Abstract: Leibniz is almost unique among early modern philosophers in giving final causation a central place in his metaphysical system. All changes in created substances, according to Leibniz, have final causes, that is, occur for the sake of some end. There is, however, no consensus among commentators about the details of Leibniz’s views on final causation. The least perfect types of changes that created substances undergo are especially puzzling because those changes seem radically different from paradigmatic instances of final causation. Building on my more general discussion of efficient and final causation in ‘Leibniz on Causation—Part 1,’ I will examine and assess some of the rival interpretations of Leibniz’s account of final causation.

In ‘Leibniz on Causation—Part 1,’ I argued that the changes in created substances have both efficient and final causes, on Leibniz’s view, and I explored various interpretations of how exactly Leibniz understands efficient causation. I also provided a rough sketch of the traditional Aristotelian account of final causation or teleology. In this tradition, the final cause is that for the sake of which a change occurs, or the purpose or end of some change. For instance, obtaining delicious coffee is the final cause of my trip to a coffee shop, and becoming a mature oak tree is the final cause of an acorn’s developments. After all, I go to the coffee shop for the sake of the coffee, and the acorn changes for the sake of becoming an oak. In both cases, according to the tradition, this end or purpose is a vital component of the explanation. Now it is time to take a keener look at what precisely the final causes of the changes in Leibnizian substances might be.

The interpretive difficulties and disputes surrounding Leibniz’s views on efficient causation that I described in ‘Leibniz on Causation—Part 1’ are nothing in comparison to the difficulties and disputes surrounding his views about teleology. As noted in Part 1, Leibniz repeatedly describes the realm of monads as the kingdom of final causes (see e.g. M §79; G 7:344/AG 319). He also associates appetitions with final causation, and since appetitions govern all changes in monads, this presumably means that all monadic changes have final causes. Yet, he is less than forthcoming about the details of his understanding of the final cause. As a result there are a number of different ways of interpreting Leibniz’s doctrine that the changes in created monads have final causes.
1 Final causation in deliberate actions

Let us start with the least controversial cases, that is, the deliberate actions of rational agents. Take, for instance, the coffee shop example: I stand outside of my favorite coffee shop, think about what to do, and decide to go in and get a large caffè latte for myself. We can explain the consequent changes in my perceptions, such as my perceptions of walking into the coffee shop and of the transaction with the barrista, in terms of final causation. I have chosen to pursue a certain end—namely, enjoying some latte—and perform a series of actions because I take them to be the best means to this end. In cases like this, the end is something that I judge to be good and that I pursue consciously.

Examples like the coffee shop, in which a rational agent chooses to pursue a particular course of action after a process of deliberation, are not terribly problematic. Yet, even here Leibniz scholars disagree about what exactly the final cause of the action is. To me it seems most plausible that the end, or perhaps the expected goodness of the end, is the final cause of the action, and that this end is the object of an appetition of the agent. In our example, then, enjoying a cup of delicious coffee is the final cause of my action. This strikes me as the most natural reading, given the traditional understanding of final causation; it is also Stephan Schmid’s interpretation (342). As seen, the Aristotelian tradition views the final cause as that for the sake of which the action is performed, that is, the end or purpose of the action. Moreover, the final cause is traditionally understood to be the object of an appetition or desire. Thomas Aquinas for instance tells us that the influence of a final cause consists in being sought or desired (Truth 3:41 [q22 a2 corp.]; cf. Pasnau, Thomas Aquinas 209). The final cause of my trip to the coffee shop, on this interpretation, is the great coffee I expect to get there, and in order for this final cause to lead to an action—that is, in order for the coffee to be a final cause—I must seek or desire it. On some traditional accounts, the agent must additionally represent the end, that is, I must represent or perceive the purpose of my visit to the coffee shop. According to those types of accounts, the end is not only the object of an appetition, but also the object of a thought or perception. And at least in paradigmatic examples of final causation, that is, deliberate, conscious agency, this appears plausible: the agent not only desires (or has an appetition for) the end, she also knows or perceives the end. In fact, she plausibly desires it because she perceives it to be good or attractive.

The textual evidence that Leibniz adopted the traditional account is admittedly thin, but so is the textual evidence against it. Leibniz simply does not explicate his understanding of final causation in much detail, at least in any text I have come across. This scarcity may itself be evidence that
Leibniz’s understanding of final causation is at least broadly traditional: if he did not find it necessary to explain what a final cause is for him, he may have assumed that his readers were already sufficiently familiar with the notion from their philosophical training. There is, however, one passage in which Leibniz does provide an explicit definition of different types of causation, and which seems to support my interpretation. This passage is from one of Leibniz’s tables of definitions, composed in the early 1700s: “The efficient [cause] is the active cause. The means is a cause which the efficient cause, intending the end, makes into a cause. The end is that, an appetition for which is the sufficient cause of striving [conatus] in the agent” (C 472). This passage does not answer our question as explicitly as one might have hoped. Still, it does suggest that for Leibniz, the end is the object of an appetition and that the agent or efficient cause brings about changes because it intends (or has an appetition for) the end. The understanding of final causation that this passage invokes thus appears to be very much in line with the traditional understanding: the final cause is identified with the end, which is the object of an appetition in the agent.

Yet, other interpreters appear to hold that prior states of the agent, rather than the objects of the agent’s appetitions, are the final causes of perceptual changes. Marc Bobro, for instance, claims that prior perceptual states are the final causes of later perceptual states insofar as they contain the reasons for these later states (§8), while Laurence Carlin argues that prior appetitions are the final causes of monadic changes (‘The Non-Aristotelian Novelty’ 83f.; ‘Leibniz on Final Causes’ 232). This would, however, be a radical departure from the traditional understanding of final causation, which identifies the final cause with the end itself, rather than with appetitions for, or perceptions of, the end. While, as seen, some traditional accounts hold that the agent must not only desire but also perceive the end in order to act for the sake of this end, they still insist that the object of this perception or desire, rather than the perception or desire itself, functions as the final cause (see Des Chene 194; Pasnau, ‘Intentionality’ 303). Hence, it strikes me as more likely that for Leibniz the end of the action, which is the object of an appetition—and plausibly also, at least in the case of rational agency, the object of a perception—is the final cause of the action.

The more a monadic change differs from the deliberate, conscious action of a rational agent, the more controversial and difficult things get. Some actions of nonrational animals, as well as some nonrational actions of human beings, might be similar enough to deliberate human actions to receive an analogous explanation. When a hungry dog dashes toward the pie I just dropped, we may be able to say that eating the pie, or the pleasure the dog expects from it, is the final cause of its running toward it. Similarly, when I absentmindedly scratch an itch, we may be able to say that relieving the
itch is the final cause of my action. Each of these examples involves a desire, or at least something desire-like, for a more pleasurable state; they are thus structurally similar to the deliberate actions of rational agents. The final causation at work in these examples is therefore still not an exceptionally challenging problem.

2 Problematic cases

The vast majority of the changes in created monads are completely different from the comparatively unproblematic examples that we just explored. First of all, many monads are so primitive that they lack memory, consciousness, and sensation altogether, and possess only very confused perceptions. They lack the capacity for pleasure and pain (M §19; G 7:529/W 504f.), and are always in a state similar to a “stupor” (M §24) or “a deep, dreamless sleep or … a fainting spell” (‘Principles of Nature and Grace’ §4, G 6:600/AG 208; cf. M §20). Unlike human and animal souls, then, such primitive monads cannot be described as pursuing pleasure in any straightforward sense. Furthermore, Leibniz holds that the perceptions of monads are infinitely complex and constantly change because he believes that each monad always perceives everything that is happening in the entire universe. Hence, even when I am sleeping and not aware of perceiving anything, I unconsciously perceive the whole world, and so do all other created monads. What are the final causes of all of the infinitely many changes in those unconscious perceptions?

Finally, even when I am awake, most of my perceptual changes are not deliberate or aimed at something that I consciously pursue. In fact, some of these changes—like the sudden throbbing pain in my leg when a dog bites me—are entirely unwelcome and undesirable. As seen in Part 1, Leibniz believes that there is no genuine interaction between created substance; instead, all changes that occur in a substance arise from its own depths. This means that even the most horrible, devastating things that happen to a created monad are strictly speaking caused by that monad itself, in accordance with its appetitions. Take for instance the death of a plant or animal (which, for Leibniz, means that the soul of that organism loses most of its body and continues to exist in a very impoverished state), such as a tree’s getting cut down. Changes of this type are farthest removed from the comparatively unproblematic case of deliberate rational agency. What might be the final causes of such unwelcome changes?
3 Three ways of dealing with problematic cases

There are at least three ways of understanding the final causation that is present in problematic cases like the ones just listed. One possibility is that even non-deliberate and unwelcome changes in created monads aim at some end that the agent unconsciously perceives as good. On that interpretation, the final cause of a non-deliberate and unwelcome change is at bottom very similar to the final cause of a deliberate rational action: in both cases, something appears good to the agent—consciously or unconsciously, as the case may be—and this apparent goodness (or, on some views, the perception of, or appetition for, this goodness) is the final cause of an action aimed at this apparent good. This interpretation is favored by a large number of Leibniz scholars, so we can call it ‘the standard interpretation.’ Another possibility is that unwelcome changes are instances of final causation only insofar as they are intended by God and serve a purpose in God’s plan for the world. Donald Rutherford (‘Laws and Powers’ and ‘Leibniz on Spontaneity’) and Marleen Rozemond defend versions of this interpretive strategy. They hold that while some monadic changes are not explained by what the agent perceives to be good, these changes are nevertheless instances of final causation because of what God knows to be good. A third possibility is that neither God’s intentions nor a perception of something as good are elements of the final causation of such monadic changes. Instead, the presence of an appetition for a particular state is sufficient for final causation. This is the interpretation that I find most plausible.

Each of the three interpretations just outlined has distinct advantages. The standard interpretation, according to which unwelcome changes aim at something that the agent unconsciously perceives as good, provides an appealingly unified account of monadic change. After all, this interpretation views all monadic changes as instances of the pursuit of something that the agent perceives as good. There is also some textual evidence for this way of understanding Leibnizian final causation. Leibniz writes to Queen Sophie Charlotte in 1704, for instance, “everything in the soul happens … in accordance with perceived good or evil. … [E]ven in our instinctive or involuntary actions … there is in the soul an appetition for good or an aversion to evil which directs it” (G 3:347/WF 224; translation altered). A similar passage occurs in a letter to Lady Masham, also from 1704: “Everything … comes down to a present state combined with a tendency towards changes, changes which are brought about … in the soul by perceptions of good and evil”
Both of these texts, taken at face value, appear to state that appetitions for, and perceptions of, the good play a role even in imperfect monadic changes. Yet, the standard interpretation also faces some problems. First of all, it is difficult to understand in what sense one of the most primitive monads—such as the unconscious monads in inanimate objects and plants—can perceive anything as good or as bad. True, Leibniz thinks that even those impoverished monads perceive everything in the entire universe. As a result, they do of course perceive good and bad things. For instance, they perceive the rain that brings relief to a drought-ridden landscape, and they perceive the earthquake that kills thousands of people and animals. But perceiving something that is good (or bad) and perceiving something as good (or as bad) seem to be two completely different things. By way of analogy, suppose someone gives me a list of statements about American football. Some of the statements are true and others false. Yet, knowing nothing about football, I cannot tell which ones are true and which ones false; I am in fact so ignorant that none of the statements even seem true (or false) to me. In this example, I perceive statements which are true and statements which are false without perceiving any of them as true or as false. The same, I submit, can be the case for good and bad: it’s possible to perceive things that are good or bad without perceiving them as good or as bad. And according to the standard interpretation, it is the latter that all monads must be able to do. In fact, they must perceive things as bad or as good all the time because they are constantly changing internally, and all of these changes are supposed to be explained by what these monads perceive as good or as bad. This, to me at least, seems extremely implausible. Being able to perceive things as good or as bad strikes me as a highly demanding type of cognition that is far beyond the capacities of the most primitive monads, and that even human beings do not engage in at all times.

Another problem with the standard interpretation is that even if all monads were able to perceive things as good and as bad, it is unclear how such perceptions could possibly explain transitions to horribly imperfect states. How does what I perceive as good explain the throbbing pain in my leg, and how does what a tree perceives as good explain its being cut down? In many cases, what happens to a finite substance simply is not good for that substance. Getting cut down, for instance, is not good for the tree in any way; it is a significant loss of perfection. Of course, the tree could nevertheless perceive the change as good for itself (assuming for the sake of argument that trees can perceive things as good). Nobody’s perfect, after all, and trees in particular are not the sharpest of creatures, to put it mildly. If they can perceive things as good, they probably make all kinds of silly mistakes. Still, the assortment of things that would have to appear good to a tree in
order to explain all of the changes that occur in it is quite astonishing. It would arguably have to include things like having its bark eaten by beetles, suffering from drought, and getting some of its roots chopped off. Do we really want to say that according to Leibniz, the tree perceives all of those things as good? Or perhaps I am not giving the tree enough credit: instead of mistakenly perceiving those things as good for itself, maybe the tree correctly perceives them as good for other creatures, or for the overall perfection of the world. It is true, after all, that every event in this world serves a purpose and contributes to the perfection of God’s creation. Might the tree undergo changes that are bad for it because the tree correctly perceives these changes to be best overall? Here again, I worry that such noble perceptions far outstrip the capacities Leibniz attributes to the most primitive monads.

Rutherford’s and Rozemond’s interpretations, according to which non-deliberate monadic changes have final causes only insofar as they serve God’s ends, can avoid these problems. After all, their interpretations are not committed to the implausible claim that all monads can perceive things as good or as bad, nor to the claim that even the least welcome changes aim at something that the agent perceives as good. Moreover, this type of interpretation fits well with certain strands of the Aristotelian tradition. As mentioned in Part 1, some medieval Aristotelians believe that only rational beings can set ends for themselves or for other beings, and that only the ends set by rational beings—ends that those beings recognize as ends—can be genuine final causes. On this picture, the actions of non-rational creatures have final causes only insofar as God has set ends for these creatures. If Rutherford’s and Rozemond’s interpretations are on the right track, Leibniz accounts for non-deliberate and unwelcome changes in a parallel way. The throbbing pain in my leg, for instance, serves a purpose and has a final cause, on this interpretation. Yet, that purpose or end is ultimately God’s: God decided to create the best of all possible worlds, and the pain in my leg is an integral part of that world. My pain contributes to the overall perfection of the world that God chose to actualize, and that is in a sense the purpose or final cause of the pain.

Rutherford’s and Rozemond’s interpretations, then, are plausible in at least certain respects. I agree with them that every monadic change is part of God’s plan and contributes to the overall perfection of this world. When God created the world in accordance with his perfect, all-encompassing plan, he imposed a teleological order on the world: everything that happens in the world serves God’s purposes and can consequently be described as having a final cause. Yet, there are reasons to doubt that this is the whole story. Take again the example of the tree getting chopped down. It seems right that we can in principle explain this event with reference to God’s plan, or to
the overall goodness of the world: even though we don’t know the details, there is a story to be told about how the world would be less good if the tree were not cut down. So it is true that the tree’s tragic end is included in God’s plan because God knew it would contribute to the overall goodness of the world. In this sense, the overall goodness is the purpose of the tree’s demise. But is that the fundamental explanation for the changes in the tree? I don’t think so. It is extremely important to Leibniz that the natural world be intelligible internally, and in fact, that the changes in each created substance be intelligible simply in virtue of that substance’s nature or internal features. Something about the tree’s nature, and perhaps its prior states, must explain all of the changes that occur in the tree. Moreover, Leibniz appears to deny that the tree has this nature or these states because of God’s plans. God could not have given a different nature to this tree; finite substances are package deals, as it were, and God simply discovered the complete concepts of these substances in his intellect. This means, I think, that God’s plan or God’s knowledge of what is best overall cannot be the fundamental teleological explanation for the changes in a created monad; it can only be the fundamental teleological explanation for the monad’s existence. If that is correct, there simply is no bottom-level, teleological explanation for unwelcome or non-deliberate changes on Rutherford’s and Rozemond’s interpretation. The only teleological explanation of these changes is a higher-level one in terms of God’s plan, which presupposes a more fundamental, internal explanation of the monad’s states. I take this to be a cost, because Leibniz talks as if final causation plays a fundamental role even in those imperfect changes.

The problems with the standard interpretation as well as with Rutherford’s and Rozemond’s alternative that I have just sketched make it worth exploring the third way of interpreting Leibnizian final causation. According to this last interpretation, being a final cause does not require appearing good to anybody, be it to God or to a created monad. The basic idea is, then, that we can simply take perceived goodness out of the equation altogether when we are dealing with unwelcome or non-deliberate changes. On the standard interpretation, as we have seen, changes in a monad are teleological only when the end that the monad is pursuing appears good or desirable to this monad. Similarly, Rutherford and Rozemond think that non-deliberate monadic changes are teleological because the monad strives for some end that God knows to be good. According to the alternative interpretation that I prefer, on the other hand, the ends that a monad strives for do not have to appear good to anybody. Instead, final causation is present in non-deliberate actions simply because the agent strives for particular states in accordance with its nature; possessing and acting on natural appetitions is sufficient for final causation. For instance, the monads in an acorn do not need to
perceive the process of growing into an oak tree as good. They naturally strive for it, and that in itself is enough to make their activity an instance of final causation. Similarly, I do not need to perceive pain as good in order to possess an appetite that aims at a painful state. Acting based on appetitions, whether or not the objects of these appetitions appear good to anybody, qualifies as final causation, on this interpretation. The object of the appetite is the final cause of the change because this object is what the agent aims at or strives for.\textsuperscript{12}

4 Conclusion

As we have seen, there is significant disagreement about Leibniz’s account of final causation. First of all, some interpreters (myself included) hold that the final cause is the object of an appetite as well as, in at least some cases, of a perception; others hold that the final cause is itself an appetite or a perception. Moreover, there are three completely different ways of explaining the teleology present in unwelcome or non-deliberate monadic changes. According to the standard interpretation, we can explain even those changes in terms of what appears good to the agent; the main difference between deliberate and non-deliberate actions is, accordingly, that the agent’s perceptions of some end as good are conscious in the former and unconscious in the latter. According to an alternative interpretation put forward by Rutherford and Rozemond, we can explain unwelcome changes teleologically in terms of what God knows to be best, but not in terms of what appears best to the finite agent. Finally, according to the third interpretation, the kind of teleology present in unwelcome changes does not require that the end appear good to anybody. Instead, the fact that these changes are explained by natural appetitions or strivings for particular states is sufficient to make these changes instances of final causation. Yet, as I have shown, there is no consensus among interpreters about any of these issues. Leibniz’s views on final causation remain an active area of research.

Works Cited


1 See M §79; ‘Principles of Nature and Grace’ §3, G 6:599/AG 207; ‘Table of Definitions,’ C 472.

3 See Aquinas, The Power of God 81 [q3 a15], Summa Theologica 2:616 [IaIIae q6 a1, corp.].

4 See Jolley 607, Bolton 191, Carriero 134, McDonough 197, Phemister 223.

5 One way out of this problem is the following: maybe all it means to say that a primitive monad perceives \(x\) as good is that this monad is motivated to pursue \(x\). In other words, maybe having an appetition for \(x\) is sufficient. The idea might be that if a monad has an appetition for \(x\), then by that very fact, \(x\) is “appetible” or worthy of being pursued for this particular monad. If that is all that it means to say that \(x\) appears good to the monad, however, this interpretation comes down to the same thing as the third interpretation I will consider below.

6 Some proponents of the standard view have a more sophisticated account of transitions to less perfect states. See for instance Bolton, who explains such transitions roughly in terms of the mutual accommodation of desire-like appetites for regular bodily motions (190f.). Yet, even if this works, my first criticism still stands: it strikes me as implausible that primitive monads can perceive anything as good.

7 In fact, for Leibniz even the motivation of rational souls is always egoistic: we can only do what we take to be good for us (see e.g. ‘Opinion on the Principles of Pufendorf,’ Dut IV.iii.277/Riley 67).

8 Thomas Aquinas at least in some passages endorses this kind of view; see e.g. The Power of God 14f. [q1 a5].

9 This is one of Leibniz’s favorite objections to occasionalism. See e.g. ‘On Nature Itself’ §15, G 4:515/AG 165; ‘Considerations on Vital Principles,’ G 6:541/L 587.

10 Leibniz sometimes talks as if God impresses his plan on creatures by giving them a particular nature (e.g. ‘On Nature Itself’ §6, G 4:507/AG 158f.), but other texts make it clear that this is only a loose way of describing God’s selection of possibles. In the Theodicy, for instance, Leibniz tells us that God’s decree to create the best possible world “does not change anything in the constitution of things. He leaves them just as they were in the state of mere possibility, that is, he changes nothing, neither in their essence or nature, nor even in their accidents, which are perfectly represented already in the idea of this possible world” (T §52).

11 After all, the monad fits into God’s plan and is chosen by God because it will spontaneously and naturally go through a particular series of states. The reverse is not true: the monad does not go through this particular series because it fits into God’s plan or is chosen by God.

12 For a much more detailed criticism of the first and second interpretation, and a more thorough defense of the third, see Jorati.