Draft of 15 April 2020.
To be published in: *Handbook of Rationality*, ed. by Markus Knauff and Wolfgang Spohn, MIT Press – in preparation.

Rationality and Morality

Christoph Fehige and Ulla Wessels

Abstract

Is practical rationality on the side of morality? Is it even the benchmark or the ground of morality? Why do what morality requires you to do? The diversity of answers that are still in the running and of considerations for and against them is astonishing. Our aim is to delineate the structure of the debate and to locate and clarify some major questions, options, and moves. We organize the presentation around a pair of prominent sample views, linking the rationality of an action to the agent's desires, but its morality to the general welfare. We expound how different the matter looks for other views of rationality or of morality. All things considered, thoughts about each of the two normative domains and about conflicting norms in general suggest that even regarding well-informed agents rationality and morality cannot be fully harmonized. To some extent, convergence will remain gappy and contingent.

Rationality and Morality*

Christoph Fehige and Ulla Wessels

How do the judgments of practical rationality relate to those of morality? Let us call that question the RM question, with "R" for practical rationality and "M" for morality. The RM question is complex because each of the two relata is controversial in its own right – what *is* the rational thing to do, and what *is* the morally right thing to do? – and so is the pecking order among them: does an action have to be morally right in order to qualify as rational, or vice versa, or is there no such connection?

Judgments of the two kinds appear at variance in many cases that have existential weight. It may well happen, for example, that morality appears to require an affluent person to donate eighty per cent of her income to a charity that saves lives efficiently, whereas practical rationality appears to require her to use the same money for completing her beloved collection of abstract paintings. We would expect a satisfactory answer to the RM question to get some kind of grip on such constellations. The answer should either show that R and M are in concord after all or, in as far as they are not, what follows from the discord for theories of normativity and for thoughtful agents.

Actions and Requirements

As usual, it makes sense to restrict the inquiry. The plan is to look at rationality and morality only in as far as they concern actions. There are other items that beckon for our attention – think of the rationality and morality of beliefs, decisions, desires, emotions, intentions, maxims, or ways of life –, but we will not extend the discussion to them. One consequence of the focus on actions is that

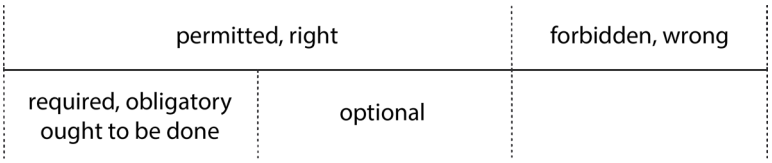

Figure 1: The main deontic terms and the logical relations among them. Terms written in the same field are roughly synonymous. Each term can occur in discourse about R and in discourse about M ("rationally permitted" vs. "morally permitted", etc.).

the rationality that pertains is *practical* rationality by definition, which enables us to omit the adjective "practical" most of the time.

We will restrict the scope further by looking at only one kind of assessment by R and M, that of actions as (rationally or morally) required, permitted, optional, or forbidden. These are known as the "deontic" assessments, and it is the "all things considered" versions of them that we will be concerned with. We present the logical relations among the deontic terms in figure 1. There is one couple of terms, not in the diagram, that we will reserve for use in the domain R: the terms "rational" and "irrational" for actions that are rationally permitted and rationally forbidden, respectively. There is another couple, in the diagram, that today we will reserve for use in the domain M: in this article, the terms "right" and "wrong" will be used for moral assessments.

Two Sample Doctrines: Instrumentalism and Utilitarianism

It will help to look at the relations between one common criterion of R and one common criterion of M, and to widen the view from there. As to the rationality of actions, many criteria that have been proposed are variations of one simple tenet: that it is rational for a person to try to get what she wants. There are competing ways of refining that outlook, and here is the version that will serve as our sample of a criterion: *An action that the agent can perform is rational if and*

only if the agent believes that no other action that she can perform brings about more fulfilment of her intrinsic desires.

The performing, believing, and desiring in the criterion should all be understood as happening or obtaining at the same moment, but the believing and desiring don't have to occur in the agent's consciousness; they may be purely implicit. An intrinsic desire should be understood as a desire of something for its own sake, not of it as a means to something else; henceforth in this article, whenever we write "desire", we will mean "intrinsic desire". The quantitative notion of fulfilment takes into account both the number and the strengths of desires.

We will treat the criterion as the defining feature of "instrumentalism".[1] The label is apt because the criterion codifies a view of actions as tools, assessing them with respect to their putative efficiency in achieving the agent's ends. Instrumentalism is a close relative of that theory of rational decision-making that puts the maximization of expected utility (MEU) centre-stage and plays a large role in the behavioural sciences. The MEU theorist's probabilities and amounts of utility correspond, by and large, to the instrumentalist's beliefs and amounts of desire fulfilment.

In the moral domain, our sample criterion will be utilitarian: *An action is morally right if and only if no other action that the agent can perform brings about more welfare, world-wide.* The utilitarian message is that welfare counts, no matter whose welfare it is, and that nothing else does. What is that all-important stuff called "welfare"? One widely held view, and one that we will assume here, is that a person's welfare is the fulfilment of her desires and that in consequence, due to conceptual connections between pleasure and desire fulfilment, pleasure is an important part of welfare.[2]

Convergence

To what extent do our sample criteria of R and M move in sync? One source of hitches can be an agent's beliefs, which play a role in one criterion but not in the other. For example, even the most fervent utilitarian can have erroneous beliefs about the impact of her actions on the amount of general welfare, and those beliefs can make it rational for her to do what is morally wrong.

If we leave aside the threat posed by deficient beliefs, we reach more significant notions of convergence and divergence. Let us call an agent well-informed if her beliefs are in such a good state that the rational thing for her to do would not change if we improved them further (that is, if we corrected false beliefs or added true ones). Using well-informedness as the stepping-stone, let us understand "convergence" and "divergence" as follows: R and M converge in as far as the actions of well-informed agents that are morally required are also rational; R and M diverge in as far as the opposite is the case.

Convergence through Moral Desires

If instrumentalism is on the right track, the principal forces of convergence will have to be desires that point in the right direction. Less metaphorically speaking, desires that, provided the agent is well-informed, have a propensity to make it rational for her to perform an action that is morally right. Let us call them RM desires. Life abounds with such desires, and one challenge is to produce a helpful classification.

Some RM desires aim directly at something that would be morally positive in itself. Figure 2 presents them as the "moral desires". Some of those attitudes even 'bring up' the topic of morality. Examples are desires that the world be a better place, to do the right thing, or to be a virtuous person. Other moral desires do not invoke morality as such, but exhibit the relevant directness all the same. Depending on what is morally positive in itself, examples might be

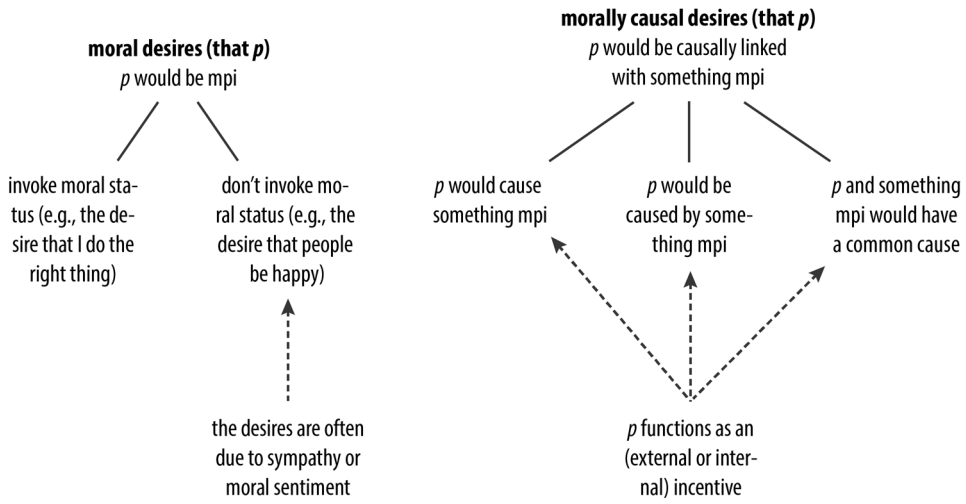

Figure 2: Some important kinds of RM desires. The abbreviation "mpi" stands for "morally positive in itself". One and the same desire can belong to more than one kind, even on the same level.

desires that some specific people be happy, that there be a lot of happiness, that everybody be treated equally – or desires to keep promises and to refrain from telling lies. The propensity that defines RM desires is present in either case. Applied to our sample moral doctrine, utilitarianism: both if Mary desires to do the right thing and if Mary desires to maximize welfare world-wide, it holds true that, provided she is well-informed, the desire will tend to make it rational for her to do the right thing.

Sympathy and moral sentiment, widespread as they are, serve as high-yield sources of those moral desires that do not invoke morality.[3] If Mary sympathizes with others, her sympathy is likely to give rise to or even to constitute a desire that others fare well. And if Mary has moral sentiments (for example, sentiments of moral admiration, indignation, or satisfaction) in view of acts of a certain kind, the sentiments are likely to give rise to or even to constitute desires for or against performing acts of that kind. The two sources are so rich that there have been proposals to feed morality from them alone, giving us an "ethics of sympathy" or a "sentimentalist ethics".[4]

Are all moral desires contingent? Is it just as possible for a person to have them as to lack them? The rationality of morality would be a more robust affair if there were numerous strong moral desires that we cannot fail to have. An argument has been devised that purports to establish the existence of those resources. Everybody, the argument aims to establish, necessarily desires that other people have pleasure, with the strength of those desires proportional to that of the pleasure at issue. The idea is that, if you fully represented to yourself that another person experiences a specific pleasure, you would (this being entailed by full representing) experience that very pleasure yourself – and would thus yourself be pleased while representing. And since a disposition to be pleased when fully representing a state of affairs to oneself is a desire that the state of affairs obtain, you desire that the other experience the pleasure.[5] If the argument works, those desires – which on conceptual grounds everybody holds regarding everybody else all the time – will do a sizable part of the work that we are anxious to see done.

Convergence and Rationality without Egoism

Our glance at other-regarding moral desires, no matter whether they are necessary or contingent, is a good occasion for setting aside egoism. One might think on the following grounds that the relations between rationality and morality are particularly strained: (i) they would be strained if rationality required us to act egoistically, and (ii) according to instrumentalism rationality requires exactly that, because the word "egoistic" stands for the property of acting in the light of one's own desires.

The train of thought is misguided. The notion of "egoism" that is advanced in claim (ii) is both unusual and apt to weaken claim (i). The usual understanding is that acting on one's own desires is not a sufficient condition for being egoistic, but that it also matters what those desires are. The usual understanding is that, if a good-hearted person strongly desires that other people

fare well and acts on *such* desires, she is not an egoist but an altruist, while an egoist is a person who lacks or fails to put into action *such* desires. To be sure, somebody might go along with that understanding itself but still link instrumentalism to egoism by adding the claim that such desires do not exist. However, we know of no sound argument for the additional claim.[6]

The misclassification of rationality as egoistic can also result from sloppy thinking about the *combination* of instrumentalism and the theory of welfare as desire fulfilment. It is true that the combination entails some connection between rational action and the agent's welfare: in acting rationally at point of time t, the well-informed person stage Mary-at-point-of-time-t maximizes the fulfilment of its desires and cannot help thereby maximizing its own welfare. We should beware of letting that connection blur the picture. In the first place, the rational person (or person stage) at issue need not be concerned with her (or its) own welfare and can act, due to the nature of her (or its) desires, highly altruistically. Secondly, the person stage that acts can still fail to maximize the welfare of the entire person, who is extended over time and may have different desires later. Thirdly, even the limited connection we are looking at requires as one ingredient a certain view of welfare; if we adopt a hedonistic view instead (seeing the welfare of an entity as the pleasure that the entity has and not as the fulfilment of her desires) and keep instrumentalism, the connection vanishes.

According to most conceptions of rationality, instrumentalism included, rationality does not require us to act egoistically. We need to distinguish the question how rationality relates to morality from the questions how egoism, how prudence, and how an agent's self-interest, self-love, or welfare relate to morality.[7] Many well-known discussions from the history of philosophy are largely about questions of the second kind. Plato, for example, assigns a key role in the *Republic* to the case of Gyges, who is ruthless in using for his own advantage his power to become invisible; Aristotle writes about virtue as a constituent of an agent's flourishing in the *Nicomachean Ethics*, and Henry Sidgwick

about individual vs. universal happiness in *The Methods of Ethics*. Those discussions apply to our question at best partly or indirectly.

Convergence through Morally Causal Desires

Some RM desires, also indicated in figure 2, are related to the good or the right in a different way. If a person desires p and p would be causally linked to something that is morally positive in itself, we will speak of a "morally causal desire". We distinguish subgroups of such desires by distinguishing three kinds of the linkedness.

A desire that p is in the first subgroup if p would cause something morally positive. For example, a person desires to cultivate a garden, to keep the kitchen clean, or to mend broken bones, and her doing these things would cause pleasure in the beholders, eaters, patients and would thus cause something that is morally positive in itself. A desire that p is in the second subgroup if, conversely, something that is morally positive in itself would cause p. For example, Mary desires that her parents treat her to cake, which the parents do only when, and in that case because, they are happy. The treat is desired and is the effect of the happiness, with the happiness being morally positive in itself. A desire that p is in the third subgroup if p and something that is morally positive in itself have a common cause. Consider, for example, heartless Mary, who does not care about the victims of malaria but who desires to be praised at a charity ball for donating to the fight against malaria. If Mary donates, neither does the praise cause the thing that is morally positive in itself (the praise does not causally affect the saving of the lives) nor vice versa, but the two have a common cause: Mary's donation. In such cases, too, the one comes with the other, and that matters for RM purposes.

The key feature, wherever morally causal desires are in play, is the indirectness. A desire can fail to be directed at anything that would be morally positive in itself and still tend to make actions rational that are (or that cause things

that are) morally positive in themselves. Links of the relevant kind are legion: you do the right thing and some 'other' desire of yours is fulfilled. In that sense, large chunks of morality are connected to rationality through carrots and sticks.

The incentives can be external – think of fellow-citizens who in response to your morally positive behaviour honour and help you and refrain from ignoring, deriding, jailing, or lynching you. The incentives can also be internal – think of the joys of believing to have done what was good or right and of the absence of pangs and remorse. Even when the incentives are internal, the constellation differs from the one that characterizes the first main group of RM desires, the moral desires. Two distinctions apply. In the first place, when you desire *your joy* of doing good and when you desire *to do good*, those are two different desires.[8] Secondly, we should in some cases distinguish even for one and the same desire between the reason to put it into one group and the reason to put it into another. Consider, for example, your desiring your own joy of having done good. The reason to count that desire as a morally causal one (your joy is caused by something morally positive) is different from the reason to count it as a moral one (your joy, too, is something morally positive in itself). One desire can be both.

Incentives are studied by game theorists in particular. We understand quite well by now, with regard to several moral standards, how even agents who have no moral desires and are at the same time without ifs and buts instrumentally rational find themselves drawn into actions and outcomes that are morally positive. Under various conditions repeated encounters in the same group of such agents become, not least because rewards and punishments can emerge, become a breeding-ground for various amounts of co-operation, equality, justice, solidarity, trust, and public good. While significant general results, most prominently a family known as "the folk theorems" of game theory, have been established by mathematical methods of the more traditional

kind, there is also evidence from agent-based computer simulations, sometimes involving entire artificial societies: the rational thing to do for the virtual agents, it turns out, is often to spare or even to assist each other.[9]

So much for the many devices in human minds and societies that make it rational for people to do the right thing a lot of the time. On the other hand, given criteria of R or M resembling our two samples, there is not much hope of full convergence. In the example from our opening paragraphs, the agent might favour abstract paintings over human lives, which in the absence of competing considerations would make it the rational thing for her to violate her moral obligation. When incentives and the agent's moral desires do not add up, rationality will take a stand against morality.

Resisting Divergence by Aligning Rationality with Morality

The dominant impetus among philosophers is to keep the divergence of R and M in check. Some philosophers go as far as to let their thinking on R itself or on M itself be governed by the premiss of "moral rationalism" – the claim that every action that is morally required is rational.[10] With that sweeping premiss or without, answers to the question "Why be moral?", understood as the question "Why *act* morally?", are sought-after, and the risk that even with regard to well-informed agents sometimes no good answer emerges is often perceived as an invitation to rethink R or M. Those who want to rethink in order to reduce or even eliminate divergence have essentially two options: they assimilate rationality to morality or vice versa. They could also combine the two moves, moralizing rationality *and* rationalizing morality, but for the bigger picture it will suffice if we treat each of the two in turn.

Smaller Departures from Instrumentalism

As to alternative views of practical rationality, modest deviations from instrumentalism make some difference regarding divergences. One example is the temporal extension of the conative basis. We could modify the instrumentalist criterion so that it covers not just the desires that the agent has at the time of acting, but all those that she had, has, or will have. The modification would prevent the rational agent from behaving ruthlessly towards her past self or her future self. Morally speaking, it would be a step in the right direction – but only a small, intrapersonal one. Anchoring in a rational agent regard for the welfare of all her own person stages is a far cry from anchoring in that agent regard for everybody's welfare.

A second example originates in the observation that, if instrumentalism is correct, rational agents can get caught in traps of practical rationality. Such traps are situations in which, if every agent acts rationally, each of them receives, predictably, less fulfilment of her desires than she would if everyone acted irrationally. General compliance with the requirements of rationality thus leads to the opposite of the moral goal, which is general welfare.

Many real-life situations appear to be traps of practical rationality. Consider societies that become much nastier because people carry weapons. Each individual reasons that to carry a weapon is likely to be more conducive to the fulfilment of her desires either way – that is, both if people whom she encounters carry a weapon and if they don't. Because of the individual decisions for weapons the considerable advantages of a weaponless society and of avoiding an arms race remain out of reach for all. Even if a weaponless society could still be achieved by changing the individual decisions through the imposing of sanctions, the sanctioning itself would gobble resources and thus welfare. The ubiquity of such traps in human interaction and the unequivocalness of the moral setbacks – less welfare for one and all – have incited a vast complex of research in ethics, economics, and psychology.[11]

Can we devise a criterion of rationality that preserves the spirit of instrumentalism but spares us the traps? David Gauthier is one of the theoreticians who have tried.[12] Gauthier suggests, controversially, that a rational agent can choose, "on utility-maximizing grounds, not to make further choices on those grounds", but to adopt a certain stance that will determine her behaviour. She will then be a "constrained maximizer". The stance that Gauthier has in mind is roughly this: I will do, provided that so will the others who are involved, my part in making possible an outcome that is better for everybody than the outcome that unconstrained maximizers could achieve. Since an agent who has recognizably adopted such a stance will sometimes produce and reap fruits of co-operation that are not available to one who hasn't, it is rational to become such an agent. Individual instrumental rationality is declared to be a tad more collective than we thought.

If Gauthier's claims are sound and help us make headway with the RM question, they do so within limits. Gauthier's ambition is restricted to establishing rational support for a morality of *mutual* benefiting. He has no ambition to provide support for moral obligations towards beings who are in need but have nothing to offer.

Larger Departures from Instrumentalism

We find more radical consequences for the shape and extent of divergence when we turn our attention to more radically different pictures of practical rationality. The main move is the dethroning of desire.

Suppose that we break with instrumentalism by replacing the appeal to desire fulfilment with an appeal to conformity to reasons. Let us agree that facts can be reasons and that constellations of facts (for example, regarding pains and cures) can make it the case that there is, altogether, for a person more reason to perform one action (say, to see her cardiologist) than another (say, to see her homoeopath). The new criterion of rationality could then say: *An action that*

the agent can perform is rational if and only if she believes that there is not more reason for her to perform another action instead.[13] Suppose further that we take many reasons to be "worldly" in that they do not involve the agent's desires. Perhaps, for example, *that Peter's education would be finished if he received a donation* is a reason to donate, even if the agent desires no such finishing or donating and desires nothing that comes with it; or perhaps *that serenading the moon would increase the glory of the moon* is a reason to serenade the moon.[14]

The impact on the RM nexus would be momentous. The task of showing that it is rational for a well-informed agent to do what morally she ought to do no longer has as its central component the task to track down a fitting constellation of desires of hers, but the task to identify a fitting constellation of reasons for her to do things. In order for the entire scheme to be successful, worldly reasons for actions need to exist *and* to have a considerable affinity to morality *and* to be connected to rationality as stated by the new criterion. Whether all or at least some of the three conditions are fulfilled is controversial.

Which morality you take to be connectible to rationality-in-the-new-spirit will depend on your views, possibly your intuitions, about the realm of reasons. If morality is concerned with promise-keeping, loyalty, or the increase of human knowledge, and so are reasons to do things and in the same proportion, then every well-informed rational agent will do the right thing: keep her promises etc. Our sample moral doctrine, utilitarianism, is no exception. If there is always most reason to maximize the amount of the fulfilment of everybody's desires, then every well-informed rational agent will do the right thing and maximize that amount.

Following a markedly different path, Immanuel Kant, too, ends up advocating some variety of rational benevolence, maximization included. We read that a rational being would try, "as far as he can, to advance the ends of others" and would come to the conclusion: "the ends of a subject that is an end in itself must, as much as possible, also be *my* ends".[15] In which way according to Kant

rationality secures so much morality is extraordinarily contentious, even among his followers. He draws on the claim that a rational being is autonomous in the literal sense of giving herself a law. To Kant that feature appears imbued with moral significance. Being autonomous, the agent's rational will is not pushed around by anything, not even by the agent's own inclinations; it finds itself with nothing left to be constituted by than the respect for rationality itself and for people who have it and for their ends; and, such being the character of laws, that will is general, not concerned with one person or group in particular. Sound statements of those connections remain a desideratum.

As expected, the picture of the relation between R and M changes when that of R changes. Our brief encounters with the "worldly reasons, not desires" approach and the "autonomy, not desires" approach have illustrated tectonic movements and the hopes of convergence that can be pinned on them.

Resisting Divergence by Aligning Morality with Rationality

If divergence is to be avoided, what about getting to work at the other end? We could truncate morality. The fewer actions are morally obligatory in the first place, the smaller the risk that an action is morally obligatory but irrational.

The demands of some moralities are so removed from most agents' conative constitution that cutting down on the demanding looks particularly promising. According to utilitarianism, for example, an agent ought to give the same weight to her own welfare and to that of her nearest and dearest as to everybody else's. Since hardly an agent's desires manifest such impartiality, utilitarian obligations that it is irrational for the obligated agent to comply with are thick on the ground.

Various moral prerogatives for agents have been suggested. We could permit the agent, for some factor $k > 1$, to attach up to k times as much weight

in her decisions and actions to her own welfare as to the welfare of everybody else; or permit her, for some $l > 0$, never to give up more than l units of her own welfare; or permit her, for some threshold value m, never to make her own welfare fall below m.[16] We could think of other permissions, too, concerning her projects rather than her welfare or concerning the welfare of those she is close to rather than just her own. Every such prerogative is a loosening of utilitarianism; it would allow the agent to be partial in the sense of granting in her decisions extra force to this or that, even if the amount of world-wide welfare suffers.

Prerogatives are apt to narrow the gap between R and M, but will not close it. A morality that involves a prerogative will still demand *something* (for instance, that the interests of others be respected to *some* extent), and there is bound to be some well-informed agent who in some situation has desires that make it irrational for her to comply even with those moderate demands.

Structurally, contract theories in the Hobbesian tradition inhabit the same middle ground. Those theories are reciprocitarian. They say, with various qualifications, that what a person morally ought to do in relation to another person is to play by rules that satisfy the following condition: the fulfilment of the desires of each of the persons would increase if each of them, rather than none, played by those rules. For example, if both refrain from insulting each other, that will save both of them some distress.

Once again we find the curtailing of moral obligation. If contractarians are right, there are fewer moral obligations than we thought and thus fewer that it might be irrational to meet. Most notably, about all persons (and other beings) who are not in a position to increase the fulfilment of her desires, the contractarian will say that she owes them nothing. And once again we also find that the curtailing does not suffice to provide full alignment with rationality. Sometimes a well-informed agent will have and see the possibility of maxim-

izing the fulfilment of her desires by breaking even the few rules of that "minimal morality" and getting away with it. And no earthly regime of sanctions will eliminate such possibilities, since all such regimes are gappy.[17]

Some thinkers have suggested that we go one step further and prevent all divergence by fully rationalizing morality. The proposal is to combine moral rationalism with the claim that instrumentalism is by and large on the right track. Gilbert Harman endorses the combination. He writes: "If *S* says that (morally) *A* ought to do *D*, *S* implies that *A* has reasons to do *D* […]." And he continues: "I assume that the possession of rationality is not sufficient to provide a source for relevant reasons, that certain desires, goals, or intentions are also necessary." Since agents might lack the relevant attitudes, Harman infers "that there might be no reasons at all for a being from outer space to avoid harm to us" and "that, for Hitler, there might have been no reason at all not to order the extermination of the Jews".[18]

The moral consequences are remarkable. If the moral "ought" requires reasons, and reasons require desires, and the desires aren't in place and thus neither are the reasons, then neither is the moral "ought". While Harman invites us to say other things about Hitler (for example, that Hitler is evil or our enemy), he claims that the moral "ought" is out of place. It is "odd to say", so Harman, that "it was wrong of Hitler to have acted as he did" or that "Hitler ought morally not to have ordered the extermination of the Jews".

And so the rationalization of the moral "ought" would be completed. The approach is Procrustean. There is no mismatch between the moral "ought" and the practical rationality of its addressees because every instance of a moral "ought" that would not conform is discarded.

Accepting Divergence

Should we accept that R and M diverge? We should if it seems to us that they do and we see no reason to resist the claim. The two parts of that condition deserve separate treatments.

Finding Some Divergence in the First Place

Not many of us enter the inquiry holding fully convergent views of R and of M each of which they deem plausible in its own right, independently of any pressure to see the two in line. It is true that there is full convergence according to some views of R and of M that we have encountered, but the independent plausibility of those views is the crux.

On the rational side, how plausible is the claim that something other than the agent's aims and projects rules the roost? The subjective picture exerts a considerable pull: if you've set your heart on something, it is rational for you to go after it; and if you haven't, it is not. On the moral side, the questions are inverse. How plausible is the claim that, provided you are indifferent to other people's welfare and have to fear no backlash from wrecking it, you are morally permitted to wreck it? On either side, the claims that would secure full convergence do not have the ring of truth.

The problem does not just arise when, at least regarding well-informed agents, one of the two domains is claimed to fully look the way we always took the other one to look, with the implausibility due to the fact that *one* view does all the reaching out. If a view of R and a view of M met half-way, the implausibility would be distributed evenly, but not lessened. Although we pointed out that some independently plausible components with a conciliatory effect may well be missing from our two sample doctrines (which we didn't call sample doctrines for nothing), no such components are in sight that would happen to dovetail, resulting in a perfect fit.

Not Theorizing away the Divergence We Find

Matters would look different if our thinking were driven by the will to rule out divergence. Maybe there are general considerations that speak against divergence and that should be taken into account when we form our views of R and M? It is surprisingly difficult to find distinct statements of such considerations in the literature, but here is a brief attempt to articulate and assess some candidates.

The argument from *the negativity of non-compliance* says: "Since people by and large act rationally, every morally required act that is irrational is a morally required act that is unlikely to be performed – which is a bad thing." We respond that any such badness would be a sad truth and that good theorizing should acknowledge truths, sad ones included. The badness at issue does not provide a right kind of reason for changing our views of R or M.

A follow-up argument adduces *the pointlessness of moral judgment due to the negativity of non-compliance*: "The point of engaging in moral judgment is to avoid the said negativity. Why bother if the project does not boost the right and the good through compliance?" Part of the answer is that not only are there many different functions that moral thinking, judging, and speaking have, but also many different paths, including indirect ones, on which the function of boosting the good and the right can be fulfilled. Some of the moral point of considering or making or uttering the moral judgment that, say, Mary ought to donate half of her spare time to a certain cause may well be independent of the factual question whether Mary ends up making that donation. Sorting out and signalling our moral view of the matter can help shaping decisions, education, outlooks, politics, relationships, and sanctions in a myriad of ways, many of which do not even relate to Mary in particular.

Next, there is the threat of *the unjustifiedness of moral judgments*: "To 'provide morality with a foundation' or to 'justify moral judgments' is or includes showing that it would be rational to act, if the opportunity arose, in line with

the moral judgments at issue. Moral judgments that are subject to divergences are therefore unfounded and unjustified." The complaint is worth pondering. Still, you justify something *to somebody*, and so the premiss of the complaint licenses at best the conclusion that in cases of divergence the "ought" judgment has not been justified to *all* persons who according to the judgment ought to do a certain thing under certain circumstances. But maybe to some of them. Secondly, the premiss of the complaint is controversial in that there are other conceptions of what it is to justify a moral judgment. Thirdly, let's not forget that generally speaking, since justification stops somewhere, the use of unjustified items might be respectable.

A final argument asserts *the inacceptability of normative impasses*: "When an agent grasps the fact of the situation and the morally required action is irrational, what is there left for us to tell her? We can tell her that two kinds of norms that apply to her impending action point in opposite directions, irreconcilably so, and that we have no third kind of norm that would adjudicate between them. We can wish her good luck at the normative crossroads and move on. None of that is of any help to her, and theorizing about normativity should do better than that. It should avoid divergences."

The most general reply to the objection is that theorizing about normativity should "do better" only if there is independent evidence that it got the lay of the land wrong. That evidence would need to be produced – and will hardly consist in the fact that something is or feels awkward for a well-informed agent. Moreover, the objection misses its target, divergence, because divergence does not entail the alleged source of awkwardness, the absence of normative guidance. Divergence allows for the possibility that there is a boss: one of the divergers or some adjudicator. The possibility is very much alive in the literature on normative pluralism, where the view is not rare that divergences between kinds of norms coexist with unequivocal overall norms.[19]

When it comes to R and M, the observation that divergence and guidance can coexist gains in stature, since one of the two divergers is practical reason itself. There is a fairly straightforward sense in which practical reason is always in charge. Practical reason deals with practical reasons – with all of them. The fascinating question whether all of them involve agents' desires makes no difference for the following consideration, which is quite general. Neither does it make a difference whether some kind of incommensurability threatens to hold *among* an agent's practical reasons. We may ignore that possibility here because we're asking whether the spectre of lacks of guidance that originate *between* R and M might justify the axiomatic excluding of a divergence of R and M. Surely no such lacks could justify such an excluding if there is lack of guidance even *within* R. Thus, what remains to be looked at is only the other case, in which all is well within R: there is some balance of all practical reasons that an agent has and thus something that the agent has most reason to do.

And now to the question where the balance leaves *moral* practical reasons. They can relate to the "most reason" verdict in two different ways, but cannot escape the verdict nor the guidance it gives. We can understand moral practical reasons either as being practical reasons of a certain kind (think of yellow bicycles, which are bicycles), in which case the balance of all practical reasons will have taken them into account – or as not being practical reasons (think of "root beer", which is not beer, and of "toy money", which is not money), in which case the balance of all practical reasons will not have taken them into account. There is guidance by the balance either way. It is guidance on the level "most practical reason", a level on which all practical reasons – that is, all reasons to do things – have been taken into account. That seems guidance enough.

From the list of objections against divergence not much is left. It seems that we should bemoan, but not deny, the existence of divergence.

Conclusion

We have explored in which sense and why it is in many cases rational for people to do what morally they ought to do, but also why with respect to many cases even of well-informed agents the diagnosis is controversial. We have sided with the common-sense view that performing actions that are stupid (to use the laity's term for "irrational") and performing actions that are morally wrong are two very different kinds of shortcomings. The action that is stupid can be altruistic, benevolent, beneficent, and morally right, and the action that is not stupid can be egoistic, malevolent, maleficent, and morally wrong. Given the duality, theoreticians of the rational and the moral will keep or turn their spotlights on the kinds, mechanics, and extents of convergence and divergence, including the metanormative challenges posed by competing norms. In practice, both desires and ways of fulfilling them ought to be shaped – and that "ought" is a moral one – so that divergence is reduced.

Notes

* More people have given us a hand regarding some aspect or other of this article than we can list here, and we are grateful to every one of them. Special thanks go to Kevin Baum, Inga Lassen, Susanne Mantel, Helge Rückert, Rudolf Schüßler, Stephan Schweitzer, and Christian Wendelborn.

1 More on doctrines like instrumentalism and on their rivals in Hutcheson 1728, sec. 1 of treatise 2, Millgram 2001, Schroeder 2007, and Schmidt 2016; Fehige 2001 defends the doctrine and provides, in note 1, further references.

2 For the equating of welfare with desire fulfilment see von Wright 1963, esp. secs. 5.9 and 5.11, Carson 2000, chap. 3, and (for numerous further sources, too) Wessels 2011; for the claim that pleasure is one kind of desire fulfilment, see Fehige 2004, esp. 143–45, Heathwood 2007, and their references. As to normative ethics: utilitarianism and its rivals are sketched in Vaughn 2013, chap. 2, and treated more thoroughly in Copp 2006, pt. 2; a helpful introduction to utilitarianism is Bykvist 2010.

3 Hutcheson 1728 covers in some detail the desires associated with the "publick sense", which is the disposition "to be pleased with the *Happiness* of others, and to be uneasy at their *Misery*" (art. 1.1 of treatise 1), and with the "moral sense", which is the disposition to have moral sentiments. More on the connections between morality, desire, sentiment, sympathy in Bricke 1996, esp. chap. 6, Fehige 2004, and Fehige/Frank 2010.

4 Single-source approaches to morality have been championed, for example, by Arthur Schopenhauer (1841, esp. sec. 16), who counts on sympathy, and by Francis Hutcheson (1725, preface and treatise 2) and David Hume (1751, app. 1), who count on the moral sentiments.

5 The argument is put forth in Fehige 2004; the book includes a discussion, in chap. 6, of the limits of even that connection between R and M.

6 Lucid treatments include Hutcheson 1728, esp. art. 1.3 of treatise 1 and the beginning of treatise 2, and Sharp 1923, sec. 2.

7 Gregory Kavka captures the difference with maximum clarity when he distinguishes the "Wider Reconciliation Project" (1985, sec. 5) from a narrower one. More on handling egoism and its relation to morality in Cholbi 2011.

8 More on the important distinction, for example, in Hutcheson 1728, art. 1.4 of treatise 1, Rashdall 1907, esp. 17–18, 28–32, and Schlick 1930, secs. 2.6–2.8.

9 For the first kind of approach, see Fudenberg and Tirole 1991, chap. 5, esp. sec. 5.1.2, and Maschler et al. 2013, chap. 13, esp. secs. 13.5 and 13.6. The agent-based simulation approach is surveyed by Gotts et al. 2003; a telling example is Hegselmann 1998.

10 Moral rationalism is also known as "the claim of overridingness" and is highly controversial; contributions to the dispute include Scheffler 1992a, chap. 4, Cholbi 2011, esp. sec. 1, Portmore 2011, and Dorsey 2012. An illuminating critical study of ways of engineering convergence is Brink 1992.

11 Petersen 2015 can serve as a gateway to the area; see also Binmore 1994.

12 A good starting point is Gauthier 1986, esp. chap. 6, which has the upcoming quotation on p. 158. McClennen's "resolute choice" (1985) bears some resemblance to Gauthier's "constrained maximization". Gauthier later espouses a conception of rationality that he acknowledges is morally charged to begin with (2013, 624); for a similar step, see McClennen 2012.

Other devices, too complex to sketch here, have been invoked against the traps: the claim that people who are in similar circumstances are bound to act similarly (see Davis 1985 and, on "mirror strategies", J. V. Howard 1988); the proposal to shift our attention from Nash equilibria to "dependency equilibria" (Spohn 2009, foreshadowed by Aumann 987); and, although with an emphasis on the explanatory rather than the normative dimension, the conceptualization of a relevant decision situation as a certain kind of "metagame" that involves partial commitments (N. Howard 1971, esp. secs. 2.5, 3.1, 3.2) or of the decision-makers as members of a group (Bacharach 2006, esp. sec. 4 of the "Conclusion", and Butler 2012). For critical thoughts on some of the approaches see, e.g., Binmore 1994, chap. 3.

13 The view that rationality is the corresponding of actions to beliefs about reasons can be spelt out in quite different ways; see, for example, Scanlon 1998, secs. 1.1.3–1.1.5, Parfit 2011, secs. 1 and 17.

14 The claim that there are worldly normative reasons for actions has acquired quite a few supporters. Examples are Thomas Nagel (1970, esp. chap. 10, and 1986, esp. secs. 8.4, 8.5, 9.2, 9.3), Thomas Scanlon (1998, secs. 1.9 to 1.11), Jonathan Dancy (2000, chaps. 2 and 5), Philippa Foot (2001, chap. 4), Frederick Stoutland (2001, secs. 3.2 and 3.3), and Derek Parfit (2011, secs. 1 to 15).

15 The quotations are from Kant 1785/1903, 430; for the moral impact of autonomy, see, e.g., pp. 405, 428–34, 444, 447–52. Kant later invokes the moral law as a "fact of reason" (1788/1908, 31–32, 42–43); whether in doing so he renounces, summarizes, or supplements his justificatory efforts is controversial. After Kant, attempts to ground morality in the respect for rationality itself have been numerous and varied; Smith 2011 is one case in point.

16 The three kinds of prerogatives are considered, one each, in Scheffler 1992b, sec. 1, Mulgan 2001, sec. 5.5, Nagel 1986, 202. For further reflections on possible kinds and shapes of prerogatives, see Wessels 2002 and Stroud 2010. Some prerogativists make it clear that anti-divergence is where they come from. It is for the sake of keeping M in the orbit of R that they conceive of M as partial

and as in that respect non-utilitarian; see Nagel 1986, 202–3, and Portmore 2011.

17 A particularly acute analysis of the connections between R and M in the contractarian project is provided by Rainer Trapp (1998), who also explains (339, 356–59) that it will not always be rational for a person to do what by contractarian lights she morally ought to do. Contractarians who acknowledge the divergence include Peter Stemmer (2017, 646–48, esp. note 34) and Gregory Kavka (1985, 305–8 and, most clearly in terms of rationality, sec. 5). That a contractarian like Gauthier might avoid the divergence by operating with a modified conception of rationality (Gauthier 1986, chap. 6) is a different matter.

18 The preceding quotations are from Harman 1975, 9; the subsequent ones, from Harman 1977, 107. Stemmer has retracted the crucial claims (2017, note 34), but used to travel a very similar path. For a while he saw norms geared so radically to the addressee's wanting that even an accidental hole in the sanctioning was considered to constitute a hole in the norm itself. Insofar as an individual action would not be followed by a sanction that the agent herself wants to avoid, so the retracted claims run, the norm not to perform that action "does not exist", and the action "is not really forbidden" (2008, 181).

19 That domain-specific "ought" judgments relate to overarching ones, which have the last word, is argued by McLeod (2001) and Woods (2018, sec. 10.2.2). In a similar spirit, Case (2016) provides a powerful argument for the conditional claim that, if you accept "source pluralism" and "conflict", you are committed to accepting "authoritative adjudication". However, support for the claim that "ought" judgments from different domains diverge is wider and comes also from authors who deny that there is an adjudicative level – see Baker 2018 and the sources given there.

References

Aumann, Robert T.,"Correlated Equilibrium as an Expression of Bayesian Rationality", *Econometrica* 55 (1987): 1–18.

Bacharach, Michael, *Beyond Individual Choice*, Princeton 2006, Princeton U.P.

Baker, Derek, "Skepticism about Ought *Simpliciter*", *Oxford Studies in Metaethics* 13 (2018): 230–52.

Binmore, Ken, *Playing Fair*, Cambridge, Mass.: MIT Press, 1994.

Bricke, John, *Mind and Morality*, Oxford 1996: Oxford U.P.

Brink, David O., "A Puzzle about the Rational Authority of Morality", *Philosophical Perspectives* 6 (1992), 1–26.

Butler, David J., "A Choice for 'Me' or for 'Us'? Using We-Reasoning to Predict Cooperation and Coordination in Games", *Theory and Decision* 73 (2012): 53–76.

Bykvist, Krister, *Utilitarianism*, London: Continuum International, 2010.

Carson, Thomas L., *Value and the Good Life*, Notre Dame, Ind.: University of Notre Dame Press, 2000.

Case, Spencer, "Normative Pluralism Worthy of the Name is False", *Journal of Ethics and Social Philosophy* 11 (2016): 1–19.

Cholbi, Michael, "The Moral Conversion of Rational Egoists", *Social Theory and Practice* 37 (2011): 533–56.

Copp, David (ed.), *The Oxford Handbook of Ethical Theory*, Oxford: Oxford U.P., 2006.

Dancy, Jonathan, *Practical Reality*, Oxford 2000: Oxford U.P.

Davis, Lawrence H., "Is the Symmetry Argument Valid?", in *Paradoxes of Rationality and Cooperation*, ed. by Richmond Campbell and Lanning Sowden, Vancouver: University of British Columbia Press, 1985: 255–63.

Dawes, Robyn M., "Social Dilemmas", *Annual Review of Psychology* 31 (1980): 169–93.

Dorsey, Dale, "Weak Anti-Rationalism and the Demands of Morality", *Noûs* 46 (2012): 1–23.

Fehige, Christoph, "Instrumentalism", in *Varieties of Practical Reasoning*, ed. by Elijah Millgram, Cambridge, Mass.: MIT Press, 2001: 49–76.

———, *Soll ich?*, Stuttgart: Philipp Reclam jun., 2004.

——— and Robert H. Frank, "Feeling Our Way to the Common Good", *The Monist* 93 (2010): 141–65.

Foot, Philippa, *Natural Goodness*, Oxford 2001: Oxford U.P.

Fudenberg, Drew, and Jean Tirole, *Game Theory*, Cambridge, Mass.: 1991, MIT-Press.

Gauthier, David, *Morals by Agreement*, Oxford: Oxford U.P., 1986.

———, "Twenty-Five On", *Ethics* 123 (2013): 601–24.

Gotts, N. M., and J. G. Polhill and A. N. R. Law, "Agent-Based Simulation in the Study of Social Dilemmas", *Artificial Intelligence Review* 19 (2003): 3–92.

Harman, Gilbert, "Moral Relativism Defended", *Philosophical Review* 84 (1975): 3–22.

———, *The Nature of Morality*, New York: Oxford U.P., 1977.

Heathwood, Chris, "The Reduction of Sensory Pleasure to Desire", *Philosophical Studies* 133 (2007): 23–44.

Hegselmann, Rainer, "Experimental Ethics", in *Preferences*, ed. by Ulla Wessels and Christoph Fehige, Berlin: Walter de Gruyter, 1998: 298–320.

Howard, J. V., "Cooperation in the Prisoner's Dilemma", *Theory and Decision* 24 (1988): 203–13.

Howard, Nigel, *Paradoxes of Rationality*, Cambridge, Mass.: MIT Press, 1971.

Hume, David, *An Enquiry concerning the Principles of Morals*, London 1751: A. Millar.

Hutcheson, Francis, *An Inquiry into the Original of Our Ideas of Beauty and Virtue*, London 1725: J. Darby.

———, *An Essay on the Nature and Conduct of the Passions and Affections: With Illustrations on the Moral Sense*, London 1728: J. Darby and T. Browne.

Kant, Immanuel, *Grundlegung zur Metaphysik der Sitten* (1785), in *Kant's gesammelte Schriften*, part of vol. 4, Berlin 1903: Georg Reimer; English quotations from *Groundwork of the Metaphysics of Morals*, transl. by Mary Gregor and Jens Timmermann, Cambridge (England): Cambridge U.P., 2011.

———, *Kritik der praktischen Vernunft* (1788), *Kant's gesammelte Schriften*, part of vol. 5, Berlin: Georg Reimer, 1908.

Kavka, Gregory S., "The Reconciliation Project", in *Morality, Reason and Truth*, ed. by David Copp and David Zimmerman, Totowa, N.J.: Rowman & Allanheld, 1985, 297–319.

Maschler, Michael, Eilon Solan, and Shmuel Zamir, *Game Theory* (first publ. in Hebrew in 2008), transl. by Ziv Hellman, Cambridge (England): Cambridge U.P., 2013.

McClennen, Edward F., "Prisoner's Dilemma and Resolute Choice", in *Paradoxes of Rationality and Cooperation*, ed. by Richmond Campbell and Lanning Sowden, Vancouver: University of British Columbia Press, 1985: 94–104.

———, "Rational Cooperation", *Synthese* 187 (2012): 65–93.

McLeod, Owen, "Just Plain 'Ought'", *The Journal of Ethics* 5 (2001): 269–91.

Millgram, Elijah (ed.), *Varieties of Practical Reasoning*, Cambridge, Mass.: MIT Press, 2001.

Mulgan, Tim, *The Demands of Consequentialism*, Oxford: Oxford U.P., 2001.

Nagel, Thomas, *The Possibility of Altruism*, Oxford: Oxford U.P., 1970.

———, *The View from Nowhere*, Oxford: Oxford U.P., 1986.

Parfit, Derek, *On What Matters*, Oxford: Oxford U.P., 2011.

Petersen, Martin (ed.), *The Prisoner's Dilemma*, Cambridge (England): Cambridge U.P., 2015.

Portmore, Douglas W., "Consequentialism and Moral Rationalism", in *Oxford Studies in Normative Ethics* 1 (2011): 120–42.

Rashdall, Hastings, *The Theory of Good and Evil*, vol. 1, Oxford: Clarendon Press, 1907.

Scanlon, T. M., *What We Owe to Each Other*, Cambridge, Mass.: Harvard U.P., 1998.

Scheffler, Samuel, *Human Morality*, Oxford: Oxford U.P., 1992a.

———, "Prerogatives without Restrictions", *Philosophical Perspectives* 6 (1992b), 377–97.

Schlick, Moritz, *Fragen der Ethik*, Vienna 1930: Julius Springer.

Schmidt, Thomas, "Instrumentalism about Practical Reason: Not by Default", *Philosophical Explorations* 19 (2016): 17–27.

Schopenhauer, Arthur, "Preisschrift über die Grundlage der Moral", in *id.*, *Die beiden Grundprobleme der Ethik*, Frankfurt / Main: Joh. Christ. Herrmannsche Buchhandlung, 1841: 101–280.

Schroeder, Mark, *Slaves of the Passions*, Oxford: Oxford U.P., 2007.

Sharp, Frank Chapman, "Some Problems in the Psychology of Egoism and Altruism", *Journal of Philosophy* 20 (1923): 85–104.

Smith, Michael, "Deontological Moral Obligations and Non-Welfarist Agent-Relative Values", *Ratio* 24 (2011): 351–63.

Spohn, Wolfgang, "Wider Nash-Gleichgewichte", in *Handeln mit Bedeutung und Handeln mit Gewalt*, ed. by Christoph Fehige, Christoph Lumer, and Ulla Wessels, Paderborn: Mentis, 2009: 131–49.

Stemmer, Peter, *Normativität*, Berlin: Walter de Gruyter, 2008.

———, "Moral, moralisches Müssen und Sanktionen", *Deutsche Zeitschrift für Philosophie* 65 (2017): 621–56.

Stoutland, Frederick, "Responsive Action and the Belief-Desire Model", *Grazer Philosophische Studien* 61 (2001): 83–106.

Stroud, Sarah, "Permissible Partiality, Projects, and Plural Agency", in *Partiality and Impartiality*, ed. by Brian Feltham and John Cottingham, Oxford: Oxford U. P., 2010: 131-49.

Trapp, Rainer W., "The Potentialities and Limits of a Rational Justification of Norms", in *Preferences*, ed. by Ulla Wessels and Christoph Fehige, Berlin: Walter de Gruyter, 1998: 327–60.

Vaughn, Lewis (ed.), *Contemporary Moral Arguments*, (first ed. in 2010), second ed., Oxford: Oxford U.P., 2013.

Wessels, Ulla, *Die gute Samariterin*, Berlin: Walter de Gruyter, 2002.

———, *Das Gute*, Frankfurt a.M.: Vittorio Klostermann, 2011.

Woods, Jack, "The Authority of Formality", *Oxford Studies in Metaethics* 13 (2018): 207–29.

von Wright, Georg Henrik, *The Varieties of Goodness*, London: Routledge and Kegan Paul, 1963.