

AGENCY THEORY RECONSIDERED

THOMAS J. DONAHUE
Mercyhurst College

In this article, we briefly outline and explain Roderick Chisholm's theory of agent causation, critically evaluate certain criticisms of it advanced by Harry Frankfurt, Clifford Williams and Irving Thalberg, and show that the theory can be improved by elaborating it in two ways, one of which is suggested by the work of Richard Taylor.

Chisholm frames what he labels "the metaphysical problem of human freedom" as follows:

Human beings are responsible agents; but this fact appears to conflict with a deterministic view of human action (the view that every event that is involved in an act is caused by some other event); and it also appears to conflict with an indeterministic view of human action (the view that the act, or some event that is essential to the act, is not caused at all). To solve this problem, I believe, we must make somewhat far-reaching assumptions about the self of the agent—about the man who performs the act.¹

The statement, explication and development of these assumptions is Chisholm's theory of agent causation.

Thus, Chisholm advances his theory because he believes that it can account for the fact that human beings are responsible agents whereas neither a deterministic nor a strictly indeterministic view of human action can do so.² We will not attempt to determine whether it is a fact that human beings are responsible agents, nor whether the existence of agent causation is a necessary condition of such responsible agency, but will confine ourselves to the more modest agenda outlined above.

Chisholm explicates the proposition that persons are agents who sometimes possess the freedom to act otherwise than they actually do act (according to Chisholm, the most fundamental assumption about the agent

¹Roderick Chisholm, "Freedom and Action," in Keith Lehrer, ed., *Freedom and Determinism* (New York: Random House, 1966), p. 11.

²By a "deterministic view of human action" we mean that view which states that every human action is produced by a sufficient causal condition. By a "strictly indeterministic view of human action" we mean that view which holds that free actions are wholly uncaused, i.e., neither produced by a sufficient causal condition nor freely initiated by an agent.

which is required to resolve successfully "the metaphysical problem of human freedom") in the following way. There are instances of free action in which there is neither a sufficient causal condition for "undertaking p" ("acting with the intention of bringing about p") nor a sufficient causal condition for not "undertaking p": i.e., it is neither causally necessary that the person perform the given action or causally necessary that the person not perform the given action. The agent is the causal source of his free actions.³ The agent freely performs these actions: i.e., he causally contributes to his performing them and is not caused to do so by antecedently existing conditions.⁴

We will now turn our attention to some criticisms of Chisholm's theory. In his "Freedom of the Will and the Concept of a Person,"⁵ Harry Frankfurt contends that an adequate theory of human freedom must be able *both* to account for our disinclination to ascribe such freedom to animals *and* to explain why having this freedom is desirable. He then argues that Chisholm's theory meets neither of these requirements and thus is inadequate.

Frankfurt's first criticism is off the mark. While it is true that Chisholm does not explain why it makes sense not to attribute agency to members of other animal species, clearly he could have used his approach to argue as follows. It makes sense to hold that persons are free agents because the weak inductive argument to the contrary (i.e., the argument from the fact that some human actions are wholly explainable in terms of event causation to the conclusion that all are wholly explainable in such terms) is outweighed by the conjunction of the fact that it is a fundamental presupposition of our ordinary activities that persons are morally responsible for at least some of their actions and the fact that such moral responsibility presupposes that persons are free agents.⁶ On the other hand, it makes sense not to attribute such agency to other animals, since the much stronger inductive argument for the total causal determination of their

³Terms such as "his" and "he" should be taken where appropriate as abbreviations for "his or hers" and "he or she" respectively.

⁴The key concept of "causal contribution" and the role of reasons, purposes and motivations in causally contributing to the occurrence of free actions will be explained more fully when we discuss the criticism leveled at agency theory by Williams and Thalberg.

⁵Harry Frankfurt, "Freedom of the Will and the Concept of a Person," in *The Journal of Philosophy*, Vol. LXVIII, No. 1, January 14, 1971, pp. 5-20.

⁶To the best of my knowledge, Chisholm has not given an argument for his assumption that our ordinary activities presuppose that human beings are morally responsible in a sense which requires that they agent-cause at least some of their actions. However, given the truth of this assumption, this line of argument is effective in rebutting Frankfurt's objection.

behavior is not counterbalanced in any way: i.e., we do not presuppose that animals are responsible agents nor are there any other rationally justified considerations which render plausible the proposition that animals are free agents.

Frankfurt's second criticism is also unfounded. As we have seen, he claims that Chisholm's theory does not explain why possession of the capacity to agent-cause at least some of one's actions is desirable. This criticism is doubly flawed. First of all, it isn't clear that the adequacy of Chisholm's theory rests in any measure upon its capacity to explain the desirability of possessing such causal power. After all, Chisholm is simply making the claim that we have such power. Since the proposition that persons are agents bears no logical relationship to the proposition that possession of such causal power is desirable, Chisholm's theory (an explication and elaboration of the former proposition) is not deficient merely if it does not substantiate the latter proposition.

In addition, it is clear that Chisholm's theory can explain why having such causal power is desirable. While Frankfurt is quite right in saying that there is no discernible phenomenological difference between the experience of a person who agent-causes a given action and that of a person who performs the same action without there being any breach of the event-causal nexus, surely Frankfurt is incorrect in assuming that this provides good reason to believe that there is no basis for preferring to be a Chisholmian agent as opposed to a deterministically conceived actor.⁷

To see that this is the case, we shall use Frankfurt's distinction between "first-order desires" and "second-order desires." Frankfurt defines first-order desires as "desires to do or not to do one thing or another."⁸ He defines a second-order desire as either a desire "to have a certain desire" or a desire that a certain desire be effective in moving oneself to action.⁹ It will be useful for our purposes to broaden the notion of "second-order desire" to include all the desires regarding the role which desires play in the production of one's own actions.

Given that one believes, as Chisholm and others do, that the existence of agent causation is a necessary condition of responsible agency, the preferability of being a Chisholmian agent to being either a deterministically conceived actor or a person whose only free actions are wholly uncaused can be understood in terms of the second-order desire to have one's desires (and the various other causally contributing factors) play that role in the bringing about of one's actions which allows for the possibility of one's possessing that human dignity which comes from being a responsible agent and that *justified* sense of pride at having been truly

⁷Harry Frankfurt, "Freedom of the Will and the Concept of a Person," in *The Journal of Philosophy*, Vol. LXVIII, No. 1, January 14, 1971, p. 18.

⁸*Ibid.*, p. 7.

⁹*Ibid.*, p. 10.

responsible for acting rightly and living well. Clearly if one desired to have such human dignity and to justifiably take pride in at least some of one's actions and one believed that one's being a Chisholmian agent is a necessary condition for each of these things, one would have a basis for preferring to be such an agent to being a deterministically conceived actor or a person whose only free actions are wholly uncaused.

The subject of uncaused, random actions brings us to a very recent criticism of agency theory. In his "Indeterminism and the Theory of Agency," Clifford Williams argues that the theory of agent causation cannot account for the possibility of moral responsibility since it does not remove the capriciousness that is said to characterize wholly uncaused actions.¹⁰ The success of his arguments hinges upon one central contention, namely that the concept of "self-movement" cannot be used to explain the occurrence of agent-caused actions at all.

Williams says that because, as stated by the theory of agency, there is no sufficient causal condition for the occurrence of the free action, the occurrence of such actions is capricious. He states that the agency theorist's attempt to explain the occurrence of the free action through recourse to the agent's bringing about the free action does not remove this capriciousness since it is the case that the agent's bringing it about either is produced by a sufficient causal condition or is not. If it is produced by a sufficient causal condition, then the agency theorist's conception of free action has been undercut. If it is not produced by a sufficient causal condition, then it is capricious and unexplainable.

Williams dismisses an apparently promising retort which an agency theorist can make, namely that the agent is causally responsible for the "bringing about of the action" as well as for the action. He says that this strategy will not work because one could ask if the agent is responsible for the "bringing about of the bringing about of the action" and so on, ad infinitum. Thus, Williams's point is that if the agency theorist persists in his claim that the occurrence of the free action is ultimately attributable to the causal power of the agent, the agency theorist commits himself to the existence of an infinite regress of alleged agent-causings which is causally ungrounded since no item in the series is "grounded" upon anything other than a capricious and unexplainable "bringing about of a bringing about of a bringing about . . ."

We will show that Williams's contention is incorrect. As we have seen, Chisholm's view is that there are cases in which it is neither causally necessary that the person perform the action which is actually performed nor causally necessary that the person not perform it. In such a case, the agent is the causal source of the action. The agent freely performs the

¹⁰Clifford Williams, "Indeterminism and the Theory of Agency," in *Philosophy and Phenomenological Research*, Vol. XLV, No. 1, September, 1984, pp. 111-119.

action: he causally contributes to his performing the action and is not caused to do so by antecedently existing conditions.

Thus, the action is not an uncaused, random or capricious occurrence. Rather, it is caused or brought about by the agent. To then ask, "But how is this bringing about of the action caused?", is to treat the agent's bringing about of the action as an event separate and distinct from the agent's action rather than as the very activity of the agent itself. (One could say that the agent's acting constitutes the bringing about of his action.) Moreover, in asking this question, Williams clearly begs the question, for he presupposes that the agent's bringing about of the action is an event whose occurrence has not been causally explained simply because there is no prior event which is said to have produced it (and thereby does not acknowledge the possibility that it was agent-caused), while this bringing about of the action is the very activity which, according to agency theory, has been caused by the agent.

Therefore, agency theory does not have to posit an infinite regress of alleged agent-causings in its attempt to explain the occurrence of agent-caused actions. This is required only on the question-begging assumption that the only type of causal explanation of an event is the specifying of an antecedent event which produced it. Hence, a theory of agency is capable of transcending the capriciousness that is said to characterize wholly uncaused actions if it can explain how the agent's bringing about of the given free action is not capricious, because, as we have seen, this "bringing about of the action" is the only exercise of agency that the agency theorist must posit. As we shall establish, this challenge can be met.

In outlining his theory of agent causation, Chisholm speaks of reasons, motivations and many other factors as causally contributing to an agent's free actions. (There are, he says, many actions which we cannot perform unless we have a purpose or motive for doing so.¹¹ If the agent performs such an action, the purpose or motive causally contributes to the agent's performing the given action by being a necessary condition of that performance.) Thus, Chisholm does not opt for a strict indeterminism in which the concept of event causation is held to be totally useless in explaining the occurrence of free actions. He simply maintains that there are instances in which such causally contributing factors do not constitute and are not part of a sufficient causal condition for the occurrence of the given action.¹² In such cases, the agent freely initiates his action.

Thus, there are two ways in which the theory of agent causation transcends the capriciousness that characterizes wholly uncaused actions. First, the theory states that reasons and motivations causally contribute to an agent's free actions: i.e., the agent is motivated to act in certain ways

¹¹Roderick Chisholm, *Person and Object* (London: George Allen and Unwin, 1976), p. 57.

¹²*Ibid.*, p. 68-9.

and acts with the intention of bringing about certain states of affairs which he believes will fulfill certain of his purposes. Such motivated and purposive free action is manifestly different from action which is wholly uncaused and thereby capricious and wholly unexplainable. Secondly, the theory of agent causation asserts that there exist agents who sometimes freely act in such a purposive manner, i.e., who initiate certain of their actions without being caused to do so. Clearly, this notion of free action differs radically from the concept of free action as wholly uncaused action. The latter type of free "action" is not, properly speaking, *performed* by anyone, but merely *happens to* the person whose "actions" they are.

This discussion of freely initiated purposive activity brings us to another criticism of agency theory. In his "Agent Causation and Reasons for Acting," Irving Thalberg raises the following objection:

I am baffled because we have a "responsible act" such that "the agent's undertaking" it has no "sufficient causal condition." Yet he has "reasons for acting as he did," which may include a desire and some "ought" thinking. Aren't these a sufficient condition or part of one? . . . But if we act for reasons, how can these constitute anything other than a sufficient causal condition, or part of one, for the movement of our body which fulfills our action plan?¹³

In order to demonstrate that an agency theorist can consistently speak of an agent's having reasons for his free actions, it will be helpful to examine Richard Taylor's distinction between two fundamentally different theories of the nature of thought:

The first theory . . . envisions thought to be a series or stream of things "within" one. My thinking, then consists of my *having* various things, such as ideas, impressions, images, feelings and other subjective or "private" things . . . My thoughts are, in any case, not literally created by me. In thinking, I do not make my own thoughts, I do not originate them . . .

The other theory is to the effect that thinking is sometimes an activity, that my thoughts are not in the ordinary sense *things* that merely arise and subside in my mind, but that they are sometimes *acts* that I literally perform.¹⁴

¹³Irving Thalberg, "Agent Causation and Reasons for Acting," in *Philosophia: Philosophical Quarterly of Israel*, VII (1978), p. 564.

¹⁴Richard Taylor, *Action and Purpose* (Englewood Cliffs: Prentice-Hall, 1966), pp. 155-156.

If a person's thinking is not simply a stream of mental entities which he is unable freely to direct, but rather is sometimes a realm of freely initiated activity, he can be causally affected by his previous mental states, e.g., his having reasons, purposes and motives at certain stages in a deliberate process, without these states being part of a sufficient causal condition for the occurrence of a given action. For if thinking is sometimes freely initiated, an agent could sometimes freely direct his deliberations by choosing some purposes, reasons and motives as being more worthy than others of guiding his deliberations and his action in a way which is not caused by his previous mental states.

Hence, in such cases, the person's having certain reasons, purposes and motives at the outset of his deliberations and at certain other stages in the deliberative process is not part of a sufficient causal condition for his acting as he eventually does. Yet, the having of those reasons, purposes and motives for the sake of which the agent freely decided to act causally contributed to his free action since they were integral parts of the deliberative process which culminated in his decision to act as he did. The existence of such factors causally contributed to the occurrence of the action in the sense of making its occurrence more likely than it would otherwise have been. (In some of these cases, this "making more likely" takes the form of being a necessary condition of the action, i.e, a condition without which the action would not only be less likely but would be impossible.)

Thus, elaborating Chisholm's theory of agent causation to include the proposition that an agent can freely initiate at least some of his thinking allows his theory to render intelligible agent-caused actions which are done for reasons. (We do not mean to suggest that Chisholm's theory unsupplemented by Taylor's thesis regarding agent-caused thinking is incapable of making clear how agent-caused actions that are done for reasons are possible, but simply that Chisholm's theory supplemented by Taylor's thesis is *clearly able* to do so.)

Having explained the nature of Chisholm's theory of agent causation and shown that agency theory can rebut successfully the aforementioned criticisms, we will conclude by suggesting two improvements of Chisholm's theory. The incorporation of Taylor's thesis that an agent can freely initiate at least some of his thinking would improve Chisholm's theory, because such an elaboration is a necessary condition of an agency theory's satisfactorily accounting for the invariable synchronization between an agent's free actions and his thoughts.¹⁵

In order to see that Chisholm's theory unsupplemented by Taylor's thesis does not make sense of the aforementioned synchronization, we shall imagine a person who is deliberating about whether to perform action A or perform action B. If he cannot agent-cause any of his thoughts, each and

¹⁵To the best of my knowledge, Chisholm does not subscribe to the view that there are agent-caused thoughts.

every step of his deliberative process (including of course the final step, the decision itself) would be produced by a sufficient causal condition. The question arises, "Will this person act in accordance with the decision arrived at through deliberation?" According to Chisholm, the subsequent action is freely initiated, even though none of the agent's thoughts are freely initiated.

When a person acts for a reason, i.e., acts with the intention of implementing a decision made through his deliberating, he often acts with various types of thoughts in mind. He may be deciding to implement his deliberative decision as he is in the very act of doing so. He may be thinking that he has made the right choice by so acting. He may be anticipating certain favorable outcomes from his so acting. He may be acting with various of the reasons for so acting before his mind. In short, the person often will be thinking in various ways which make sense given the way he is acting.

However, this congruence between thinking and acting is inexplicable if it is the case *both* that these actions are agent-caused *and* that these thoughts are not agent-caused. If each thought is produced by a sufficient causal condition, but some actions are not (and thus could have been otherwise), why does it almost always happen that those thoughts which accompany a person's actions jibe so well with those actions?¹⁶ Because the thoughts could not have been otherwise, they are on a "causal track" which has no gaps, no points at which there exists causal insufficiency. However, some of these actions are freely initiated and thus could have been

¹⁶It is true that there occasionally occur thoughts which are not synchronized with the actions they accompany, i.e., thoughts which presuppose propositions which are incompatible with certain truths about the nature of the action itself. For example, one is intending to make a lesser amount of a casserole by halving the quantities of the various ingredients. One thinks one is doing this, while one is actually doubling certain of the ingredients. Such gaps between thought and action may be severe and frequent or mild and infrequent. In the later case, they could be the result of simple inattentiveness. In the former case, they probably reflect severe psychological problems.

Nevertheless, it is the case *both* that in the *overwhelming majority* of cases, synchronization between thoughts and actions exists *and* that any plausible theory of agent causation would have to hold that many agent-caused actions are not out of sync in the aforementioned manner, i.e., that many agent-caused actions are not accompanied by thoughts which presuppose propositions which are incompatible with certain truths about the nature of the action itself. If a theory of agent causation did not hold this to be true, it would be condemned to a counter-intuitive absurdity—free actions either always or usually occur when the person believes that he is doing something else, i.e., when he does not know what he is doing.

otherwise than they were. Hence, these actions could have been out of synchronization with the thoughts which accompany them. Why is it that they *very rarely* are?

It would be ridiculously implausible to hold that this synchronization is merely coincidental. However, if it is not mere coincidence, what explanation could be provided except that these thoughts and those that precede them are, in the vast majority of cases, the dominant components of a sufficient causal condition of the action that is performed? It is not open to an agency theorist to explain this synchronization by saying that the person is not caused to act as he does by his thoughts, but rather acts on the basis of reasons. For the question is, "What accounts for the perfect synchronization (in the vast majority of cases) between thoughts which are caused by sufficient causal conditions and actions which are freely initiated and hence *not* produced by sufficient causal conditions?" Appealing to the proposition that the agent freely acts on the basis of reasons does not account for this synchronization because the agent's decision to act in X-manner on the basis of reasons A through C is merely one thought-event in the nexus of thinkings, each of which is produced by a sufficient causal condition. The question remains, "Why does a deterministic nexus of thinkings match up so well with the agent's freely initiated actions?" To put this question negatively, "Why is there not more than an *extremely occasional* lack of synchronization between such thoughts and actions?"

By not providing for the existence of agent-caused thoughts within the context of his theory of agent causation, Chisholm renders himself unable to explain the intimate relationship between thinking and acting in cases of freely initiated action. As a matter of common sense, we believe that the aforementioned synchronization is attributable to the fact that humans *act rationally* in the minimal sense of acting for reasons (as opposed to the maximal sense of acting for sufficiently good reasons). However, as we have seen, this cannot provide us with an adequate explanation of the aforementioned synchronization if it is the case *both* that the actions in question are freely initiated *and* that none of the thoughts in question are freely initiated. Thus, within the context of a theory of agent causation, if we are to be able plausibly to explain this synchronization between thoughts and free actions, we must assume that at least some human thoughts are freely initiated. As we shall see, this assumption will allow us to create such a plausible explanation.

As we have seen earlier in this article, if persons can agent-cause at least some of their thoughts, they can freely direct the course of their deliberations. By doing so, they freely initiate at least some of their actions: i.e., some human actions are such that for each of them it is the case *both* that at the time immediately preceding the agent-caused thoughts there is not a sufficient causal condition for the person to perform the given action nor is there a sufficient causal condition for the person not

perform the given action, *and* that the person causally contributes to his performing the given action.

Our explanation suggests itself. If agents freely initiate at least some of their thoughts, it is quite understandable that in cases of freely initiated actions the agent's thoughts and actions are well synchronized *both* because much of the causal contribution that the agent makes to his freely initiated action is that causal contribution he makes to his freely initiated thinking *and* because it is after all the same agent who is directing both the stream of his thoughts and the course of his actions. Of course, neither of these things would be the case if there were no agent-caused thoughts. Thus, in order adequately to account for this synchronization between thought and action which any plausible theory of agent causation must assert to exist in cases of agent-caused action, Chisholm's theory must be elaborated to include the proposition that persons agent-cause at least some of their thoughts.

There is another point on which Chisholm's theory as presently formulated is not fully adequate, namely his discussion of the ways in which reasons and motives can causally contribute to the occurrence of agent-caused actions.

Chisholm correctly notes that reasons and motives can serve as necessary conditions of agent-caused actions. He uses the killing of our friends or our family as examples of actions which are so abhorrent to us that we cannot undertake them without having some reason or motive for doing so. (As we have noted, for Chisholm "to undertake p" is "to act with the intention of bringing about p.")

Chisholm also rightly points out that reasons or motives can incline us to perform a given action without necessitating us to do so. He uses the example of a public official whose "scruples are sufficiently strong that he would never undertake to bring it about that he is offered a bribe," but "his desire for money is sufficiently strong that, if he were offered a bribe, he would be unable to refuse it."¹⁷ Chisholm defines "a motive's inclining but not necessitating a person's undertaking" in the following way:

he desires p; it is not within his power to undertake not-p; but it is within his power not to undertake p.¹⁸

Chisholm's account of these matters is satisfactory as far as it goes. However, it is not an adequate exploration of that category of states of affairs which may properly be described as, "a motive's inclining but not necessitating a given undertaking," if this is taken to mean, "a motive's making it more likely, but not necessitating, that a given undertaking occur."

¹⁷Roderick Chisholm, *Person and Object* (London: George Allen and Unwin, 1976), p. 69.

¹⁸*Ibid.*, p. 69.

If the theory of agent causation is true, surely there are cases in which a motive inclines but does not necessitate a given undertaking (in the sense of making that undertaking more likely while not necessitating that it occur) that do not fit Chisholm's aforementioned definition of "inclination without necessitation." Imagine the following situation. Joe is sitting in a park studying for an upcoming exam. Joe enjoys playing softball and has some desire to play softball at the present time. Imagine also that there are many softball games taking place in the park at present. It is within Joe's power not to undertake to bring it about that he is invited to play in a softball game. It is also the case that Joe would be able to resist an invitation to play softball if he were to receive one.

All of this is both internally consistent and conceivable. We would be correct in saying that Joe's desire to play softball "inclined but did not necessitate" him to play because it made him more likely to play than he would have been if he did not have this motive, but did not make it causally necessary that he play.¹⁹ However, this example would not qualify as "inclination without necessitation" as Chisholm defines it because Joe had the power to resist acting on the inclining desire even when invited to do so. Thus, Chisholm's definition of "inclination without necessitation" is too narrow to include all of those cases in which motives incline but do not necessitate us to perform a given action. Therefore, his account does not outline all of the ways in which motives or reasons may causally contribute to an agent's free actions.²⁰

¹⁹What we mean by, "made him more likely to play," is simply, "added additional causal power to his motivation to play." This does not imply that upon this occasion or any other occasion the agent in question will be unable to successfully resist such a motivation.

²⁰I would like to thank Peter Hare for his constructive criticism and his many helpful suggestions.