**Negligence and self-trust**

Samuel Murray

*Providence College*

*Abstract*

Why are we accountable for negligent wrongdoing? This paper develops a contractualist account of accountability for negligent wrongdoing rooted in maintaining self-trust. Displays of negligence threaten the self-trust needed to exercise planning agency. People thus have reason to take responsibility for being negligent to defeat higher-order evidence about the unreliability of one’s planning agency. Individuals are rationally required to take responsibility for negligence in virtue of the demands of planning agency. One novel implication of this view is that taking responsibility for negligence grounds our holding others responsible for negligence.

*Keywords*: negligence; blame; self-trust; planning; control

It is only the precise knowledge of his own empirical character that gives the human being what we call *acquired character*: someone possesses this who knows his own properties good and bad, precisely, and thereby knows for certain what he may entrust to himself and demand of himself, and what he may not.

* Schopenhauer, ‘Prize essay on free will’, pp. 69-70.

**0**

Suppose Randy promises to meet a friend for lunch tomorrow. The next morning, he gets distracted by work and forgets about lunch. Randy seems culpable for forgetting. We wouldn’t think twice if we heard that Randy’s friend was angry about being stood up or that Randy was embarrassed about forgetting. *That* Randy is culpable seems obvious. But it is puzzling *why* this is so.

Blame seems appropriate for some negligent misdeeds, and Randy is no exception. But the standard justifications for such reactions seem equally *inappropriate*. People are responsible for their choices and the actions following on them. People are responsible for actions that say something about *who they are*. People are responsible for what is in their control. None of these apply in Randy’s case. He does not decide to forget the lunch. Forgetting the lunch need not tell us much about who Randy is. In fact, Randy might typically be conscientious, although knowing that would not assuage the feeling of culpability.[[1]](#footnote-1)

Instances of negligent omission—wrongdoing where one is unaware of doing anything wrong until it is too late—lack the typical anchors for attributing accountability.[[2]](#footnote-2) Thus, we have a puzzle: while people seem sometimes to be accountable for their negligent wrongdoing, there is nothing that readily justifies imputing accountability to them. If anything, standard *excuses* seem to apply. Randy seems a victim of faulty memory, which implies that he is unfortunate, but not accountable. Nevertheless, it is part of our ordinary repertoire of moral appraisal to sometimes find people accountable for negligence, but not because of anything they do.

Recent work on normative theories of blame have focused on two different first-order responses to this puzzle. One response tweaks the conditions on responsibility derived from some overarching ontology of responsible agency. The other response is to deny that people are ever culpable for negligent omissions. The debate over these responses is lively because different theories of responsible agency make opposed predictions about the normative status of negligent wrongdoing (Sripada, 2015; Murray et al., 2019).

The present essay does not directly address this first-order debate. Instead, I presume my favored solution to the puzzle, which states that people are culpable for negligence in virtue of exhibiting control over it. Generally, an individual is responsible for some wrongdoing only if they could control whether the wrongdoing occurs. Thus, my solution falls under the “Fair Opportunity” model of responsibility (Brink, 2021). Insofar as we retain our ordinary deliberative and choice-making capabilities in many circumstances of negligent wrongdoing, we *can* often control whether some negligent wrongdoing occurs and, hence, we will sometimes be culpable for negligent wrongdoing. In Randy’s case, he is culpable because he should have recalled the lunch and could have recalled the lunch. In the abstract, if an agent should have avoided being negligent and could have avoided being negligent, then the agent is *prima facie* culpable for being negligent. This control-based (or capacitarian) account has two relevant elements: (1) a moral principle that prohibits negligence, and; (2) an ability principle that grounds blame for negligent wrongdoing in agential capacities to avoid negligence. For ease of expression, let’s call this the *Control Principle*.[[3]](#footnote-3)

In lieu of fully defending this account, I offer three reasons why I prefer it to others. First, the Control Principle seems psychologically apt, as moral judgments of negligence are sensitive to perceived agential control (Kissinger-Knox et al., 2018; Murray et al., 2019; Murray et al., 2023) and information-processing underlying judgments of blame are hierarchically organized in a way predicted by the Control Principle (Malle et al., 2014). Second, the Control Principle is conservative with respect to culpability for negligent wrongdoing, unlike alternative accounts of the metaphysics and epistemology of moral responsibility (Rosen, 2004; Doris, 2015; Sripada, 2015; Levy, 2017; Talbert, 2017). Finally, the Control Principle has, arguably, the same degree of explanatory power as alternative principles (see Clarke, 2017).

My focus here is on an underappreciated issue about the metaethical foundations of the Control Principle. Recent defenses of the Control Principle assume that it is justified on instrumental grounds (Pettit and McGeer, 2015; Murray and Vargas, 2020; Vargas, 2020; Vargas, 2021).[[4]](#footnote-4) Non-instrumental frameworks tend to pair with first-order views of responsibility that do not locate the source of responsibility in control. This pairing of metaethical and first-order views stems partly from a seeming incompatibility between control-based accounts and non-instrumental justifications of our responsibility practices.

In what follows, I want to outline the apparent incompatibility of the Control Principle within a non-instrumental framework (§1) and how instrumentalism justifies the Control Principle (§2). I then argue that when we understand how planning agency gives rise to the possibility of negligence, we can explain why planning agents are rationally required to take responsibility for negligence. This provides a contractualist foundation for the Control Principle (§3). Finally, I compare the account developed here to similar proposals to draw out some implications of the view (§4).

**I**

The rituals of responsibility fall under the part of morality that characterizes how to live well with others. As such, I am interested in whether the Control Principle is consistent with the basic features of moral principles. This section argues that there are reasonable grounds for thinking that the Control Principle could not be part of any internally coherent *moral* practice of responsibility (while some merely coercive practices of responsibility might not generate the same problem, any moral practice cannot be merely coercive).

Moral principles are fundamentally fair, where fairness consists in mutual acceptability. Principles are mutually acceptable when our use of them in practical reasoning can be justified to others and we can enjoin others to use these principles in similar circumstances. As such, fair principles are apt for use in practical reasoning and moral appraisal.[[5]](#footnote-5) Scanlon notes that fairness as mutual acceptability arises from our recognizing others as rational agents:

Morality is a system of co-deliberation, and moral reasoning is an attempt to work out principles that each of us could be asked to employ as a basis for deliberation and accept as a basis for criticism. Seeking such principles is part of what is involved in recognizing each other’s value as rational creatures (1998, p. 268).

Thus, moral principles must be suitable for use in practical reasoning and moral appraisal because of our respect for rational agency. What makes principles suitable as grounds for moral criticism and deliberation is that rational agents see such principles as furnishing reasons to act. Accordingly, when people violate reasonable principles, one can demand an account of how the individual’s practical standpoint seemed to license such norm-violating behavior.[[6]](#footnote-6)

The Control Principle seems ill-suited for practical reasoning in two ways. Note that the deliberative dimension of fairness implies two different criteria for assessing whether a principle is fair. First, because fair principles are fundamentally *reason-giving*, grasping the content of a principle must provide some distinctive motivation for compliance. Second, fair principles should make possible rationally settling on what to do. These are distinct criteria:

(*Motivation*) If P is a fair principle, then an individual’s grasping that P should motivate behaving in P-compliant ways.

(*Deliberation*) If P is a fair principle, then an individual’s grasping that P should inform rational deliberation about what to do.[[7]](#footnote-7)

Does the Control Principle provide distinctive motivation to act non-negligently? It seems not. In acting negligently, an individual fails—unwittingly—to follow through on some intended course of action. This means that, for any instance of negligent behavior, the agent had some previously acquired (and unrelinquished) intention that fails to shape their behavior at the appropriate time. But the agent’s motivation for action can be accounted for exclusively in terms of the underlying intentions. For example, Randy is motivated to go to lunch not because of understanding prohibitions against negligence, but because he intended to go. Any motivational force of the Control principle seems redundant with the motivational force of the mental states that manifest intentional agency.

Even if the Control Principle provided distinctive motivation, it also seems unsuitable for deliberation because it generates a kind of practical neuroticism. A plan, as a structured set of intentions to act in the non-immediate future, necessarily represents temporally extended activity. In the simplest case, the agents intends at *t*1 to A at *t*2. This intention must be held in mind either occurrently or dispositionally. Either way, the agent is susceptible to negligence. If the intention is maintained occurrently, then some interference factor or distraction might result in the agent’s “losing track” of the intention. If the intention is maintained dispositionally, then the agent must rely on cues to bring the intention to mind at the right time. These cues may be endogenous or exogenous, but in either case one might not perceive the cues or perceiving the cues might fail to trigger the disposition. In each case, the plan is disrupted in virtue of factors grounded in personal agency. Even when plans are derailed due to faulty external cues, such cues only become relevant because the agent decides to maintain the intention dispositionally. Hence, any plan is susceptible to neglect and, thus, we always have reason to adjust plans accordingly.[[8]](#footnote-8) Additionally, as this discussion shows, all negligence is preventable in some sense. Intentions could be managed differently or more cues could have been utilized. For example, Randy could have left an extra note for himself about the lunch. There is always something that one could have done to avoid being negligent.

This might seem too strict. Surely there are stages where a plan is reasonably constructed, and adjusting due to the possibility of negligence is, well, neurotic. What makes some plan reasonably constructed is that potential plan-disruptions are too remote to merit revision. But the mere remoteness of a considered plan-disruption is not obviously a reason to dismiss the consideration. Absent some account of why remoteness permits non-consideration, it seems that grasping both the Control Principle and the preventability of negligence either prompts continual redeliberation *or* requires arbitrarily ending deliberation despite considered reasons in favor of continuing deliberation. There is no possibility of rationally settling on a plan and, hence, no possibility of *rational* deliberation. For that reason, the Control Principle cannot (seemingly) structure rational deliberation.

These could be issues only with the moral prohibition against negligence. However, there might be no strict moral prohibition against negligence. The moral prohibition against negligence is just shorthand for whatever prescriptions and proscriptions are associated with more fundamental obligations arising from our promises and assurances. Similar problems about the *other* dimension of the Control Principle—the ability element—can also be raised. The ability element states that some individual is *prima facie* culpable for their negligence only if they could have avoided being negligent. An agent could have avoided being negligent only when it is up to them whether they are negligent or not. The case of Randy illustrates that negligent wrongdoing need not reflect the balance of one’s preferences and values, even at the moment of wrongdoing. If we think that the actions that are up to us are just those that causally depend on our values, desires, and decisions, then negligent wrongdoing, by definition, falls outside the scope of what is up to us.

This highlights a higher-order issue with the Control Principle. People participate in rituals of responsibility because doing so secures various kinds of otherwise unattainable social goods. These goods justify our bearing the occasional burdens of blame and sanction. We agree to open ourselves up to blame only because we decide whether we are penalized. We attribute to autonomous agents only those actions that reflect on their values and commitments and, thereby, hold one another accountable for only those actions that are truly “ours”. Each of us, as rational and autonomous agents, expects—most of the time—to act on the basis of defensible decisions (Lenman 2006, pp. 23-24). But just as one’s rational autonomy grounds the reasonableness of participating in an occasionally harmful practice of responsibility, the same autonomy would lead one to eschew participating in a system that occasionally assigns penalties for negligent wrongdoing (Lenman 2006, p. 25). Such wrongdoing is seemingly unpredictable (Rudy-Hiller, 2019; Sripada, 2019), deliberatively unassailable (as the argument for practical neuroticism shows) and need not indicate anything about who we are deep down (Amaya, 2022).

**II**

Some reject the Control Principle in favor of principles that explain culpability in terms of relations of value- or self-expression. On these alternative views, someone is culpable for wrongdoing only when such wrongdoing is causally produced by objectionable evaluative commitments with which the agent identifies (Sripada, 2016; Doris, 2015). In short, if one is *prima facie* accountable for some behavior, then that behavior expresses one’s self. Call this the *Deep Self Principle*. The Deep Self Principle shares with the Control Principle the idea that culpability for wrongdoing requires some robust relationship between agent and action. The Deep Self Principle grounds culpability in a causal relation between (some subset of) values and actions, while the Control Principle grounds culpability in a counterfactual relation of manipulability. Either way, the basic normative structure is preserved. The explanatory scope of both Principles is nearly identical (Sripada, 2015). One point of divergence is over culpability for negligent wrongdoing (Murray, 2017). The Control Principle seems to have the conceptual resources to explain such culpability, and this aspect of the Principle has been touted as a virtue of the framework (Clarke, 2017). The Deep Self Principle explicitly precludes accountability for negligent wrongdoing, unless such wrongdoing manifests objectionable lack of care or habitual negligence (Sripada, 2015).[[9]](#footnote-9) Randy’s negligent wrongdoing, by stipulation, indicates neither an objectionable lack of care nor habitual negligence. If such cases are possible—and they seemingly are—then accepting the Deep Self Principle comes at the cost of explaining accountability for negligent wrongdoing.

Many adherents of the Control Principle have done so under the banner of instrumentalism about responsibility (McGeer and Pettit, 2015; Vargas, 2020; Murray and Vargas, 2020). Instrumentalism refers to a family of views that justify attributions of responsibility in terms of their effects (Jefferson, 2019). Importantly, the justifiability of blame derives not from any notion of fairness or desert, but from the general capacity for moral improvement, as Anneli Jefferson explains: “…the fact that their moral agency is susceptible to being fostered and scaffolded by being held responsible is also what makes an agent…the right kind of target for these practices” (2019, p. 559). This suggests that no particular specific capacity is strictly necessary for being responsible for wrongdoing. Instances of negligent wrongdoing bring out this implication. Randy might lack the specific capacity to recall the meeting because he forgot. However, if Randy has a general receptivity to moral criticism and subsequent improvement, then the instrumentalist view might characterize blame as justifiable. On most instrumentalist views, the justification for blaming Randy does not derive from anything about Randy’s particular situation; rather, blaming is justified because in general agents will exhibit improved moral sensitivity and get better at avoiding negligence in similar kinds of situations overall in virtue of the practice’s prescribing blame in this way.

This criticism might not seem to apply to instrumentalist frameworks that specify the conditions on responsible agency independently of the teleological aims of the practice (Vargas, 2013; 2021). However, even on Vargas’ instrumental framework, susceptibility to blame is a mechanism by which people tune their moral sensitivities over time: “One way we extend our capacities into new contexts is to, at some point, be vulnerable to blame because we had a capacity that went unexercised” (2020, p. 410). For Vargas, an agent has an unexercised capacity to respond to reasons when, from the perspective of an ideal observer, the agent recognizes and responds to such reasons in a suitable proportion of relevant words, where thresholds of relevance and suitability are a function of some combination of the agent’s intrinsic features and the observer’s overarching interests in cultivating moral agency over time (Vargas 2013, p. 220). This generates the result that Randy has the capacity to avoid being negligent because in a suitable number of nearby worlds, Randy remembers the meeting.

While Vargas can explain some aspects of capacity talk on his account, the notion of capacity at issue does not secure a robust notion of fairness. Because some blame is justifiable in virtue of *extending* capacities, S could justifiably be blamed for failing to *A* at *t* because a practice that sanctions such blaming tends to bring it about that agents like S tend less often to fail to *A* in the future. Thus, one must have a general capacity where successfully *A*-ing falls within the scope of the capacity, but one need not possess the specific capacity for successfully *A*-ing. What is fundamentally important for blame on the instrumentalist account is that when S fails to A, she is responsible if attributing blame on the basis of failing to A generally brings it about that agents like S fail to A less often in virtue of being so blamed. Of course, Vargas might respond that what makes it the case that such attributions tend to have this effect is that S-like agents have the capacity to recognize and respond to reasons in favor of *A*-ing. Note that on this view, however, acting *as if* S had the relevant capacity would be an equally justifiable basis on which to impute culpability to S for failing to A. This is what makes instrumentalist frameworks fundamentally incompatible with deontological constraints on fairness. Letting instrumental considerations dictate the scope of permissible blame blurs the distinction between having a specific capacity and acting as if one has a specific capacity. Blaming constrained by fairness can target only those individuals who *have* specific capacities relevant to their wrongdoing. Instrumentalists can flout this constraint.

In short, instrumentalist frameworks explain responsibility for negligent wrongdoing at the expense of a fundamental fairness constraint. This is not a reason for or against instrumentalism. There is no decisive intuitive evidence in favor of a non-instrumental theory of responsibility over an instrumentalist one, and there is even evidence that instrumental considerations are associated with differences in folk judgments of responsibility (De Brigard and Brady, 2013; Nadelhoffer et al., 2013). But there remain good reasons to explore alternatives to the instrumentalist framework of responsibility for negligent wrongdoing. One such reason is that instrumentalism assumes the existence of practices as part of the instrumental calculus. Thus, we are asked to consider the aims of a *practice*: “If responsibility *practices* are to be what they present themselves as being, it is not enough that we *find ourselves with them*. They must be justified, at least if we are to retain them in *anything like their current form*” (2020, pp. 405-6; emphasis added). Practices can be better or worse, according to various metrics, but they must all be coherent if they are truly practices. The question of whether norms assigning culpability for negligent wrongdoing are *instrumentally preferable* is distinct from whether such norms could be part of any practice of responsibility whatsoever. Given the arguments from motivation and practical neuroticism, it seems that the Control Principle would fundamentally destabilize deliberation in creatures whose powers of reasoning resemble ours. Thus, setting aside questions of fairness, instrumentalist explanations of blame for negligent wrongdoing do not settle questions about the coherence of the Control Principle within practices of responsibility more generally.

**III**

The Control Principle seems incompatible with fundamental considerations of fairness. The Deep Self Principle is compatible with these considerations but characterizes the accountability relation between agent and action in a way that precludes accountability for negligent wrongdoing. The Control Principle might be coherent within an instrumentalist framework, but this framework is incompatible with deontological constraints on fairness. Is there a way to make sense of the Control Principle while respecting constraints on fairness? I believe there is. To show this, I propose to assess the puzzle of culpability for negligent wrongdoing from a different starting point. Often, discussions of the puzzle begin with some formulation of the conditions on responsibility and show that, despite our intuitions to the contrary, standard conditions do not warrant our attributing blame for negligent wrongdoing. The discussion then focuses on how to amend the conditions to accommodate our intuitions (Ayars, 2021; Bernecker, 2018; Clarke, 2014; Murray, 2017; Rudy-Hiller, 2019). Rather than starting from the conditions of responsibility, I start from the conditions on negligence: what about us makes negligence possible? In following this thread, we can eventually develop an adequate contractualist framework that encompasses the Control Principle.

The rudiments of intentional psychology are insufficient for making negligence possible. Lots of organisms can be plausibly interpreted as having intentions. But although many organisms exhibit intentional agency, fewer appear capable of negligence. In acting negligently, an agent fails to act in a way that aligns with available considerations that a reasonable person in similar circumstances would have noticed. Thus, negligent failure cannot arise due to environmental interference or lack of skill—a reasonable person could be expected to miss relevant considerations under these conditions. But neither can the failure stem from deliberate disregard—such would be reckless rather than negligent failure. What, then, makes possible those failures constitutive of negligence? My suggestion is that the possibility of negligence arises from the unique psychology of planning agency, which consists in the ability to form intentions for the non-immediate future that structure subsequent deliberation (Bratman, 1987). These intentions can be stored in memory, but they must be recalled at the right time. In forming future-directed intentions, planning agents typically coordinate between many different activities simultaneously. This means that, occasionally, one gets absorbed in a particular activity and thereby fails to take note of considerations relevant to a different plan. Other times, a previously formed intention fails to be recalled from memory at the right time. These situations arise from our ability to plan—in forming intentions and storing them in memory, we open ourselves to the possibility of neglecting some parts of our plans in pursuing other parts.[[10]](#footnote-10) While intentional agents can sometimes fail to achieve their goals, only planning agents can fail in the distinctive way that amounts to negligence.

We can fail in different ways as agents, but not all failures constitute negligence. We can fail due to weakness of will, lack of skill, or bad luck. But negligence is a failure of planning agency. We can see this by considering an analogy to weakness of will, which is a failure of *vertical* integration, a misalignment between one’s judgments about what one has most reason to do and what one does. It constitutes a break between judgment and action. Negligence, on the other hand, is a failure of horizontal integration, a misalignment between prior intention and current action. As such, it constitutes a break between intention and judgment. But such breaks are only possible for agents who can form intentions but not always think occurrently about them. Consider Randy who, based on his lunch plan, should not judge that he should go to the gym because the prior plan excludes gym-relevant considerations. Randy’s negligence reflects a failure of his prior intention to shape his subsequent deliberation in relevant ways. This is just what planning agency consists in: prior intentions structuring subsequent deliberation and guiding action over time. For this reason, negligence is a failure of planning agency, not simply a failure of planning agents.

Some instances of negligence seem not to reflect our planning agency. Suppose that Randy fails to adequately clear the snow-covered sidewalk outside his house. Later, a postal worker injures herself by slipping on the ice. Randy is negligent, but this does not seem to reflect his planning agency. That is, Randy could be negligent in clearing his sidewalks even if he never considered the possibility of the postal worker slipping. But if he never considered the possibility, then he did not have any prior plans that fail to get realized.[[11]](#footnote-11)

Planning agency can extend widely because agents inhabit various social roles, thereby taking on different responsibilities. Inhabiting a role partly consists in adopting, either explicitly or implicitly, the policies associated with fulfilling the role. Because these commitments can be implicit, many different activities can reflect one’s planning agency even if one has not formed an explicit intention. For example, Randy, as a homeowner, is implicitly committed to certain policies associated with his role. This includes a policy of clearing the sidewalks, maintaining the lawn, and so on.[[12]](#footnote-12) These policies structure subsequent deliberation in the same way as explicit intentions. Thus, negligence with respect to these policies reflects a failure of planning agency even in the absence of explicit intentions. For this reason, Randy’s failing to clear the sidewalks constitutes negligence even in the absence of an explicit intention or thought regarding the postal worker.[[13]](#footnote-13)

The relationship between planning agency and commitments explains why negligence is *prima facie* wrong. Planning agency enables us to adopt a more complex and variegated set of commitments that shape deliberation and action over time. Without planning agency, we could not form the types of commitments toward which we could be negligent. Planning agency, then, grounds distinctive norms of planning and vigilance that we violate when we are negligent (Murray, In press). In this way, negligence constitutes violating the distinctive duties we have toward our commitments that we possess in virtue of being planning agents.

Hence, if something acts negligently, then it is a planning agent. Failure to exercise planning agency implies, however, the possibility of success. If, in being negligent, we fail to adequately exercise planning agency, this means that successfully exercising planning agency was possible. This is a claim about *specific* capacities. My failure to *X* right now implies, in a way that merely not doing *X* right now does not, that I could have *X*-ed right now. This means that being negligent indicates that we can sometimes successfully exercise planning agency. If there were no possibility of successfully exercising planning agency, then a failure to follow through would not constitute negligence—it simply wasn’t within the individual’s power to bring about the planned outcome. If there is no possibility of success, then a failure is not negligent (though it might be a kind of unintentional omission).[[14]](#footnote-14)

Successfully exercising planning agency requires the capacity to commit to future actions. Planning agency consists in forming intentions for the nonimmediate future, where these intentions then structure subsequent deliberation. Having these future-directed intentions constitutes a commitment toward a future course of action. We can understand this commitment in two different ways: (1) the agent is committed to having the intention figure in practical reasoning, thereby excluding considerations related to other courses of conduct that are inconsistent with executing the intention at the relevant time, and; (2) the agent is committed to act on the prior intention at the relevant moment without redeliberating. One’s prior intending settles what one will do at the relevant moment (assuming there has been no change of heart or new evidence that triggers reconsideration of the original intention).

The fact that one can execute an intention formed earlier without redeliberation suggests that planning depends on rationally holding attitudes of self-trust. In foregoing deliberation at the moment of action, one exhibits trust in the earlier self’s appraisal of the reasons in favor of so acting. In committing to a future course of action, one also trusts the later self to follow through (or revise the commitment for the appropriate reasons). Without this forward- and backward-facing self-trust, one could not plan. This does not mean that temporally extended action would be practically impossible; instead, it would mean that one could, in beginning to engage in temporally extended action, hold attitudes that fall short of commitment. One could *hope* to act in the future or *prefer* to act, but neither of these constitute a commitment to action. This is because in so committing, my intention shapes subsequent deliberation. My intending to act can, in virtue of this, serve as a fixed point in deliberating about what to do, while hoping or preferring to act cannot serve as *fixed points*. If one lacks forward-facing trust, then in what way could one take this future activity as a fixed point in deliberation? Thus, the possibility of planning requires backward- and forward-looking self-trust.

When one is negligent, it provides higher-order evidence that one is untrustworthy in following through on commitments. The evidence is higher-order because it bears on the kind of person one is, rather than the material circumstances of particular plans. The evidence also serves as an inferential basis for forming beliefs about one’s reliability in following through on any plan, as most plans are structurally similar.[[15]](#footnote-15) Thus, instances of negligence provide evidence as to one’s unreliability generally. Negligence erodes self-trust by undermining general reasons for trusting oneself.

The capacity for planning depends on rationally holding attitudes of self-trust. Over time, one could no longer exercise planning agency because the self-trust needed to form commitments is defeated by this higher-order evidence. It is not just practically irrational, but practically *impossible* to form commitments in the face of undefeated higher-order evidence about one’s unreliability. Self-trust is *constitutive* of one’s capacity to make commitments. Without such self-trust, there is no way to exercise planning agency. And, given the ubiquity of negligence for planning agents like us, it seems we are all doomed to a life of self-suspicion.

In response to this, individuals can take responsibility for their negligent wrongdoing as a way of defeating higher-order evidence provided by negligence. Taking responsibility is here understood as the self-directed analogue of holding another responsible. On a broadly Strawsonian account of responsibility, for example, an individual who takes responsibility for wrongdoing is one who holds some culpability-imputing reactive attitudes that takes the individual as its object.

In taking responsibility, negligent individuals signal something important. People signal that the neglected commitments are important, and that the negligence is not due to insufficient concern for such commitments. They also signal that the victims (if there are any) deserved better. Taking responsibility also signals that the individual expected better of themselves. These signals resulting from taking responsibility are importantly related to negligence. In signaling an expectation that one should have done better one implicitly communicates that one *could* have done better in the past and *should* have done better. This is what the expectation of doing better amounts to. In signaling the importance of the neglected commitments, one indicates that one *will* do better in the future.

These signals jointly constitute an undercutting defeater for the belief in one’s unreliability. In taking responsibility, the negligent wrongdoer provides a reason for relinquishing the belief that the wrongdoer is unreliable. This does not remove the evidence; rather, it undermines an entitlement to infer unreliability from past negligence. Beyond removing this entitlement, taking responsibility also provides a reason both to oneself and others to willfully ignore the aspects of one’s agency that gave rise to the negligence. This is part of a general pattern of strategic ignorance that is essential to socially-embedded planning agency (Murray and Finocchiaro, 2021). Taking responsibility for negligence, then, is another form of confabulation that rationalizes willful ignorance. In taking responsibility, we are inviting ourselves and others to (reasonably) ignore the inherently fallible aspects of ourselves.

But why is anyone motivated to dismiss evidence of their own unreliability? Recall that, as one acquires higher-order evidence about the unreliability of one’s agency, one can no longer rationally hold attitudes of self-trust, which attitudes are necessary for planning. The motivation to take responsibility, then, stems from the more fundamental motivation to be a planning agent. We have overriding reasons to preserve our planning agency grounded in two inter-related kinds of goods. One kind of good concerns living well with others. Some have argued that shared intentionality is the basis of communal living (Tomasello, 2016). Insofar as shared intentionality requires basic forms of planning agency (Bratman, 2014), our reasons to be planning agents are parasitic on our reasons to live communally. Additionally, one might think that much of our shared life with others presumes the ability to plan—to coordinate future behavior as a way of sharing goals. Our ability to participate in and enjoy the goods of shared living would be practically impossible without planning. The other kind of good concerns individual aspirations. The kinds of activities and values that we think constitute a meaningful life require planning agency. Getting married, raising kids, earning a degree, writing a book, planting a vineyard, and so on each (plausibly) require planning agency. But to accomplish these things within the scope of one’s life certainly requires a sophisticated degree of planning. Without planning, our highest aspirations would be reduced to whatever we could hold in mind for as long as we could hold it there.[[16]](#footnote-16)

From this, we can explain why people accept accountability for negligence. We must take responsibility as a way of defeating higher-order evidence about our unreliability. Insofar as our reasons to preserve self-trust rationally compel taking responsibility, any principle that underwrites attributions of blame for negligent wrongdoing cannot be reasonably rejected simply because it sanctions people to hold one another responsible for negligence. In §2, we considered that some people might reject participating in a system of responsibility that licenses attributing blame for negligent wrongdoing precisely because we only have reasons to participate in a system that attributes blame for intentional wrongdoing. But we now see why that initial characterization is mistaken. The principles sanctioning blame for negligent wrongdoing reflect reasons we have to take responsibility for our negligence.

The appropriateness of first-personal blame grounds the aptness of second-personal blame. Every planning agent has a defeasible reason to accept accountability for negligence to preserve self-trust. Because of this, no planning agent can reasonably reject principles that map culpability to instances of negligent behavior. This means that the Control Principle, as one such principle that maps culpability to tokens of negligence, can be a basis for appropriate moral criticism, including second-personal blame. In this way, the contractualist dimension of the account grounds the permissibility of second-personal blame for negligence in the rational requirements of planning agents to take responsibility for their negligence. If nobody can reasonably reject the principle, then nobody can object to others who use it as a basis for second-personal blame.[[17]](#footnote-17)

But why, it might be asked, must we hold ourselves *accountable*? Couldn’t we commit to holding wrongdoers responsible in the attributability sense (where we express disappointment at the unreliability of others) without getting angry? However, the account presented here need not commit to a substantive claim about the moral emotions implicated in blaming the negligent; instead, a purely formal claim suffices: whatever emotions are implicated in blaming the negligent must be those that are apt for defeating higher-order evidence by signaling the relevant facts about the negligent agent (e.g., commitment to norms, commitment to victims, commitment to improve, etc.). I think the emotions that are apt to play this role—at least for individuals in Western, educated, industrialized, rich, and democratic societies—are the emotions implicated in accountability, namely anger and resentment (Nichols, 2007; Shoemaker, 2018). The form of responsibility, at least with respect to negligence, is calibrated to the vehicles best suited to restore the sense of reliability in ourselves and others.

**IV**

According to the Control Principle, individuals are culpable for negligent wrongdoing (when they are) because they should have avoided being negligent and could have avoided being negligent. In §2, three problems were raised about the moral justification of the Control Principle:

1. There is no unique motivational basis for avoiding negligence.
2. There is no way to rationally deliberate such that one avoids being negligent.
3. Because there is no way to rationally deliberate such that one avoids being negligent, it is not up to us whether we are negligent or not.

The self-trust account provides a way to address each problem. The first two claims concern the role of the Control Principle in deliberation. The underlying idea is that the Control Principle does not frame rational deliberation about *what to do* and cannot, for that reason, provide a fair basis for attributing blame for negligence. But, having situated the occurrence of negligence within the psychology of planning agency, we can see that this inference is mistaken. The Control Principle structures *retrospective* practical deliberation rather than *prospective*. The Control Principle is useful not for deliberating about *what to do*, but for deliberating about *how to respond* given what has been done. This means that the Control Principle does not provide distinctive motivation for avoiding negligence; instead, it provides distinctive motivation for taking responsibility when we are culpable for negligence. Likewise, the Control Principle does not figure in deliberation about what to do. It figures in deliberation about how to respond to negligence when it occurs. So construed, the Control Principle can figure in a co-deliberative morality because our planning agency rationally requires taking responsibility for negligence.

Finally, it did not seem to be “up to us” whether or not we are negligent. This conclusion was based partly on the lack of rational deliberation in determining whether to act in ways that eliminate the risk of negligence. But more generally the conclusion seemed to follow from the lack of action-theoretic anchors for negligence: when someone is negligent, they do not intend to be negligent or desire to be negligent or know that they are being negligent. So, in what way does negligence reflect agency? As we have seen, negligence is a manifestation of planning agency. What makes negligence “up to us” is that it stems from our planning agency. However, this does not imply that the occurrence of negligence is somehow within our deliberative reach. Thus, the argument presented here has the curious implication that what is “up to us” extends beyond the scope of deliberation. The claim is not that negligence is something we could—but often do not—deliberate about (like the pace of one’s breathing); rather, negligence is something we cannot avoid simply through sufficient deliberation.

The fact that negligence, despite being deliberatively unassailable, is up to us reflects the unique role of self-trust in planning agency. When we commit to future courses of action, we implicitly trust ourselves to carry out these actions (and, at the moment of action, trust our earlier selves in their assessment of reasons). But we are fallible agents with astonishingly limited capacities for memory, attention, and control. These limitations become more apparent when we consider the volume and complexity of our plans. This fallibility makes negligence inevitable. There is no way to eliminate fallibility, so part of maintaining self-trust is accepting this fallibility as a manifestation of our agency. The implication is that trusting yourself demands a certain degree of self-acceptance. Negligence, then, is “up to us” because we have reason to attribute negligence to ourselves.[[18]](#footnote-18)

With this, we can defuse the practical neuroticism argument from §2. There, the problem was that no matter how one deliberated, one’s plans would always leave open the possibility of negligence, meaning that deliberation could never be rationally terminated. The argument works only because we presume the deliberator lacks self-trust. However, as noted before, we can rationally settle on some plans despite the considered possibility of negligence because the possibility of negligence is too remote and can, for that reason, be dismissed. We can see now why this dismissal is rationally permitted. In trusting ourselves, we are entitled to consider certain events as too implausible to provide reasons for re-deliberation. Thus, what makes some considerations irrelevant to re-deliberation is that deliberating on the basis of them would betray a lack of self-trust. This lack of self-trust is what makes constant re-deliberation seem practically neurotic in the first place.

Joseph Raz (2011) articulates an account of responsibility where individuals are responsible for conduct that arises from manifestations of their rational agency (p. 243). For negligence, we are responsible for those activities that fall within a domain of secure competence, which consists of a range of activities that are connected to one’s sense of identity. As Raz explains it, the domain of secure competence defines the range of one’s responsibility and the rationale behind our accepting responsibility for occurrences within this domain:

…central to our way of Being in the World is a permanently evolving sense of our own mastery and its limitations. Our sense of ourselves includes awareness of a domain within which we are confident that, barring competence-defeating events (a seizure, a biased teacher, etc.), if we set ourselves to do something we will. I call it the domain of secure competence. A crucial feature of that domain is that acts within it are ones that we are entitled to undertake without reflecting on the prospects of successfully performing them…Our sense of who we are, which underpins our self-esteem, as well as our inclination to take or avoid risks and therefore our aspirations and ambitions, is tied up with our success in establishing a domain within which our powers of rational agency are securely reliable (pp. 244-45).

Later, Raz adds that: “We assert our competence by taking responsibility for [failures to exercise adequate control]” (p. 245). For Raz, one’s sense of self is the basis for taking responsibility, while the self-trust account holds that the structure of planning agency rationally requires taking responsibility. Raz appears to concede that we can sometimes eschew responsibility for negligence by relinquishing a particular claim to one’s domain of secure competence (see p. 245). The self-trust account does not allow for this: insofar as negligence manifests the general fallibility of planning agency, taking responsibility is rationally required to maintain the self-trust necessary for planning.

Amaya and Doris (2015) discuss cases of moral performance mistakes. They argue that individuals are sometimes responsible for the consequences of these mistakes even though there is no positive moral defect that gives rise to the mistake. Insofar as traditional theories of responsibility rely on the presence of a positive defect to anchor attributions of responsibility, Amaya and Doris propose an alternative account of responsibility, what they call ‘accountability without answerability’:

Even though [accountability without answerability] is not dependent on agentic appraisal, it goes further than a mere negative assessment of behavior. It reflects the fact that competent moral agents are bound by reasonable expectations; that, in the absence of excusing conditions, violations of these expectations are disapproved; and that the disapproval is fittingly expressed in attitudes of blame and guilt (p. 269).

Consider their example of helping behavior. Researchers found that individuals exposed to modest degrees of white noise were less likely to help than individuals in a control condition (72% compared to 36%; Matthews and Canon, 1975). Suppose this is an instance of negligent wrongdoing. It seems not to be a failure of planning agency, and thus the need to preserve self-trust does not provide sufficient grounds for attributing responsibility for all negligent wrongdoing. But this example still might call into question our self-trust. It indicates an inherent fragility to our capacities. We sometimes execute our intentions poorly or act on our commitments half-heartedly. When this amounts to negligence, it calls into question our capacity to follow through on our plans and commitments. And that implies the same lack of reliability that our taking responsibility is meant to undercut.[[19]](#footnote-19)

Finally, Elinor Mason has recently developed a sophisticated account of the conditions under which people might take responsibility for inadvertent wrongdoing, such as negligence or unconscious bias. For Mason, blame for negligence is sometimes appropriate because we are invested in personal relationships. When negligence hurts someone else, taking responsibility exhibits an investment in securing their trust and continuing the relationship. Taking responsibility for negligence, then, is required by our respect for others (Mason, 2018: p. 194). While I agree with many aspects of Mason’s account, I disagree with making personal relationships the fundamental basis for taking responsibility for negligent wrongdoing. It seems possible to commit “self-contained” negligence. Suppose I commit to volunteering on the weekends cooking breakfast at the homeless shelter downtown. However, Saturday morning rolls around and it slips my mind. I am required to take responsibility for failing myself in my commitment, but this is not because I have violated the duties of a personal relationship. The shelter was not counting on my being there and I have no track record of showing up on the weekends. If there is no relationship, but I can still be required to take responsibility, then the duties of personal relationships cannot be the basis on which I take responsibility. Second, navigating the demands of personal relationships often requires exercises of planning agency, in part because planning facilitates forms of interpersonal coordination that are essential to sustaining personal relationships. However, the demands of planning agency seem more fundamental than the demands of interpersonal relationships because the latter asymmetrically depends on the former: there could be no interpersonal relationships without planning agency, but there could be planning agents in a world without interpersonal relationships. Accordingly, the demands of interpersonal relationships are optional in a way that the demands of planning agency are not.[[20]](#footnote-20)

**V**

The self-trust account explains why the Control Principle is a mutually acceptable basis for holding negligent wrongdoers accountable. When we properly situate the occurrence of negligence within the psychology of planning agency, we can extract a contractualist foundation for the Control Principle. This account explains why taking responsibility for negligence is rationally required, but it also explains individual motivation to take responsibility. It also offers a novel account of the role of taking responsibility in the life of a planning agent: taking responsibility is a tool that we use to compensate for the inherent fragility of our agency. Saddled with inescapable fallibility, the next best thing is to own our failures and trust ourselves.

**References**

Amaya, S. 2022. Negligence: Its moral significance. In M. Vargas and J. Doris (eds.) *The Oxford Handbook of Moral Psychology* (Oxford: Oxford University Press).

Amaya, S. and Doris, J. 2015. No excuses: Performance mistakes in morality. In J. Clausen and N. Levy (eds.) *Handbook of Neuroethics* (Dordrecht: Springer), 253-72.

Amaya, S. and Murray, S. “Vigilance.”

Ayars, A.A. 2021. Blaming for unreasonableness. *Journal of Ethics and Social Philosophy* 19:1, 56.

Bernecker, S. 2018. On the blameworthiness of forgetting. In D. Debus and K. Michaelin (eds.) *New Directions in the Philosophy of Memory* (London: Routledge), 241-58.

Bratman, M. 1987. *Intention, plans, and practical reason*. Stanford: CSLI Publishing.

Bratman, M. 2014. *Shared agency*. Oxford: Oxford University Press.

Brink, D.O. 2021. *Fair opportunity and responsibility*. Oxford: Oxford University Press.

Clarke, R. 2014. *Omissions: Responsibility, Agency, and Metaphysics*. Oxford: Oxford University Press.

Clarke, R. 2017. Ignorance, revision, and commonsense. n P. Robichaud and J.W. Wieland (eds.) *Responsibility: The Epistemic Condition* (Oxford: Oxford University Press), 233-51.

De Brigard, F. and Brady, W.J. 2013. The effect of what we think may happen on our judgments of responsibility. *Review of Philosophy and Psychology* 4:2, 259-69.

Doris, J. 2015. *Talking to Our Selves*. Oxford: Oxford University Press.

Graham, P.A. 2014. A sketch of a theory of moral blameworthiness. *Philosophy and Phenomenological Research* 88:2, 388-409.

Irving, I., Murray, S., Krasich, K., Glasser, A. 2023. The Catch-22 of forgetfulness: Responsibility for mental mistakes. *Australasian Journal of Philosophy*.

Jefferson, A. 2019. Instrumentalism about moral responsibility revisited. *Philosophical Quarterly* 69:276, 555-73.

Kissinger-Knox, A., Aragon, P., and Mizrahi, M. 2018. Does non-moral ignorance exculpate? Situational awareness and attributions of blame and forgiveness. *Acta Analytica* 33:2, 161-79.

Lenman, J. 2006. Compatibilism and contractualism. *Ethics* 117:1, 7-31.

Levy, N. 2009. Culpable ignorance and moral responsibility. *Ethics* 119, 729-41.

Levy, N. 2017. Methodological conservatism and the epistemic condition. In *Responsibility: The Epistemic Condition*, 252-65.

Malle, B., Guglielmo, S., and Monroe, A.E. 2014. A theory of blame. *Psychological Inquiry* 25, 1-40.

Mason, E. 2018. *Ways to be blameworthy*. Oxford: Oxford University Press.

Murray, S. 2017. Responsibility and vigilance. *Philosophical Studies* 174:2, 507-27.

Murray, S. In press. The nature and norms of vigilance. *American Philosophical Quarterly*.

Murray, S. and Finocchiaro, P. 2021. These confabulations are guaranteed to improve your marriage! *Synthese* 198, 10313-339.

Murray, S., Krasich, K., Irving, Z., Nadelhoffer, T., and De Brigard, F. 2023. Mental control and attributions of blame for negligent wrongdoing. *Journal of Experimental Psychology: General* 152:1, 120 - 38.

Murray, S., Murray, E., Stewart, G., Sinnott-Armstrong, W., and De Brigard, F. 2019. Responsibility for forgetting. *Philosophical Studies* 176, 1177-1201.

Murray, S. and Vargas, M. 2020 Vigilance and control. *Philosophical Studies* 177:3, 825-43.

Nadelhoffer, T., Heshmati, S., Kaplan, D., and Nichols, S. 2013. Folk retributivism and the communication confound. *Economics & Philosophy* 29:2, 235-61.

Nichols, S. 2007. After incompatibilism. *Philosophical Perspectives* 21:1, 405-28.

Pettit, P. and McGeer, V. 2015. The hard problem of responsibility. In D. Shoemaker (ed.) *Oxford Studies in Agency and Responsibility*, vol. 3 (Oxford: Oxford University Press), 160-88.

Raz, J. 2011. *From normativity to responsibility* (Oxford: Oxford University Press).

Rosen, G. 2004. Skepticism about moral responsibility. *Philosophical Perspectives* 18:1, 295-313.

Rudy-Hiller, F. 2019. Give people a break: Slips and moral responsibility. *Philosophical Quarterly* 69:277, 721-40.

Rudy-Hiller, F. 2022. A capacitarian account of culpability for negligence. *Manuscrito* 45:2, 118-160.

Scanlon, T.M. 1998. *What we owe to each other* (Cambridge, MA: Belknap Press).

Scanlon, T.M. 2008. *Moral dimensions* (Cambridge, MA: Belknap Press).

Sher, G. 2006. *In praise of blame* (Oxford University Press).

Shoemaker, D. 2011. Attributability, answerability, and accountability. *Ethics* 121, 602-32.

Shoemaker, D. 2015. *Responsibility from the margins* (Oxford University Press).

Shoemaker, D. 2018. You ought know: Defending angry blame. In M. Cherry and O. Flanagan (eds.) *The moral psychology of anger* (Lanham, MD: Rowman and Littlefield), 67-88.

Smith, A.M. 2005. Responsibility for attitudes: Activity and passivity in mental life. *Ethics* 115:2, 236-71.

Smith, H.M. 1988. Making moral decisions. *Noûs* 22, 89-108.

Sripada, C. 2015. Moral responsibility, reasons, and the self. In D. Shoemaker (ed.) *Oxford Studies in Agency and Responsibility*, vol. 3, 242-64.

Sripada, C. 2016. Self-expression: a deep self theory of moral responsibility. *Philosophical Studies* 173, 1203-32.

Sripada, C. 2019. The fallibility paradox. *Social Philosophy and Policy* 36:1, 234-48.

Talbert, M. 2013. Unwitting wrongdoers and the role of moral disagreement in blame. In D. Shoemaker (ed.) *Oxford Studies in Agency and Responsibility*, vol. 1 (Oxford: Oxford University Press), 225-45.

Talbert, M. 2017. Akrasia, Awareness, and Blameworthiness. In *Responsibility: The Epistemic Condition*, 47-63.

Tomasello, M. 2016. *A Natural History of Human Morality* (Cambridge, MA: Harvard University Press).

Vargas, M. 2013. *Building better beings*. Oxford: Oxford University Press.

Vargas, M. 2020. Negligence and social self-governance. In A. Mele (ed.) *Surrounding self-control* (Oxford: Oxford University Press), 400-420.

Vargas, M. 2021. Constitutive instrumentalism and the fragility of responsibility. *The Monist* 104:4, 427-42.

1. People sometimes blame the habitually non-negligent *more* for negligent wrongdoing than the habitually negligent (Irving et al., 2023). [↑](#footnote-ref-1)
2. My target is whether and why people are ever accountable for negligent wrongdoing. An individual is accountable for some wrongdoing only when their behavior makes resentment, indignation, or moral anger appropriate (Shoemaker, 2015). I presume that accountability is non-trivially related to both culpability-imputing judgments and instantiating some Strawsonian reactive attitudes, although my account could be adapted to fit non-emotive (Scanlon, 2008) or hybrid accounts (Sher, 2006) of blame. Hereafter, the terms ‘responsibility’ and ‘culpability’ will be treated as synonymous with ‘accountability’. [↑](#footnote-ref-2)
3. A moral principle is a function that maps moral status (wrongness, permissibility, culpability) to act-tokens or some properties of those tokens. For example: Necessarily, if some act is an *injuring*, it is *prima facie* wrong. The Control Principle, then, is a package of the following principles:

   Necessarily, if some act is negligent, then the act is *prima facie* wrong.

   Necessarily, if the direct consequences of a wrong act are preventable, then failing to prevent those consequences is *prima facie* culpable. [↑](#footnote-ref-3)
4. Notable exceptions are Raz (2011), Amaya and Doris (2015), Mason (2018), and Rudy-Hiller (2022). [↑](#footnote-ref-4)
5. Thus, genuinely moral principles function as both evaluative standards and practical guides (Smith, 1988). [↑](#footnote-ref-5)
6. The implication is that appropriate moral criticism presumes a shared understanding between criticizer and criticized of fair principles underlying the criticism. Additional considerations bear on the appropriateness of moral criticism, such as whether the criticized could, in principle, answer why she violated, rather than complied with, the norm (Shoemaker, 2011). This “answerability” condition on appropriate criticism implies that criticism cannot be justified simply in terms of the principles underlying the criticism. [↑](#footnote-ref-6)
7. This leaves open the possibility that the fairness of some principle alone might not exhaust how it shapes deliberation and action. [↑](#footnote-ref-7)
8. Thanks to Walter Sinnott-Armstrong for discussion of the practical neuroticism argument. [↑](#footnote-ref-8)
9. Other versions of the Deep Self Principle focus on *attributability* (Smith, 2005; Graham, 2014). On this view, negligent wrongdoing might merit feelings of disdain or disappointment that prompt adjusting one’s relationship independently of whether one is accountable for the negligence. However, these views fail to capture our ordinary tendencies of feeling anger toward negligent wrongdoers, even when such wrongdoing manifests nothing about one’s deep self (see Irving et al., In press). [↑](#footnote-ref-9)
10. See Amaya & Murray (Under review). [↑](#footnote-ref-10)
11. Thanks to David Shoemaker for providing the objection and the example. [↑](#footnote-ref-11)
12. Some policies might arise only from the interactions between *n-*adic combinations of different roles. As a simple example, the roles of *being a homeowner* and *being a resident of Rhode Island* might combine to produce a policy of clearing the sidewalk of ice, whereas the roles of *being a homeowner* and *being a resident of Florida* do not. [↑](#footnote-ref-12)
13. The example suggests another way negligence reflects a failure of planning agency. Randy is negligent not because he fails to implement a policy associated with his role, but because he fails to adequately execute the corresponding intention. Sometimes negligence consists of *poorly executed intentions* rather than just *non-executed intentions*. And it seems possible for non-planning agents to execute their intentions poorly. This is discussed in §4. [↑](#footnote-ref-13)
14. This, in turn, reinforces the assumption (stated earlier) that all negligence is preventable. [↑](#footnote-ref-14)
15. This evidence can also ground second-personal judgments. Insofar as negligence is publicly observable, the same higher-order evidence is available to others. [↑](#footnote-ref-15)
16. Rudy-Hiller (2022, pp. 137-38) notes a third possible good, namely a unified conception of oneself (or wholeheartedness). In failing to take responsibility, one risks being alienated from parts of one’s agency (i.e., those fallible aspects of oneself that give rise to negligence). Being wholehearted seems to be a constituent of both forward- and backward-looking self-trust. [↑](#footnote-ref-16)
17. Does the present account presume that contractualism is the correct normative theory? I do not think so. The distinctively contractualist elements of the account could be substituted for analogues in other deontological theories without undermining the view presented here. For example, one difference between a straightforward deontological theory and contractualism is whether fair principles are just those every rational agent would accept as binding or that no rational agent would reasonably reject as binding. Given our special reasons to be planning agents, it is plausible that every rational agent would accept the Control Principle as articulated here. Thus, as far as I can tell, nothing about the present account depends on a specifically contractualist formulation of deontological normative theory. [↑](#footnote-ref-17)
18. This answers a common objection that memory failures are “glitches” that cannot ground blame for negligent wrongdoing (Rosen, 2004; Levy, 2009; Talbert, 2013). Ayars (2021), for example, claims that remembering “does not seem to involve a kind of rational activity” (p. 75). Remembering does not result from deliberation. But the rational activity underlying remembering is trusting oneself to remember. This self-trust is the basis for attributing memory failures to agents (absent excusing conditions). The self-trust account, then, widens the attributability base in a way that makes memory failures attributable to agents. Hence, memory failures are not mere glitches. [↑](#footnote-ref-18)
19. Thus, instances of negligence that consist in poorly executed intentions can also reflect our planning agency insofar as they can undermine one’s confidence in their resilience over time. [↑](#footnote-ref-19)
20. This is not intended as a decisive objection. Rather, I take it as suggestive that it is unclear how planning agency and socially embedded agency are related. I consider the former to be fundamental, but it is possible that the two are inextricably intertwined. Nevertheless, I agree with Mason that interpersonal relationships augment the ways in which we can be negligent and shape the modes of expression that appropriate blame for negligence can take. [↑](#footnote-ref-20)