to intentional action , i.e., P2, whereas Davidson tries to square P1 and P2 with the occurrence of incontinent action, i.e., P3. For Davidson, his principles (P1 and P2) jointly imply the following single argument:

P: If an agent judges that it would be better to do x than to do y, and he believes himself free to do either x or y, then he will intentionally do x if he does either x or y intentionally.

This argument (P) might seem to preclude the possibility of *akratic* action. However, Davidson defines *akratic* action as that which clashes with all things considered to be better judgment—the *akratic* agent intentionally does B while judging it better to do A. But he does not (at the time of choosing) judge it better to do A. Hence, his action does not falsify P. But, for Mele, the agent may hold the judgment that it is best to do A while no longer intending to do A. It is his attitude towards A-ing: it is no longer one of intending (1995, 23). According to Mele, best judgment is "*the best judgment with a corresponding intention*" (Ibid., 19).

Intentions are one thing and evaluative judgments another. Mele states that the conduct at issue is intentional conduct, and intentional conduct depends on its occurrence in the presence of intentions. He has argued that an action-guiding function to evaluative judgment does not devote one to suppose that the judgments are themselves *logically* or *causally* sufficient for the presence of corresponding intentions. In this context, Mele adds that one can say that our best judgments are capable of *influencing* intentions formation even though they do not uniformly give rise to intentions. Focusing on the judgements, it may seem that we have just two things to work with - the *contents* of the judgments and the judging *attitude* toward those contents (Ibid., 25).

Mele believes that all best judgments have a motivational dimension. Akratic action against such judgments would be irrational (Ibid., 26). There are many normal agents who may have a *generic*, standard desire to act as they judge best. They may desire generically to do whatever they judge best. Such a desire would dispose them to intentions in accordance with their best judgment. Mele states

that a best judgment is a "*best belief*" (Ibid., 28). Sometimes, best judgments would be distinguishable from best beliefs.

From a general understanding, self-control can be exercised against both *anticipated temptation* and *present temptation* (Ibid., 32). Every case of intentional action is a case of motivated action. Motivation is required specifically for intelligent or learned behavior, not for behavior in general (e.g., the mating behavior of flies). Motivation is required for the capacity to learn strategic behavior by way of reward and punishment. In this context, memory is required, but memory is distinguishable from motivation. Memory enables a being to form and retain an associative link between a particular type of behavior and its consequent reward, whereas motivation disposes a being to seek out certain kinds of consequences and to seek to avoid other kinds (Ibid., 35–36).

Decisive best judgments are subject to defeat by opposing motivation and are supportable by exercises of self-control. Consider a case where an agent judges it best to do A, but wants to do B, where A and B are compatible acts. Thus he may reasonably be seen to act in accordance with the stronger desire in exercising self-control to resist B. For example, a man with limited funds may want to see an expensive play in a theatre more than he wants to see a movie, but would also prefer to see a movie and eat dinner at a restaurant more than he would like to see the play. He may realize that given his financial situation, he cannot see the play and eat dinner at a restaurant. Unless a desire is irresistible, it is subject to the agent's control (Ibid., 42–44). For instance,

> Ian turned on the television about half an hour ago when he started eating lunch. He decided to have a quick meal so that he would have time to finish painting his shed before his wife came home from work. Ian has just finished eating and he is thinking that he ought to get back to work now. However, he is enjoying the golf tournament on TV and he remains seated. He tells himself that he will watch until the next commercial break, but the commercial comes and goes and Ian is still in front of the set. Thinking that he had better drag himself away from the television now, Ian utters this self-command: "Get off your butt, Ian, and paint that shed!" Ian turns off the set, picks up his painter's cap, and walks into his backyard.

Strict *akratic* actions are motivated by resistible desires, and if self-control cannot be exercised against preponderant proximal temptations, such temptations are irresistible (Ibid., 55). Our capacity for control over our desires extends even to proximal desires, including many proximal desires to A that compete with our decisive better judgments and are stronger than our proximal desires not to A. Agents who lack a capacity for control over a desire that competes with their decisive better judgment are cut off from autonomy in that connection.

Mele's view of *akrasia* is based on two principles, in which he argues: (1) one can manifest *akrasia* by performing an act that coincides with one's best judgment, and (2) one can exercise self-control on behalf of an act that goes against one's best judgment (Ibid., 60). He cites an instance:

> Young Bruce has decided to join some wayward cub scouts in breaking into a neighbor's house, even though he decisively judges it best not to do so. Suppose that at the last minute Bruce refuses to enter the house and leaves the scene of the crime. His doing so because his decisive judgment has prevailed is one thing; his refusing to break in owing simply to a failure of nerve is another. In the latter event, Bruce has arguably exhibited weakness of will; he 'chickened out'.

In the above case, Bruce experiences some trepidation about breaking into the house and tries to steel himself for the deed. Although he has exhibited strength of will, some exercises of self-control are not performed in the service of a best judgment.

We can exercise self-control in support of decisive better judgments that derive from our appetites, emotional commitments, and continent and incontinent behavior, which might be viewed as outcomes of 'intellect and passion'. Self-control is the ability to master motivation that opposes one's intentions or decisions: an agent can display self-control in abandoning an intention. An agent may exercise self-control in support of his making a decision or intention that fits his better judgment (Ibid., 61–62). To say that 'S' performed an *akratic* action is not to say that 'S' is an *akratic* person. Just as a generally honest person can lie without properly being branded a liar, someone can act *akratically* without properly being termed an *akratic* person. Hence, the above example of Bruce's participation in a break-in can be an *akratic* deed without inferring that he is an *akratic* boy. After all, Bruce may exhibit considerable strength of will in mastering the fear that opposes his decision. Better judgments may be based on and supported by passions, emotions, appetites, etc. But in some instances, *akrasia* may issue in the long term or even permanent modifications of one's conception of the good life. *Akratic* reasoning can also influence one's better judgments more directly, without altering one's conception of the good life.

There are mainly two conditions under which the displayed self-control of a person is true: (1) an agent's displaying self-control does not depend on his being a self-controlled person. Even a minimally self-controlled person may display the powers of self-control that he has, and (2) exercises of self-control oppose something in support of something else, e.g., an agent exercises self-control in opposing motivation that threatens to generate conduct, that is contrary to his better judgment (Ibid., 64).

An agent is *akratic* if and only if his performance involves the defeat of a higher-order desire by a lower-order one. The full range of continent and incontinent action can be captured only by viewing the possession of second-order desires of a certain kind as essential to each. In "Freedom of the Will and the Concept of a Person," Harry Frankfurt has given more importance to the notion of a second-order desire. According to his account, a person "has a desire for the second order either when he wants simply to have a certain desire or when he wants a certain desire to be his will" (Frankfurt 1971, 10). Wanting a certain desire to be one's will is, wanting that desire "to be the desire that moves effectively to act". Such a second-order desire Frankfurt terms a "second-order volition". For him, both second-order desires and volitions are indisputable. Sometimes, we can find that if continent action were to require intentionally resisting desires, continent action would seem to be essentially tied to higher-order desires.

Now we can see how belief plays a crucial role in the following context. Biased beliefs are divided into two general categories, *unmotivated* and *motivated*. Psychologists have identified a number of sources of unmotivated biased belief, such as vividness of information, heuristic availability, confirmation bias, and tendency to search for causal explanations. These sources are capable of functioning independently of motivation. For example, motivation can enhance the vividness or salience of certain data. Motivation can bias our beliefs; there might be a place for doxastic self-control in our lives. It depends on whether motivation can bias our beliefs relative to our own doxastic standards or principles, or whether we are capable of exercising self-control in such a way as to moderate the effects of motivation on the belief (1995, 86–88).

Akratic action manifests weakness of will, or at least an associated imperfection. Similarly, continent action manifests self-control. *Akrasia* is a deficient capacity to contain or restrain one's desires. Continent belief requires self-control. The agent who performs a strict *akratic* action acts against a consciously held better judgment. What sorts of better judgment might be violated in cases of *akratic* belief or served by an exercise of self-control instances of continent belief? There are two possibilities: J1 is the judgment that it is best not to allow what one wants to be the case to determine what one believes (P) is the case, e.g., non-strict *akratic* belief—belief formation, belief retention, and belief revision. J2 is the judgment that it is best to believe that (P), e.g., strict *akratic* belief.

There are two central ranges of cases for characterizations of non-strict *akratic* and continent believing (Ibid., 94). First, a central range of cases of *non-strict akratic believing* is constituted by instances of motivated believing that (1) violate a doxastic principle that the believer 'S' either explicitly believes to be correct or undeniable implicitly accepts, which principle is not in competition with any other principle of 'S' and (2) were suitably avoidable by 'S' by means of an employment of self-control. Second, a central range of cases of *non-strict continent believing* is constituted by instances of believing that (1) accord with a doxastic principle that the believer 'S', either explicitly believes to be correct or undeniably implicitly accepts, the principles that are not in competition with any other principles of S, and that (2) were suitably promoted by S's exercising self-control.

In *Irrationality*, Mele talks about self-deception, and discusses some common ways in which desire can lead to biased belief in cases of self-deception, such as negative misinterpretation, positive misinterpretation, selective evidence gathering, selective focusing, or attending. He also claims that 'self-deception commonly is not intentional' (Ibid., 95–96). Just as traditionally conceived *akratic* actions are distinguished from compelled actions performed against the agent's better judgment, *akratic* feelings conceived are also uncompelled. Like *akratic* beliefs, *akratic* feelings are not *akratic* actions; a proper conception of *akratic* feeling must accommodate the difference. The preceding remarks suggest the following characterization of *strict akratic feeling* (Ibid., 102–105):

> S's being or remaining in a feeling-state, x during t is an instance of strict *akratic* feeling if and only if S's being or remaining in x during t is uncompelled, and during t, S consciously holds a judgment to the effect that there is

good and sufficient reason for his not being or remaining in x, again feeling-states include states of not having certain feelings.

Strict *akratic* feeling necessarily derives from the strict *akratic* action. Specifically, a strict *akratic* action associated with one's not bringing one's feelings in line with one's better judgment. Strict *akratic* actions violate *practical* better judgments, judgments about what it would be best to do or refrain from doing.

Values (intrinsic and extrinsic) are trickier (Ibid., 118). An agent *intrinsically* values X if and only if he values X for its own sake, or as an end. Some things are valued both as ends and as means. For example, Pujarini values daily exercise both for its own sake and because it is conducive to good health. We can distinguish between intrinsic and extrinsic values. X is among person S's *intrinsic values* if and only if X is among S's values and S intrinsically values X—for example, justice is one of Pujarini's intrinsic values if and only if justice is among her values and she intrinsically values justice. Similarly, X is among S's *extrinsic values* if and only if X is among S's values and S extrinsically values X: physical fitness is a value of Pujarini's and she values it as a means to happiness. In the case of valuing, the same personal value can be both intrinsic and extrinsic: Pujarini may value physical fitness both as a means to happiness and for its own sake.

Mele states that the ideally self-controlled person will not assign value *akratically*. People can revise or embrace their values on the basis of principled reflection, but the principles that inform reflections can themselves be shaped by their values. The ideally self-controlled person *perfectly* manifests what Mele calls *perfect self-control*, which has four dimensions (Ibid., 121–122):

1. *Range* Perfect self-control is perfectly global. It is manifested in overt actions, mental actions, intentions, beliefs, and emotions, in practical reasoning and decisive better judgments, and in the assessment, acceptance, and revision of values and principles. It has a *maximal categorical range*.
2. *Object* The ideally self-controlled person has never exercised self-control errantly, but only in support of "pure" items that non-*akratically* hold decisive better judgments, values, principles, and the like. The objects of exercises of self-control in the ideally self-controlled person are *perfect*.
3. *Frequency* The ideally self-controlled person exercises self-control whenever he reflectively deems it appropriate to do so.
4. *Effectiveness* The ideally self-controlled person's exercise of self-control always succeeds in supporting what they are aimed at supporting. They are perfectly effective. This is not a matter of luck or causal deviance; rather it is the ideally self-controlled person consistently and intentionally bringing about the success of his exercises of self-control.

Part II

Personal autonomy is a property of a person, e.g., values, principles, beliefs, desires, intentions, and plans. For Mele, 'psychological autonomy' is an autonomy with

regard to various aspects of one's mental or psychological life, including one's *pro-attitudes*, e.g., one's values and desires. There are three distinguishable species of autonomy that agents may enjoy with respect to their pro-attitudes: (1) developing—an agent autonomously *developing* a pro-attitude over a stretch of time, (2) possessing—an agent autonomously *possessing* a pro-attitude during a stretch of time, and (3) influencing—an agent being autonomous with respect to the *influence* of a pro-attitude on his intentional behavior (Ibid., 138).

There are at least two capacities involved in personal autonomy, namely (1) outer-directedness, i.e., one's environment, and (2) inner-directedness, e.g., capacity for decision-making, critical reflection on one's values, principles, preferences, and beliefs. Both types of capacity have at least a partly psychological basis (Ibid., 144). Psychological autonomy has two types of conception. One is the *externalist* perspective: there is more to being psychologically autonomous over a stretch of time than what goes on inside a person during that time. The psychological autonomy of some individuals also depends on how they came to possess the values and desires that guide self-reflection and decision-making, and this also depends on agents' *causal histories*. The other is the *internalist* perspective: the history is relevant only insofar as it yields functional capacities for such things as decision-making, self-reflection, identification, and self-modification (Ibid., 146).

Mele views *P-autonomy* (for 'possession') as autonomy regarding the *possession* of a pro-attitude that one has an integral psychological autonomy (Ibid., 149). There are a number of points about ability on which compatibilists and incompatibilists can agree, such as (1) the ability to do A intentionally has a *representational* dimension, e.g., John currently has no representation of a means of impressing his new neighbors, and it has not even occurred him to try to impress them; (2) the ability to do A intentionally has a *motivational* dimension, e.g., John now has no motivation to leave his office soon, because he is in the midst of a thoroughly enjoyable project. But if there is no bar to his acquiring such motivation, he is intentionally able to leave his office soon, (3) the ability to do A intentionally has a distinct *executive* dimension, e.g., owing to illness, (4) the ability (during *t*) to A intentionally (during *t*) often is contingent upon *environmental* conditions, e.g., John who knows how to play tennis (during *t*). He may be stranded in a desert, miles from the nearest tennis equipment and so, he has the relevant motivation and executive properties may be unable to play tennis (Ibid., 149–150).

Mele argues that autonomous possession of a pro-attitude requires authenticity regarding that pro-attitude. Similarly, Joel Feinberg, in his book *Harm to Self* (1986, 33), writes:

> A person is authentic to the extent that … he can and does subject his opinions and tastes to rational scrutiny. He is authentic to the extent that he can and does alter his convictions for reasons of his own and does this without guilt or anxiety…. He will select his lifestyle to match his temperament, and his political attitudes to fit his ideals and interests.

For Feinberg, autonomy does not entail being a wholly self-made person. The ideally self-controlled agent is not infected by *akrasia*. In chapter 5 (Self-control and Belief), Mele characterized an ideally self-controlled person is that which he wants

to be, and his influences are what he believes to be true, albeit only in the ways permitted by his principles (one's subjective beliefs on one's own principles). Since deliberation is partly a matter of belief formation, including the formation of evaluative beliefs about prospective courses of action, and the observation applies to the deliberation of ideally self-controlled individuals, such individuals will not *akratically* violate their deliberative principles, which are the principles concerning how they should conduct their deliberation. Ideally self-controlled deliberators need not be *perfect* deliberators (1995, 178).

The non-autonomy of an agent with respect to a deliberative process is explicable in two ways: (a) facts about the input to deliberation, and (b) facts about what the agent does and is capable of doing with that input. Even an agent possessed of ideal self-control can have some compelled pro-attitudes. Such pro-attitudes can function as input to deliberation, rendering the agent non-autonomous with respect to a particular deliberative episode. An ideally self-controlled person can be cut off from autonomous deliberation, and the autonomous deliberation is handled by the attribution of authenticity to an agent.

We have a psychological autonomous agent if the following three conditions are met (Ibid., 187): (1) the agent has no compelled motivational states or some coercively produced motivational states, (2) the agent's beliefs are conducive to informed deliberation about all matters that concern him, and (3) the agent is a reliable deliberator. These are not the necessary conditions for psychological autonomy, but they are sufficient for such autonomy. Mele has not offered a compatibilist analysis of autonomy, but he has advanced a compatibilist set of sufficient conditions for psychological autonomy, and it is supplemented to provide sufficient conditions for autonomous action, which excludes covert non-constraining (CNC) control by another agent. A CNC controller can operate on three agential fronts (Ibid.): (1) the victim's motivational attitudes, (2) the victim's informational attitudes, and (3) the victim's executive qualities. The satisfaction of the "compatibilist trio" of conditions suffices for the absence of CNC control by another agent on each of these fronts. Robert Kane (1989) argues that no compatibilist account adequately addresses the possibility of certain sorts of manipulation. He has argued that compatibilism cannot give a good response to the possibility of 'CNC control.' Mele discusses various cases of CNC control and argues that his conditions provide a good response to Kane's challenge to compatibilism. However, Mele further points out his view in two ways: First, the CNC control problem can be handled by compatibilists. They can offer plausible compatibilist grounds for the judgment that agents manipulated by a CNC controller in representative cases do not conduct themselves autonomously. Second, the CNC control problem can be handled by them in a way that not only provides sufficient conditions for the absence of causality determining CNC control over a specified choice (or decisive better judgment, or intention) but also provides sufficient conditions for the absence of autonomy-blocking CNC control that leaves various relevant options open.

There are both advantages and disadvantages to autonomy. Compatibilist (soft-determinist) belief in autonomy can help to render secure our belief in ourselves as autonomous agents, whereas an agent cannot be autonomous if each agent-internal state or event and the other is the product of some deterministic causal chain. The

advantage of incompatibilist belief in autonomy (libertarian sense of free will) is that it can accommodate the intuition (i.e., an agent is not autonomous if his states and actions are the result of any deterministic factors) with the agent's having ultimate responsibility and control. However, the advantages of agnostic autonomism (agnostic about compatibilism belief in autonomy) are that it does not insist that autonomy is compatible with determinism, nor does it insist that we are internally indeterministic in a way that is of use to libertarians. With regard to its disadvantages, (1) no set of sufficient conditions for compatibilist autonomy is sufficient for autonomy, and (2) human beings are not internally indeterministic in a way that is theoretically useful to incompatibilist believers in autonomy. The advantage of non-autonomism is that no human being is autonomous, and its disadvantage is that it must provide a convincing argument for incompatibilism and convincing grounds for the thesis that human beings are not internally indeterministic in a way that is required for incompatibilist autonomy (Ibid., 252–253). Lastly, Mele accepts agnostic autonomism, i.e., both compatibilism and libertarianism; the agnostic autonomist may believe that if internal indeterminism is not actually obtained, then compatibilism is true.

My Views on Mele's Understanding of Autonomy and Self-control

With regard to this book, I want to stress that an autonomous agent has some sort of agential power, whereas self-control is necessary to enhance the agent's autonomy. I do argue that an ideally self-controlled person does not lack autonomy, because he is not always trying to control only the continent and incontinent behaviours, actions, and feelings. Someone's power of "control" is grounded in his own personal autonomy and without personal autonomy, that agent cannot control himself with respect to executing any kind of continent acts. For example, Pujarini is an ideally self-controlled person, and she knows how to control or hold her anger in public places; for executing such an act, she should be an autonomous person with a strong capacity or power to control her anger attitudes. I would prefer to express that in the case of an intentional action, both self-control and autonomy are merged, that is, there is an essential connection between these two notions. It seems that both autonomy and self-control are necessary conditions for exercising any kind of intentional action.

I shall claim that there is a very close relationship between 'autonomy and moral responsibility,' on the one hand, and 'self-control and moral responsibility,' on the other hand. In general, moral responsibility requires an autonomous agent and a self-controlled person, but it does not encompass that all events for which one is morally responsible at a time during which should be an autonomous agent or a self-controlled person. One can contend that an agent is not *morally responsible* for an act—and may be only *responsible* for that act—at some relevant time, even if that agent is an autonomous agent and/or a self-controlled person. The relevant time need not be while an agent is actively pursuing an end. It may be a time during which an agent passively neglects to do something; it may be the time of 'not-doing'. For an instance, an autonomous agent may be morally responsible for not making an effort to save a child he sees

drowning and for the child's death by drowning. Similarly, a lifeguard who carelessly falls asleep at his post may be morally responsible for the drowning of a swimmer in his charge. If the lifeguard were not an autonomous agent or not a self-controlled person, then he would not be morally responsible (rather we can say only responsible) for the drowning. I think, one's notion of moral responsibility will encompass not only one's notion of free will or continent feelings and behavior, but also some notion of *moral significance* where one can determine some notion of *what morality is*. Thus there is an intimate relationship between autonomy, self-control, and moral responsibility.

What I have noted in this book is that Mele is trying to build a bridge between reason and passion in the case of decision-making or better judgement. Consider the example of emotion. Emotions play an important motivational role in our lives and we have the power/ability to control our emotions or other affective feelings, that is, how we respond or react on them. In many cases, our emotions are the product of our desires, habits of interpretation, learned patterns of emotional response, other affectives, and physiological conditions. Similarly, a self-controlled individual who is characteristically guided by his better judgments, even in the face of strong competing motivation, does not devoid of emotion or passion. As I have already discussed in part I regarding Mele's view that the 'human being is more holistic in nature, and the self of self-control is not exactly identified with reason; it is preferable to be identified with the person or *agent* (my word)' (Ibid., 6). Therefore, we claim that self-control can be exercised in support of better judgment, based in part on a person's appetites and emotional commitments, as well.

I have noted that Mele's better judgment is a general human tendency, i.e., the human attitude, especially one that favors one alternative over others. As we are existential, emotional, social, rational, and selfish human beings, we need beneficial or profitable options over the alternatives, and for that we are always giving more priority to the better judgment which is decided by our reason, desire, belief, emotion, or whatever we feel pleasurable, according to the different perspectives. We all want to fulfill our own desires throughout the course of our lives—a seemingly infinite series of our good or bad desires.

Finally, I would prefer to articulate that 'being a self-controlled person is a necessary capacity or ability of being an autonomous person', instead of Mele's view that 'being a self-controlled person is insufficient for being an autonomous person'. Consider an example that Pujarini is a self-controlled person who can control everything in the sphere of her life, but she has no autonomy to act. In this context, do you think that she is really a self-controlled person who has no autonomy? I would say "NO" because on the one hand, controlling one's desire over others is a power of a self-controlled person, and on the other hand, that power of "trying" itself has some sort of autonomy, and we cannot ignore this kind of attempting of the mental act. I cannot, hence, support Mele's definition of an autonomous person is a self-controlled person and not vice-versa. If we agree on Mele's view in this context, I think that it will be seen as a paradox in itself.

Conclusion

The concept of "autonomy" suggests any uncompelled, uncoerced, intentional action of a self-conscious and self-reflective agent, and it may be held that all such actions are free unless there are some freedom-blocking properties, such brainwashing, compulsion, coercion, insanity, or relevant deception. An autonomous agent and self-controlled person are involved in self-reflection, self-assessment, careful deliberation, and successful attempts at self-modification. Free agents are first and fundamentally free with respect to what they do.

Self-control is not only exhibited in behavior that is in accordance with the agent's better judgment in our thinking, but is also that which can be or ought to be exercised because it is exhibited in both *motivational* and *evaluative* ways. It can be exercised in order to balance one's motivations with one's better judgment, and is also encouraged by one's own rational evaluation based on both *intrinsic* and *extrinsic* desires. I would say that the autonomous agent and self-controlled person have the power to make choices which can only and ultimately be explained in terms of their own self-will, i.e., character-traits, motives, and efforts.

Last, but not least, this book, *Autonomous Agent: From Self-Control to Autonomy*, is meticulously written and well-argued from both a philosophical and psychological perspective. It would be more effective for those scholars who are interested in the philosophy of action, especially regarding free will, agency, self-control, moral responsibility, and *akrasia*. I highly appreciate Alfred Mele's philosophical argumentation, vast literary contributions, and academic intellectual depth.

Acknowledgments With a great sense of joy and pleasure, I take this opportunity to thank my research supervisor, Dr. Vineet Sahu (IIT Kanpur, India), for his comments and suggestions of an earlier version of this paper. My sincere thanks also go to the anonymous referees of the JICPR for their valuable corrections, remarks, and suggestions.

References

Ainslie, G. (1975). Specious reward: A behavioral theory of impulsiveness and impulse control. *Psychological Bulletin, 82,* 463–496.

Davidson, D. (1969). How is weakness of the will possible? In J. Feinberg (Ed.), *Essay on actions and events*. 2nd Edn. Oxford: Clarendon Press.

Frankfurt, H. (1971). Freedom of the will and the concept of a person. *Journal of Philosophy, 68,* 5–20.

Joel, F. (1986). *Harm to self.* New York: Oxford University Press.

Kane, R. (1985). *Free will and values*. Albany: The State University of New York Press.

Kane, R. (1989). Two Kinds of incompatibilism. *Philosophy and Phenomenological Research, 50,* 219–254.

Mele, A. R. (1987). *Irrationality: An essay on akrasia, self-deception, and self-control*. New York: Oxford University Press.

Mele, A. R. (1992). *Springs of action*. New York: Oxford University Press.

Mele, A. R., & Agents, Autonomous. (1995). *From self-control to autonomy*. New York: Oxford University Press.

Mischel, W. (1973). Effects of attention to symbolically presented rewards on self-control. *Journal of Personality and Social Psychology, 28,* 172–179.

Pettit, P., & Smith, M. (1993). Practical unreason. *Mind, 102,* 53–79.

Rorty, A. (1980). Where does the akratic break take place? *Australasian Journal of Philosophy, 58,* 333–346.

Ross, WD. (1915). Translator. *The Nicomachean Ethics*. By Aristotle, New York: Oxford University Press.