7-26-2010

Self-Governing Policies: A Critique of Bratman

Robert J. Muckle
University of Massachusetts at Amherst

Follow this and additional works at: http://commons.pacificu.edu/rescogitans

Part of the Philosophy Commons

Self-Governing Policies: A Critique of Bratman

Robert J. Muckle
University of Massachusetts at Amherst

Michael Bratman in his essay "Reflection, Planning and Temporally Extended Agency"[1] espouses a reductive view of agency based largely upon what he calls "self-governing policies." Combining self-governing policies with a Lockean approach to personal identity over time, Bratman's view stays within the bounds of event causation and attempts to reduce agency to mere attitudes. First I will give the problem Bratman's approach to agency intends to solve, then move through his view beginning with the planning theory of intention, then on to his Lockean approach to personal identity and finally self-governing policies. I will then pose some problems for his theory.

Bratman is attempting to create a reductive theory of agency, and in this paper is concerned with agential authority. With his theory he tries to show how persons qua agents can meaningfully decide to do some action, thus giving them authority over their actions. He wants to explain how an agent can "take a stand as an agent to see where [she] stand[s] with respect to a first-order desire."[2] Bratman uses his idea of self-governing policies to account for agential authority without specifying exactly one thing in the brain to call "the agent."

The theory Bratman presents is reductive in the sense that he wants to take sentences of the form "Jenny chose to drink soda" and explain them away without reference to one thing to call "the agent." Self-governing policies are the explanation that removes the reference to one thing that is "the agent."

Bratman rejects the agent-causal view of action rather quickly.[3] One issue he takes with the agent-causal view is that it is difficult to say what the agent *is* as separate from the relevant agent-functioning of the brain. His second objection to the agent-causal view is that it leaves behind the event-causal order. Human agents are assumed to be largely different agents from cats and dogs, but cats and dogs seem to be stuck in the event causal order. Bratman does not want to account for this difference by allowing for human agents to have agent-causal powers that affect the event-causal order. He thinks his theory accounts for that very difference, without reference to agent causation.

With the motivation for Bratman's paper in mind, it is time to dig in. Bratman's planning theory of intention is rather simple: when an agent plans to do some future action, she is thereby committed to the intention to do that action. Plans are not irrevocable, but they do have stability: "there is, normally, rational pressure not to reconsider and/or abandon a prior plan."[4] Plans are in many cases specific in a means-end manner, which means that plans will organize the actions needed to happen to achieve a certain desired end.

Some plans, indeed the ones that guide much of our lives, are general plans, which Bratman labels "policies."[5] One example of such a policy is the widespread ritual of morning coffee. A policy of coffee in the morning, perhaps on the way to work, is not unbreakable as sometimes one runs late and cannot make a pot, or stop at the desired coffee spot. However, the wide majority of the time the agent with a policy of coffee in the morning gets her coffee. So, policies are like plans in that they are breakable sometimes, although policies are more general in the sense that policies guide actions in recurring situations.

A discussion of future plans leads Bratman to discuss personal identity, since a view of agency that is reliant on future actions had better have the actions carried out by the same agent who planned them. As mentioned, Bratman has a contemporary Lockean view of identity over time.[6] He believes that not only are backward looking memories part of identity over time, but forward looking motives and plans are part of identity. Naturally, he thinks that the forward-looking plans and their later execution are (loosely) the grounds for personal identity. Bratman does not take the time to spell out every part to his personal conception of the contemporary Lockean approach, but assumes one is available and moves on.

The important piece that he does spell out is his view of a Lockean connection. He thinks that a Lockean connection is one between an intention and its later execution: connections require temporal cross-reference. Temporal cross-reference, for Bratman, is simply an agent planning to do something at some time t1, then at a later t2 when that plan is executed, the agent has an understanding of the completed plan she made at t1. In this sense, the agent has reference from one time to another. This is an interesting point for him to make since it seems that policies such as morning coffee might not be subject to such temporal cross-reference. Surely there are people who each night set themselves up for a cup of coffee in the morning, however what of the people who do not? Suppose someone has been stopping at the same Dunkin' Donuts each day at roughly the same time to get the same coffee. Over time she will naturally stop thinking about getting her coffee in the morning, she will just do it out of habit. This person seems to have a policy in place, which should be part of a person's identity over time, but yet does not have the Lockean connection that Bratman seeks, since she does not actively plan to get coffee, she just does it.[7]

Accepting Bratman's view of identity, we trudge on into self-governing policies. Self-governing policies are higher-order policies that are concerned with how to act generally. Distinguish this from merely having a policy of drinking coffee each morning: a self-governing policy is one that would be more to the effect of "weaning myself off of caffeine." This policy is higher-order because I have a desire about a desire: I want caffeine, but I also want to want less caffeine. So, my self-governing policy will endorse actions that lead to fending off my desire for more caffeine by way of "half-caff" or some such.

Bratman asserts that an agent cannot have two contradicting self-governing policies.[8] Two self-governing policies P and Q contradict one another when a justifying act for the desired end P goes against the policy Q. For example, having a self-governing policy to not indulge in desserts after a meal will be contradicted by a self-governing policy of always having ice cream after a hot meal to cool ones throat.[9] Bratman uses a technical term to express when an agent's policy is without contradiction amongst her other policies: "satisfied." An agent is "satisfied" with a policy P when there is no policy Q that an agent already has, such that Q contradicts P.

One might think that there are times when an agent does have contradicting policies due to contradictory beliefs about the world. Suppose an agent believes that Barack Obama will be a great president, but also believes that anyone with "Hussein" as part of their name is a terrible president, then it seems she is contradicting her own beliefs. Suppose further that she (like most, hopefully) has a self-governing policy to always vote for the candidate she thinks is the best one, but also that (due to her hatred of Saddam Hussein) that she will never vote for someone with "Hussein" in their name. These policies are contradictory since the action of voting for Obama, whom she believes is the best candidate, will go against her policy of not voting for someone with "Hussein" in their name. This contradiction is clearly due to the agent's being in a bad epistemic situation: she has not made the mental connection that "Hussein" is a part of her favorite candidates name, so she does not know that voting for Barack will contradict her other policy.

However, adjust the example a bit and it is problematic: suppose that she is very deeply interested in the American political landscape, she knows all of the likely candidates for the presidency for several elections to come. She knows that Barack is (just shy of beyond a doubt) going to be the only candidate with "Hussein" in his name. Bratman does assert that self-governing policies are not immune to being broken, sometimes one just gives in to a desire that contradicts the policy, but she does not adjust the policy according to that whimsical desire. So having adjusted the voting case as such, it seems that our fair agent is in a position to contradict one of her policies via voting for Barak, but it is allowed since it is clear to her that it will be just this once. So perhaps it is better to tighten up the matter of "satisfaction" and let it be thus: an agent is satisfied with a self-governing policy P when there is no policy Q such that *any action* policy P

will endorse will be rejected by Q. This would allow for voting-type cases, but reject the cases where an agent has two directly contradictory policies (e.g. the dessert/ice cream example from before).

So having explained what a self-governing policy is, now it is time to give its role.[10] A self-governing policy is what speaks on the agent's behalf when the agent takes a stand with respect to a certain desire. Like many, Bratman refuses to pick out the agent for fear of agent causation, so it is the job of the self-governing policy to endorse one desire over another. A self-governing policy endorses a certain desire "by the endorsement… of the relevant functioning of that desire by a self-governing policy with which the agent is satisfied."[11] When Bratman uses "relevant functioning of that desire" he means the actions associated with that desire. Recall my self-governing policy from earlier: I want to wean myself from caffeine. The relevant functioning of that policy will be actions like skipping afternoon coffee, or water in place of caffeinated soda. The self-governing policy will endorse those actions, since they will lead to the desired end of being weaned from caffeine. The policy will reject actions that go against the desired end, such as getting a cup of coffee in the afternoon, or a caffeinated beverage with dinner. The policy is neutral on actions that neither lead to the desired end, nor against the desired end.

Here, "endorse" and "reject" seem to be metaphorical ways to express the role that self-governing policies play. To say that a self-governing policy "endorses" one action over another is to say that the action in question is considered to be a means to the desired end of the self-governing policy. So, a self-governing policy to wean oneself from caffeine "endorses" actions by the actions' being means to the end of less dependence on caffeine. "Reject" works in a similar manner: a self-governing policy rejects actions when the action inhibits the ability to get to the desired end.[12]

Self-governing policies are also not immune from rational revision: I might over time decide that I do like my afternoon coffee enough to change my policy to endorse that action, or actions similar to it.

It is here where I object: on Bratman's view, when an agent is going to "take a stand" it is the self-governing policy that speaks on her behalf. However, self-governing policies are subject to rational reflection. He cannot account for who (or what) is doing the reflecting in that sense. In a more concrete sense, Bratman cannot account for what endorses a self-governing policy. Merely calling them "self-governing" does not exempt them from a need for an endorsement of their own. By this I mean that self-governing policies cannot really *be* self-governing in precisely the sense that as an action needs endorsing, so too do self-governing policies.

There needs to be some point at which the higher-order self-governing policies can come to a stop, or it is a regress. Bratman supposes that self-governing policies are

subject to revision, which means that the agent will decide between two self-governing policies which one she shall accept. So, she is going to endorse one over the other. The problem is that Bratman cannot say it so simply since he rejects agent causation, thereby rejecting the idea that there is a part of our metaphysics, this thing, the agent, that can endorse one over the other. He will need another higher order self-governing policy about which policies can be accepted or rejected. That policy that decides between lower ones will have to have been endorsed by yet another one. This higher policy, too, will have to be endorsed, and thus each time on the way up the policy will need endorsement.[13]

Bratman supplies the aforementioned "satisfaction" in order to stop a possible regress. If an agent is satisfied with self-governing policy P that means that there is no policy Q that contradicts P, so she has only P endorsing actions. However, in the process of rational reflection satisfaction becomes irrelevant. If I am to revise my self-governing policy to wean myself from coffee then I must bring in another self-governing policy to conflict with it, and decide between the two. Satisfaction then, is left behind in the process of reflection; reflection will bring the agent to be satisfied with the policy she chooses.

The agent does not even need to bring in a completely different policy to conflict with the one she wishes to revise: suppose that an agent has a policy of never giving table scraps to dogs (since, if she does, the dog will never stop begging). However suppose the dog her friend brought over was just so darned cute, albeit rather thin, she decides to feed him a few bites of chicken breast. Seeing the dog's joy, she then decides to adjust her old policy: only give dogs table scraps if the dog that is begging for food is really really cute. The agent has not brought in a completely new policy, it is only an adjustment to a current one, however the objection still stands: what endorses the new policy? It must be a higher-order self-governing policy, which also needs endorsing. Thus, the regress is still present.

This regress provides a serious problem for Bratman. If I am right about the regress, then when an agent attempts to adjust a self-governing policy she cannot do so since higher order policies must also be endorsed. Of course he could accept that agents decide those things, but he is openly against agent causation. If, in light of the regress, we decide to give up that self-governing policies can be reflected upon then the whole idea of an agent changing what she wants over time is lost, since she cannot change her long term wants.

Another objection can be made that stems from his view of identity over time. Suppose that Sally is a strict Republican in the most stereotypical sense. She hates any and all things that are associated with the political Left-Wing. She is also a fundamentalist Christian who hates any and all people who are not. Any policies that she has regarding these aspects of her life are subject to rational revision, as Bratman allows. Suppose one

day she is at a restaurant bar and strikes up a conversation with a union carpenter who explains the trials of being a workingman in such a society. Inspired by this she decides to read the Communist Manifesto. Upon completion of the Manifesto she drops all of her prior beliefs and is now a stereotypical Marxist, even dropping her religious beliefs citing them as "the opiate of the masses." It is fair to say that Sally has changed every single one of her policies that guided her life upon completion of the Manifesto. The question then becomes whether or not Marxist Sally is the same person as Republican Sally.

It is easy enough to say that she is indeed a different person after her life changing experience. Everything that guides her life going forwards is now completely different. The Lockean connections between plans and their executions are no longer there from the plans she had when she was still a Republican. Her plan to vote for the Republican senator in the next election is certainly out of the question now, as is her plan to go to church on Sunday. Since she is not going to execute these plans, the forward ties from when she made those plans is lost, and it seems as though she is a different person.

Bratman could respond to my story as follows: although she has lost the forward-looking connections, she has the backward looking memory of the times when she was a Republican and was an avid churchgoer. She still has thoughts of "I used to be so ignorant of the workingman" or "What was I thinking in voting for that guy!?"So, although her policies have changed, she remembers the times when they were different. However, adjust the story a bit and she has no backward looking memory. On her drive back to work Sally was involved in a nasty car accident and lost much of her memory. She remembers the pleasant conversation with the carpenter, but nothing else. The carpenter hearing of the accident got her a copy of the Manifesto to read while in the hospital, and it changed her life as before. Now Sally has no backward looking memory to tie herself into the times when she was a Republican, she only has the forward looking (and completely different) connections of her new self-governing policies to rouse rabble and protest large corporations. Now Bratman is stuck with Sally being two different persons in the same life: the one who was a republican before the accident, and the one who was a Marxist after the accident.

---

[1] "Reflection, Planning and Temporally Extended Agency" Bratman, Michael. Philosophical Review vol. 109.1 (January 2000)

[2] Bratman (2000): 38

[3] Bratman (2000): 39

[4] Bratman (2000): 40

[5] Bratman (2000): 41

[6] Bratman (2000): 43-45

[7] There is more to be said here about personal identity, but I will hold off on it until later on.

[8] Bratman (2000): 49-50

[9] Assuming ice cream after a meal is always considered a dessert. Also, I think the example needs to be such that it is not direct contradiction (e.g. P & ~P) since I think having directly contradictory policies would very likely create some kind of cognitive dissonance and call for immediate revision.

[10] There is obviously more to be said about the specificities of self-governing policies, but that will only complicate things, and is not necessary here.

[11] Bratman (2000): 50. I set aside the more nitpicky complexes with satisfaction Bratman covers.

[12] I am decidedly unsure of how Bratman wishes to precisely spell out a self-governing policy's endorsement since he does not do so in his paper. My own thoughts in his defense are thus: an action A is endorsed by self-governing policy P when A is a member of a class of actions C that are means to the end P. So my policy to wean myself from coffee would endorse actions that are a member of the class of actions that lower my caffeine intake. I think this is a fair way to see endorsement, but it is not found in his paper so I will not pursue it further.

[13] A quick, but perhaps unfavorable, response on behalf of Bratman could be that there are basic policies that we just have. Perhaps third-order policies are so general that they can handle the revision of the lower policies. These third order policies cannot come under reflection since they are brute facts about the way that a person is. This seems unfavorable since those would be better candidates for personal identity over time, but Bratman does not want something of that sort to be our identity.