Agency in the Net of Reasons

Mike Arnautov, 2025

A persistent strand in philosophical conversation maintains that human actions are driven by reasons rather than causes, and that this distinction somehow preserves genuine agency in ways that a purely causal account cannot do. This view suggests that when we act for reasons – as opposed to being moved by causes – we exercise a form of control that escapes the deterministic web that governs the physical world. While I am happy to accept that conflating reasons and causes can be problematic, I shall argue that this distinction, however appealing, fails to deliver the promised could-have-done-otherwise libertarian agency. When we examine the structure of reason-based action carefully, we discover that it reproduces, in the mental sphere, the same problematic features that supposedly plague causal determination – only more so.

Needless to say, this is not a worry for compatibilists, for whom agency is simply an ability to make choices, and who thus see no tension between a physicalist universe and our experience of genuine agency. Similarly any attempts to construct a libertarian version of free will within the realm of causes (e.g. by invoking quantum randomness) are also unaffected.

Before proceeding, I need to define my terms for maximal clarity.

- By ‘reason’ I do not mean the faculty of reasoning, but rather a consideration that counts in favour of an action – for example, wishing to watch something on television constitutes a reason for turning it on. Reasons need not be singular or atomic: taking an umbrella because it might rain *and* because I do not wish to get wet can be treated as a single, compound reason. Emotional states may also qualify: how I feel about something – anger, affection, anxiety – may function as a reason for acting, whether or not I acknowledge it as such. More broadly, I include any factor – conscious or unconscious, affective or cognitive, internal or external (as in externalist views of mind), psychological, metaphysical, or otherwise – that an agent, or a theory of agency, might treat as explaining an action without appeal to causation. My aim is not to defend this maximalist conception of reasons, but to show that even on this most inclusive account, the structure of reason-based explanation fails to deliver what voluntarists mean by ‘genuine agency.’

- By ‘action’ or ‘acting’ I do not mean only physical action. Making a decision also counts as an action. Deciding to take an umbrella is an action whether or not I actually take the umbrella in the end. Not making a decision is also an action – a null one.

- Finally, I use ‘arbitrary’ and ‘arbitrarily’ to mean acting for no reason – not for a capricious reason, or an unconscious one, but genuinely no reason at all. This is structurally analogous to ‘randomness’ in the sphere of physical causation, though I do not claim any deeper identity. It is the absence of internal rationale, that I wish to highlight.

With these distinctions established, let us consider the relationship between reasons and actions. The following analysis traces the logical structure that any complete rational explanation must have, not the temporal psychology of actual decision-making. When I examine how reasons function in justifying action, I am concerned with logical relationships between different considerations, not with any sequential process of deliberation.

When I act, there are two fundamental possibilities: I act for no reason (i.e., arbitrarily), or I act for a reason.

If the second case obtains, two further possibilities arise: I may act for a reason necessarily, or I may act for a reason contingently.

If I act for a reason necessarily, then either I have but one available action, or I have a range of actions to choose from. In the latter case, I must have further reasons, to narrow the range of options until just one option is left.

If I act for a reason contingently, then either I act arbitrarily within the constraints imposed by my reason, or there is a further reason involved – requiring us to add that reason to the reasons already included in the overall reason and consider the expanded set.

Strictly speaking, we could imagine an infinite regress of reasons. However, since real agents are finite, this possibility reduces in practice to paralysis by analysis: a failure to act, which in itself is a ‘null action’ – determined by reasons in their infinite regress.

Thus, in practice, rational justification cannot regress indefinitely. The logical structure of any complete explanation of action must bottom out in one of two ways: either in complete determination by the totality of my reasons, or in placing a constraint on available actions.

The upshot is that my actions are either fully determined by my reasons, or I choose for no reason within constraints imposed by my reasons, or I simply choose for no reason. Appealing to deliberation, to reasoning, does not resolve the problem, since reasoning is itself a series of decisions based on reasons and thus does not escape the above logic.

Since my reasons must form a part of my overall decision-making state, regardless of how that is defined, this analysis reveals something rather deflating. What emerges looks entirely like a mixture of causal determinism and randomness, merely translated into the mental sphere. Crucially, this conclusion does not rely on establishing any connection between reasons and causes. It merely points out a structural parallel between determination by reasons and determination by physical causes.

If anything, the move from the realm of causes to the realm of reasons is problematic for libertarians. In causal discourse, one can at least gesture towards some mysterious ‘agent causation’ or other exotic metaphysical entities that allegedly transcend the determinism/randomness dichotomy. Whether these succeed is another matter, but the conceptual space seems available for such moves.

However, with reasons the logical constraints are much tighter. The very concept of a reason is defined by its role in explanations. One cannot coherently posit some influence on action that is ‘reason-like’ but not actually a reason.

This creates what we might call a ‘logical closure’ problem for libertarians who retreat to the reasons-based account. Any factor they invoke to explain how actions can be neither fully determined by reasons nor arbitrary will automatically count as an additional reason, thus falling under the above analysis. One cannot step outside the logical space of reasons while remaining within a reasons-based account of action.

It may be objected that I have defined 'reasons' too broadly. But the argument does not depend on the particular breadth of the definition. It holds so long as ‘reason’ includes whatever any given voluntarist theory might invoke to explain an action in non-causal terms. The point is not to insist on a preferred taxonomy of reasons, but to show that even on the most inclusive understanding, the structure of explanation fails to secure the agency it is supposed to preserve.

It could be further objected that I am implicitly assuming that states of mind are complete and exhaustive descriptions of the agent's decision-making state. However, the above argument does not depend on this assumption – as can be seen from a thought experiment that makes the dilemma particularly stark.

Suppose that at the point of making a decision, the universe is cloned into a large number of perfect copies. Extending this thought experiment beyond its usual scope, by 'universe' I do not mean merely the observable world, but the totality of everything that anybody might wish to include in the agent's deliberative state — internal or external (as in externalist views of mind), psychological, metaphysical, or otherwise, along with the complete history that led to the moment of decision. Do I make the same decision in all those cloned universes, or not?

The ‘free will’ hypothesis, if it is to deliver genuine agency, seems to require that the answer be ‘no’ – that I might choose differently across the various copies. But this leaves us with an uncomfortable question: for what reason do I act differently in some of the cloned universes? Since I have, by definition, exactly the same reasons in every case – including the same full set of internal and external factors – there appears to be no reason for the divergence. The only answer left is: I act differently for no reason at all.

This crystallises the dilemma in its clearest form. Either my reasons determine my choice – in which case I choose identically across all universes – or my reasons fail to determine my choice – in which case the variation across universes is unmotivated by any reason. There is no third option available within a reasons-based model that preserves genuine agency.

Perhaps this should not surprise us. After all, if reasons are to do any explanatory work – if they are to account for why we act as we do rather than otherwise – they must constrain our behaviour in ways that mirror the constraining power of causes. The moment we grant reasons this explanatory role, we invite back through the window the same deterministic worries we sought to banish through the door. They may wear the new metaphysical garb of determination by reasons rather than determination by causes, yet still appear to leave no room for could-have-done-otherwise, i.e. for voluntarist agency.

Prior Work and Context

The relevant literature is vast, and the above argument inevitably has parallels and echoes in various prior works on this subject. While this essay does not directly quote or explicitly engage in polemics with specific authors, its central dilemma regarding the determination by reasons and the ‘could-have-done-otherwise’ problem draws on a long tradition in free will debates. Readers familiar with the ‘standard argument against libertarianism’ (often associated with figures like David Hume, Peter van Inwagen, Derek Pereboom and Galen Strawson) will recognise conceptual similarities, although this paper differs in its critique of reasons-based accounts by basing it solely on their internal logical structure. Similarly, the paper does not explcicitly engage with the broader philosophical terrain surrounding agent causation and compatibilist theories of agency. The aim here is to present a freestanding logical argument, which, to the best of my knowledge, offers a novel, purely internal critique of the reasons-based voluntarist position without reliance on considerations like moral responsibility or empirical science to demonstrate its incoherence.

Disclosure: This essay was written with the aid of AI. Anthropic's Claude, Google's Gemini, and OpenAI's ChatGPT were invaluable in helping to organise the initial draft, offering critical responses, and suggesting improved phrasing for clarity and conciseness. All philosophical content, argumentation, and final responsibility remain my own.