

Divided Agency, Manipulation, and Regret

Jonathan D Payton

1. DIVIDED AGENCY?

Bazargan-Forward conceives of agency as consisting of two functions: a deliberative and an executive function. The job of the deliberative function is to decide what to do. While Bazargan-Forward is neutral on many of the details of how deliberation works, here is a sketch which I find attractive and which seems to fit much of what he says. When I deliberate about whether to φ , I consider (what I take to be) my reasons for and against φ -ing, with an eye towards acting as my reasons dictate. Reasons come in two kinds: *first-order* and *second-order*. First-order reasons are considerations which count directly for or against φ -ing. Second-order reasons concern my reasons for or against φ -ing; they're considerations of the strength and overall balance of my reasons for and against φ . Deliberation ends when I answer for myself the question, "Shall I φ ?" on the basis of all these reasons.

The job of the executive function is to do what I've decided to do and, at least typically, to do it for the reasons that I decided to do it; the reasons which figured in my deliberation become *motivating reasons*. For p to be a motivating reason for me to φ is, roughly, for p to figure in the correct rationalizing explanation of my φ -ing; I φ *because* p , in the special sense of "because" which

10.25365/jso-2024-9147

Published online November 30, 2024

Jonathan D Payton, Bilkent University, E-mail: jonathanpayton@bilkent.edu.tr

Open Access. © 2024 Author(s) published by the Journal of Social Ontology. This work is licensed under the Creative Commons Attribution 4.0 International License.

is operative in the explanation of action.¹

Bazargan-Forward's driving idea is that these two functions can be divided across multiple agents: the deliberator and the executor needn't be the same person.

It is possible [...] for *you* to “outsource” the executory functions to *me*. In such a case, you attribute to me the role of enacting the practical reasons you take there to be (Bazargan-Forward 2022, 24, emphasis in the original).²

In such a case you attribute to me the active function of enacting the practical reasons you take there to be pertaining to φ [...] [M]y conduct has the function of acting on the practical reasons you take there to be (39).

The result is an “interpersonal division of agential labour”: one person does the deliberating while another “enacts” (23, 24) or “acts on” (39) the reasons the first person takes there to be.

Bazargan-Forward takes this idea to have important moral consequences. For Bazargan-Forward, an agent's motivating reasons are “deontically relevant” to the moral evaluation of her action (38, 60). That is, when asking whether she acted permissibly, we must consider the set of her motivating reasons, P , and ask whether it's permissible to φ -because- P . However, he's willing to entertain the idea that motivating reasons are merely “aretaically relevant,” that is, relevant to our evaluation of the agent as a person (64). Either way, in a case of divided agency the deliberator's reasons have an important role to play in our evaluation of the executor and/or her behaviour: “we must repair to the deliberator's motivating reasons in morally evaluating the executor's conduct” (37). Consider the following example:

Performance Review. Manager is ordered by Owner to promote Employee, who happens to be a man. This act is permissible in itself: by any fair set of promotion criteria, Employee ought to be promoted. Manager follows the order, and promotes Employee with no ill intentions. However, unbeknownst to Manager, Owner is a sexist who

¹ Here, I'm drawing on a way of thinking about action and motivating reasons which traces back at least to Anscombe (1963) and Davidson (1963); for recent criticism, see O'Brien (2021, 2023).

² Reference to Bazargan-Forward (2022) will be shortened to page number only throughout this paper.

wants to ensure that men are promoted before women, and issues the order with that intention.³

Clearly, Owner's sexist motivations show something bad about his character, and we may even want to say that he has acted impermissibly in issuing the promotion order *because* doing so promotes his sexist agenda. But, Bazargan-Forward thinks, Owner's sexist motivating reasons affect the moral quality of *Manager's* behaviour, not just his own. Owner's motivations are deontically, or at least aretaically, relevant to our assessment of Manager's act of promoting employee.

2. SOME REASONS FOR SKEPTICISM

Much of what Bazargan-Forward says suggests that, in a case of divided agency, the reasons which figured in the deliberator's deliberation become the executor's motivating reasons. Intuitively, to "enact" or "act on" a reason is to for it to be among one's motivating reasons, and so, to say that the function of "enacting" or "acting on" the deliberator's reasons has been outsourced suggests that the executor φ -s *for* those reasons.

Moreover, this idea would seem to make clear sense of Bazargan-Forward's claim that the deliberator's reasons can be deontically or aretaically relevant to the executor's behaviour. If the executor literally acts on the deliberator's reasons, then those reasons are also the *executor's* reasons, and so there's no mystery how they could be deontically or aretaically relevant to her behaviour.

Finally, the idea is strongly suggested by the analogy Bazargan-Forward draws between divided agency, on one hand, and *temporally extended agency*, on the other. Often, I don't wait until the time of action to do my deliberating. Instead, I do my deliberating in advance, effectively extending my agency across time: I exercise the deliberative function at an earlier time and the executive function at a later time. Drawing on Ferrero (2010), Bazargan-Forward claims that, at least under certain conditions, the fact that I previously decided that I would now φ acts as a reason that I have *now*, to φ . But, he thinks, I don't φ *merely* for the reason that I already decided I would φ . Rather, the fact that I previously decided that I would φ acts as a kind of device of "anaphoric reference" back to the reasons for which I originally decided to φ : when I act on my previous decision, I'm "ultimately acting out of the original and unadulterated assessment of the merits of the case" (Ferrero 2010, 13; see also

³ The example is Bazargan-Forward's (46), but the phrasing is mine.

Bazargan-Forward (2022, 30). Likewise, *mutatis mutandis*, for divided agency.

I don't think this is the right thing to say, either about temporally extended agency or about divided agency. Motivating reasons—i.e. the reasons we “enact” or “act upon”—are expressible as propositions which figure in the correct rationalizing explanation of our behaviour; the agent φ -s *because p*. But often, in cases of temporally extended agency, it's wrong to say that I do what I do *for* the reasons I previously had. When the time comes to act, I may have forgotten what the original reasons behind my decision were, and so I can't be said to be acting on them *now*. The correct explanation of my behaviour doesn't appeal to the reasons which motivated my original decision, but rather to the fact that I made that decision: my reason to φ now is *that I previously decided to φ* . Nor does this reason seem to act as a device of anaphoric reference back to my previous reasons, if that idea is taken strictly seriously. I might *add* a reference to my previous reasons, something like, “I previously took myself to have good reasons for φ -ing and I decided to φ , for *those reasons*.” But this isn't enough to get them to figure as motivating reasons: I can hardly say to myself, “I'm now φ -ing for the reasons for which I originally decided to φ , which I've now forgotten.”⁴

Likewise, in cases of divided agency, it seems incorrect to say that the executor is ultimately acting on the deliberator's reasons. In the case of *Performance Review*, for instance, it would simply be incorrect to say that Manager promotes Employee *because* doing so promotes Owner's sexist aims, or anything of the sort.

Now, it seems that Bazargan-Forward actually agrees with me, here. Although much of what he says *suggests* that the deliberator's reasons are among the executor's motivating reasons, he explicitly rejects this view:

[T]he reasons the deliberator takes there to be determine what the executor's protected reasons are and thus what the executor's purpose is. But this claim falls short of the claim that the deliberator's motivating reasons are the executor's motivating reasons (55).

Rather than bestowing *motivating reasons* on the executor, the deliberator bestows a *purpose* (40, 55–6). A purpose is different from an intention: in *Performance Review*, Manager certainly doesn't *intend* to promote Owner's

⁴ At one point, Ferrero seems to agree; he denies that the reasons for my earlier decision to φ figure in the *content* of my current intention to φ , suggesting that they're not among my motivating reasons *now* (Ferrero 2010, 14). But this leaves the idea that I'm “ultimately” acting on those reasons unclear.

sexist aims. Nonetheless, Bazargan-Forward thinks, that is the *purpose* of Manager's behaviour. We might think of having a purpose as something like being aimed in a direction. Typically, I'm the one aiming myself and my behaviour in certain directions: *my* intentions dictate the purpose of my behaviour. However, in a case of divided agency, it's the deliberator who aims the executor and her behaviour in particular directions, thereby bestowing a purpose on them. Thus, no matter how pure Manager's own intentions and motivating reasons might be, it remains true that the *purpose* of his promotion of Employee is to promote Owner's sexist agenda.

The problem is that the notions of deontic and aretaic relevance on which Bazargan-Forward draws, when morally evaluating the executor's behaviour, seem to be very strongly tied to the notion of motivating reasons. If Owner's reasons *aren't* among Manager's motivating reasons, it's not clear how those reasons could render his behaviour impermissible. By hypothesis the act of promoting Employee, considered apart from Manager's reasons, is permissible in itself. And by hypothesis Manager's own motivations were pure. How, then, could he have acted impermissibly? It's even less clear how purposes could be aretaically relevant, if they don't bestow motivating reasons on the executor. If Manager does something permissible for permissible reasons, how could his *character* possibly be impugned? Bazargan-Forward writes:

[W]hen Manager acts as Owner's executor, it's *Owner's* motivating reasons that constitutively determine what *Manager* is supposed to do and why. This means *the relationship Manager bears to Owner is functionally analogous to the relationship in [sic] which Manager bears to his own self when he is acting in his own capacity*. As such, Manager can no more disavow the Owner's motivating reasons than he can his own [...] [A]n executor doesn't just help the deliberator achieve her goals; in addition, the deliberator's goals functionally count as the executor's. (65, emphasis in the original)⁵

In a sense, this seems right. Having a purpose is like being aimed in a direction, and whatever Manager's own intentions might have been, there's simply no denying that his behaviour was aimed, by Owner, in a morally reprehensible direction. What's less clear to me is how this point can have the moral consequences Bazargan-Forward suggests, once it's been stripped of any suggestion that Manager is acting on morally reprehensible *intentions* or *reasons*.

5 See also (230–1).

In sum, much of what Bazargan-Forward says suggests a picture of divided agency on which the executor literally acts on the deliberator's reasons: that picture fits his picture of temporally divided agency; and it makes good sense of the moral judgments he wants to make about cases like *Performance Review*. However, that picture is problematic and doesn't seem to reflect Bazargan-Forward's considered view. But his considered view doesn't seem to support his moral judgments as easily.

3. "GET YOUR ASS TO MARS"

I'd like to suggest an alternative way of thinking about what's going on cases like *Performance Review*.

Consider a different case. In Paul Verhoeven's 1990 film, *Total Recall*, Douglas Quaid lives as a construction worker on Earth but dreams of moving to Mars. Through a series of hijinks he learns that he actually *has* been to Mars, but that his memory has been erased. He receives a video recording from his previous self, who identifies himself as "Hauser." Hauser says that he's a secret agent who has been working to bring down the mining-tycoon-*cum*-dictator Vilos Coahaagen and to support the mutant uprising on Mars. He instructs Quaid to pick up where he left off.

Unfortunately, Quaid has been tricked. Worse, he's been tricked by himself. Hauser in fact works for Coahaagen, and voluntarily had his memory erased in order to become the perfect double-agent. By returning to Mars and contacting the members of the mutant uprising, Quaid has led Coahaagen right to them.

In some ways, this is like a case of temporally extended agency: it's like what happens when I leave myself a note reminding me to do something, but completely forget what my reasons were when the time to act comes. But it's also importantly different from an ordinary case of temporally extended agency. It feels more like a case of manipulation, albeit one in which the manipulator and the manipulated are the same. That's certainly how it feels to Quaid: he feels manipulated and betrayed, and spends the rest of the film attempting to thwart both Coahaagen and his previous self.

The explanation for this feeling of manipulation, I suggest, has to do with Quaid's and Hauser's different *aims*. While I've criticized some aspects of Ferrero's (and hence Bazargan-Forward's) model of temporally extended agency, he has something important to say on this matter. Typically, when I decided what to do in advance, I do so with an eye to doing what I *would* do if I waited to deliberate in the moment. Or, if I suspect that my decision-making

will somehow be impaired when the time for action comes, I make my decision with an eye to doing what an idealized, unimpaired version of me would do, if he deliberated in the moment. In short, I'm trying to make a decision that I can, and should, stand behind when the time for action comes. What makes it rational for me to act on this previous decision is the thought that I succeeded in this aim: if it's rational for me to think, in the moment, that my previous decision was the correct one, then it's rational for me to do what I decided to, *simply for the reason that I decided to*, rather than re-open the question for further deliberation (Ferrero 2010, 8–10). Quaid certainly believes that Hauser is instructing him to do what he would (or should) decide to do anyway, if he had all the relevant information. Moreover, it seems perfectly rational for him to believe this. (Why would he lie to himself?) But this belief is false; his aims and Hauser's are out of alignment. Moreover, Hauser *knew* they would be out of alignment, and has been directing Quaid's behaviour to satisfy *his* aims, not Quaid's. Hence the feeling of manipulation.

I want to suggest that when the aims of the order-giver and the order-follower are out of alignment in this way,⁶ the order-follower *can* disavow the order-giver's aims, in the sense of "disavow" which is relevant to a moral assessment of their behaviour and/or character. There's simply no denying that Hauser pointed Quaid in a certain direction; that's simply among the facts of the case. But to "disavow" Hauser's aims, Quaid needn't deny this obvious fact. He need only refuse to take those aims as his own, and refuse to let them serve as the standard by which his behaviour is to be judged.

Recall that, in an ordinary case of temporally extended agency, I don't think it's correct to say that I "enact" or "act upon" my previously held reasons for deciding to φ . Nonetheless, Ferrero (2010, 13–4) and Bazargan-Forward (2022, 30) are correct in thinking that, when I act on my previous decision, the rationality or goodness of my behaviour can be judged with reference to those reasons. This is, I think, because although I'm not *acting on* those reasons, I'm volunteering to let those reasons serve as the standard against which my behaviour is to be judged. If I'm committed to thinking that my previous decision was a good one, then I'm also committed to thinking that the overall balance of reasons, as I then saw it, favoured φ -ing. If that turns out not to be the case, then my behaviour is impugned. But this is *only* because I committed myself to thinking that the reasons which motivated my decision were good ones, a commitment which I took on when I decided to act on my previous

⁶ Here I don't say "deliberator" and "executor," since those labels suggest Bazargan-Forward's model of divided agency.

decision rather than re-open the question for deliberation.

Similarly, I suggest, when one agent acts on another's orders, they take on a commitment to thinking that the reasons which motivated those orders were good ones, and to having their behaviour evaluated on the basis of those reasons.

You might wonder how this is supposed to help Quaid. In acting on Hauser's orders, didn't he commit himself to letting Hauser's reasons serve as the standard of evaluation for his behaviour? How can he now disavow them?

The answer I want to suggest is that Quaid's commitment wasn't unqualified. He didn't commit himself to thinking that Hauser's reasons, *whatever they were*, were good ones, and so he didn't commit himself to letting Hauser's reasons, *whatever they were*, serve as the standard against which his behaviour could be judged. Of course, he probably didn't have any clear exceptions in mind; he didn't need to explicitly think to himself, "... Well, unless Hauser is really trying to bring down the mutant uprising." But exceptions arguably don't need to be made explicit in order to be in force. If a friend asks me for a favour, and I reply, "I'll do anything you want," I don't mean *anything*; helping them to cover up a murder is probably out. Similarly, Quaid can rightly claim not to have committed himself to Hauser's aims, and so he can disavow them.

Now return to *Performance Review*. It may be that Manager really didn't give himself an escape clause, and really did consent to being pointed in *whatever* direction Owner chose. Sometimes, perhaps, people really do give their superiors *cart blanche* in this way. But that's already a morally problematic decision which reflects poorly on Manager's character. In any case, it's much more likely that Manager's commitment to letting Owner point him in a certain direction, and hence to thinking that Manager's reasons are good ones, isn't unqualified; exceptions, even if not made explicit, are in force. Manager can rightly say that it didn't occur to him that Owner might have sexist motives, and that he never intended to let such motives serve as the standard by which his behaviour could be judged. In that sense, he can disavow Owner's aims.

4. REGRET

Does this mean that Manager can breathe easy? Not necessarily; certainly, if I were in Manager's position, and discovered that my act of promoting Employee (which was perfectly permissible in itself) had been ordered for sexist reasons, I'd have trouble sleeping at night. But perhaps what I'd be feeling isn't best conceptualized as guilt over having done something impermissible

or having acted out of bad character, but as what Williams (1981) calls “agent-regret.” Intuitively, I ought to feel something like regret, a deep sadness about the role I played in Owner’s machinations; after all, there’d simply be no denying that it was *I* who promoted Employee. It would be inappropriate for me *not* to feel this way. But it would be equally inappropriate for me to blame myself or feel guilty, as though I’d acted impermissibly or displayed a corrupt moral character. After all, I acted in good faith, and Employee *ought* to have been promoted. Regret is permissible here, but guilt and blame aren’t.⁷

It’s a cliché that there are no knockdown arguments in philosophy. Nor have I attempted to provide any. Rather, what I’ve attempted to do is raise some concerns about Bazargan-Forward’s model of divided agency and his application of it to derive moral consequences, and to sketch an alternative picture.

REFERENCES

- Anscombe, G. E. M. 1963. *Intention*. Cambridge, MA: Harvard University Press, 2 edition.
- Bazargan-Forward, S. 2022. *Authority, Cooperation, and Accountability*. Oxford: Oxford University Press, URL <https://doi.org/10.1093/oso/9780192862419.001.0001>.
- Bittner, R. 1992. “Is it Reasonable to Regret Things One Did?” *Journal of Philosophy* 89 (5): 262–273, URL <https://doi.org/10.2307/2027168>.
- Davidson, D. 1963. “Actions, Reasons, and Causes.” *The Journal of Philosophy* 60 (23): 685–700, URL <https://doi.org/10.2307/2023177>.
- Enoch, D. 2012. “Being Responsible, Taking Responsibility, and Penumbral Agency.” *Luck, Value, and Commitment: Themes from the Ethics of Bernard Williams*, edited by U. Huerer, and G. Lang, Oxford: Oxford University Press, 95–133, URL <https://doi.org/10.1093/acprof:oso/9780199599325.003.0005>.
- Ferrero, L. 2010. “Decisions, Diachronic Autonomy and the Division of Deliberative Labour.” *Philosophers’ Imprint* 10 (2): 1–23, URL <http://hdl.handle.net/2027/spo.3521354.0010.002>.
- O’Brien, L. 2021. “Answerability without Reasons.” *Oxford Studies in Agency and Responsibility. Volume 7*, edited by D. Shoemaker, Oxford: Oxford University Press, 32–53, URL <https://doi.org/10.1093/oso/9780192844644.003.0003>.

⁷ For discussion of Williams’ notion of agent-regret see e.g. Bittner (1992); Wolf (2004); Enoch (2012); Sussman (2018).

- O'Brien, L. 2023. "Actions and Questions." *Analysis* 83 (2): 260–268, URL <https://doi.org/10.1093/analys/anac069>.
- Sussman, D. 2018. "Is Agent-Regret Rational?" *Ethics* 128 (4): 788–808, URL <https://doi.org/10.1086/697492>.
- Williams, B. 1981. "Moral Luck." *Moral Luck*, Cambridge: Cambridge University Press, 20–39, URL <https://doi.org/10.1017/CBO9781139165860.003>.
- Wolf, S. 2004. "The Moral of Moral Luck." *Setting the Moral Compass: Essays by Women Philosophers*, edited by C. Calhoun, Oxford: Oxford University Press, 113–128, URL <https://doi.org/10.1093/oso/9780195154740.003.0007>.