CANADIAN JOURNAL OF PHILOSOPHY
2018, VOL. 48, NO. 5, 812–830
https://doi.org/10.1080/00455091.2017.1415105

Routledge
Taylor & Francis Group

Integration and authority: rescuing the 'one thought too many' problem

Nicholas Smyth

Philosophy, Simon Fraser University, Burnaby, Canada

ABSTRACT

Four decades ago, Bernard Williams accused Kantian moral theory of providing agents with 'one thought too many'. The general consensus among contemporary Kantians is that this objection has been decisively answered. In this paper, I reconstruct the problem, showing that Williams was not principally concerned with how agents are to think in emergency situations, but rather with how moral theories are to be integrated into recognizably human lives. I show that various Kantian responses to Williams provide inadequate materials for solving this 'integration problem', and that they are correspondingly ill-positioned to account for the authority of morality, as Williams suspected all along.

ARTICLE HISTORY Received 28 May 2017; Accepted 5 December 2017

KEYWORDS Philosophy; morality; Kant; Bernard Williams; one thought too many; ethics

A resolute humanist, Bernard Williams refused to see moral philosophy as a purely theoretical enterprise. He was correspondingly critical of many attempts at moral theory, and he reserved special ire for philosophers whose theorizing became too detached from the exigencies of human social and psychological reality. This is the theme that unifies his disparate critiques and places him in a critical tradition which includes Hegel, Nietzsche and Freud. This tradition tends to ask difficult questions about the *authority* of moral philosophy. The basic concern is this: given that our practical perspective is that of a socially and psychologically situated human being, why do moral theorists have the *right* to shape our practical activity? As Williams put it in the introduction to *Moral Luck*,

> *By what right* does [moral theory] legislate to the moral sentiments? The abstract and schematic conceptions of 'rationality' which are usually deployed in this connection do not even look as though they were relevant to the question – so soon, at least, as morality is seen as something whose real existence must consist in personal experience and social institutions, not in sets of propositions. (Williams 1981b, I)

CONTACT Nicholas Smyth nick.a.smyth@gmail.com, nsmyth@sfu.ca

It was this general concern which lead Williams to articulate what has come to be called the 'One Thought Too Many Problem' (hereafter OTTMP), a problem which was said to infect impartial moral theories. In this paper I will clarify and refine his argument, after which I will survey various responses to it offered by Kantian moral theorists in particular. I will argue that Williams supplies the Kantian with two distinct tasks, which I label the *justification problem* and the *integration problem*. The justification problem concerns the ways in which Kant's moral theory might vindicate our right to preserve and promote our special ties to others, while the integration problem concerns the manner in which this justificatory story is to be integrated into the deliberative perspective of an actual human agent. This requirement is, I think particularly urgent for the Kantian theorist, and this is why I do not, in this paper, discuss consequentialist replies to Williams. Consequentialists are often tempted to simply deny that integration is important, arguing that their theory merely supplies a criterion of right action, and not a decision-procedure or a description of correct moral thought.[1] Kantians, on the other hand, ordinarily follow Kant himself in trying to say how their own criterion of rightness is to *show up* in the deliberative experience of a good moral agent. Neo-Kantian Barbara Herman is particularly clear on this point. In Kant's moral theory, she writes, we must 'find an account of how one is to integrate the requirements of morality into one's life' (Herman 1985, 193, See also Korsgaard 2009). In short, the Kantian definitely needs a solution to the integration problem, and in my view this renders them vulnerable to the OTTMP in a way that the consequentialist is not.

These preliminaries aside, I will now proceed to the main discussion. After outlining Williams' argument, I will describe the general solution to the justification problem offered by the Kantians who have most directly responded to the argument: Marcia Baron, Robert Louden and Herman herself. I will proceed to show that there is no corresponding solution to the integration problem, and I conclude that Williams' questions about the authority of Kantian moral theory remain unanswered.

Williams' argument

Williams' name is often raised in connection with a well-known thought experiment, and here I'll offer a slightly stylized version of the case. It features a pair of unfortunates who are drowning after a shipwreck. Luckily for one of the unfortunates, the sole potential rescuer – who of course can only save one person – is her husband, while her unlucky drowning counterpart has no relation to this potential rescuer at all. Assuming that their marriage is in good working order, she can quite naturally count on being rescued, since loving husbands tend to act in order to preserve the lives of their spouses. Indeed, we can imagine her panic subsiding as she treads water, safe in the knowledge that she will surely be rescued first.

https://doi.org/10.1080/00455091.2017.1415105 Published online by Cambridge University Press

However, suppose she thinks to herself: *Wait. Wasn't my husband just reading some mid-twentieth century Kantian moral philosophy during the buffet breakfast?* Her panic returns, augmented by a new, terrible thought, namely, the thought that her own husband might be the one who is going to allow her to drown in the name of impartial morality. Sure enough, she sees him remove a coin from his pocket and place it gingerly on the tip of his thumb. As the ocean continues to drag her towards its depths, she thinks, despairingly and perhaps correctly: *this man has never actually loved me*.

Of course, the case is farcical, since the husband's reaction to the situation is comically unrealistic and even morally disturbing. Yet, it is worth reminding ourselves that certain well-regarded moral philosophers had, in the 1960s and 1970s, theorized themselves into the conclusion that the coin-flipping husband's behaviour was perfectly acceptable. This is because their theoretical models very nearly implied that the husband ought to reflect on the situation in just the way that this robotically impartial person does. Indeed, we must recall that this is not Williams' case, rather, it is derived from one offered by Charles Fried at the end of his *An Anatomy of Values*. There, Fried had argued that a properly Kantian moral philosophy was committed to such principles of action as, '[t]he interests, preferences or desires of the agent have no special status or higher priority just because they belong to the agent,' and, 'the interests of no named party may be preferred *because* he is that named party' (Fried 1970, 111). Such principles, for Fried and the Kantian more generally, arise because moral value is grounded solely in *humanity*, in our capacity to set and pursue ends. Since this capacity is said by Kant to be identical in (virtually) all mature human agents, the rational agent must weigh the lives of strangers and loved ones equally – or so it was thought (Kant 2011, 4:429). Near the end of the book, Fried admits that his advocacy of such principles has left him with a problem, since, 'surely it would be absurd to insist that if a man could, at no risk or cost to himself, save one or two persons in equal peril, and one of those in peril was, say, his wife, he must treat both equally, perhaps by flipping a coin' (Fried 1970, 227).[2]

In response, Williams writes, 'the most striking feature of this passage is the direction in which Fried implicitly places the onus of proof: the fact that coin-flipping would be inappropriate raises some question to which an "answer" is required' (Williams 1981a, 17). In other words, Williams has questions about this 'question', and about the ways in which an impoverished model of character can make the question seem much more salient than it actually is (or could be). Surely, Williams suggests, if the marriage is at all typical, the man will simply save his wife without much of a thought for the *permissibility* of the action itself. Thoughts of permissibility seem so out of place because the dispositions that characterize a normal, loving human relationship are such that they will automatically prompt the rescue of one's beloved in such situations. The recognition that one's wife is drowning is, for Williams, all that any agent should be required to register at moments of this sort. This psychological phenomenon,

https://doi.org/10.1080/00455091.2017.1415105 Published online by Cambridge University Press

which Williams elsewhere labels *practical necessity*, is, for him, a central part of having a character at all. If you do not have any overwhelming disposition to act in certain ways in certain sorts of situations, then we are licensed to wonder whether you care about anything at all. Since caring about various things is part of what gives us our character, Williams concludes that it is Fried's thin and abstract model of character which has lead him to miss the ubiquity and significance of this sort of practical necessity.

However, Williams made two errors in his presentation of the problem. Here is the key passage, where he delivers what he takes to be the basic point:

> Surely this is a justification on behalf of the rescuer, that the person he chose to rescue was his wife? … the consideration that it was his wife is certainly, for instance, an explanation which should silence comment. But something more ambitious than this is usually intended, essentially involving the idea that moral principle can legitimate his preference, yielding the conclusion that in situations of this kind it is at least all right (morally permissible) to save one's wife … But this construction provides the agent with one thought too many: it might have been hoped by some (for instance, by his wife) that his motivating thought, fully spelled out, would be the thought that it was his wife, not that it was his wife and that in situations of this kind it is permissible to save one's wife. (Williams 1981a, 18)

Williams' first error lies in an excessive focus on how the rescuing husband is supposed to think and feel at the time of action. Though some of his defenders have sought to smooth over this fact, his references to the husband's 'motivating thought' make it very clear that Williams was specifically worried about how Kantian theory will require us to think and feel *when we are acting*.[3] And, as many commenters have subsequently argued, a moral theory need not ask agents to have thoughts of permissibility in mind when confronted with these sorts of situations (For example, see Louden 1992, 67). It is striking that *most* theories of practical activity recommend that we develop dispositions to respond instinctively in various types of situations.

Williams' second error is his apparent opposition to the thought that a theory can or should 'legitimate' the husband's saving his wife out of love. This is far too strong. It cannot be that the theoretical legitimation of spousal love is absurd *as such*. Indeed, it is hard to see how this sentiment is anything other than question-begging against moral theorists who want to tell us why *this* is a morally good husband. Moreover, we should remember that Williams himself was attracted to the view that one's deepest or most authentic desires provide powerful reasons for action, and this *itself* is a higher-order theory of rationality which legitimates the husband's preference.[4] So, it cannot be that the legitimation project is intrinsically mistaken. Rather, as I will now argue, OTTMP is a problem that is faced by specific moral theories: Williams must say that the Kantian encounters particular difficulties in trying to legitimate the husband's actions. Fortunately, I think that Williams supplies us with the beginning of an argument for precisely this conclusion, and I'll now start to say what that argument is.

https://doi.org/10.1080/00455091.2017.1415105 Published online by Cambridge University Press

Let's begin by distinguishing two distinct problems brought out by Fried's example. The first is the *justification problem*. This is the question of why the loving, rescuing husband is right to save his wife. Ideally, this should not merely involve establishing that he is *permitted* to rescue her, it also involves establishing that there are positive reasons in favor of his action. The second problem is the *integration problem*. To solve this problem is to say how our answer to the justification problem is to be integrated into the practical lives of real agents, how they are to see their own decisions in the terms it provides. In the end, I think that Williams was convinced that even if the Kantian can solve the justification problem, the integration problem will remain insoluble. An unsolved integration problem indeed leaves the agent with one thought too many: they will have simple, natural thoughts about the value or importance of their loving relationships, and those thoughts will clash with the solution to the justification problem, with the theoretical legitimation of the agent's preferences.

Before I move to my discussion of the two problems, a reader may be wondering why the integration problem demands a solution. This, I think, derives from a basic concern about the authority of moral theory. Williams wanted us to have something to say to the agent who is being asked to structure their practical life around the considerations provided by such a theory. By way of illustration, we might see this as a basically *Hobbesian* sort of demand. Before Hobbes, it was not commonly thought that states had to relate their own justifications for power to the practical reasons possessed by of each of their subjects. For example, according to one historically influential story, monarchical political structures are derived from the moral-metaphysical structure of the patriarchal family, which in turn was said to mirror the moral-metaphysical structure of the universe itself (i.e., with God as the foundational patriarch).[5] Agents who did not subscribe to the particular religious worldview deployed in this type of story were often thought to be simply mistaken or defective. Hobbes, by contrast, tried to show that basic motivations possessed by *any* human being – for example, desires for survival and security – were best served by the existence of a powerful central authority. Thus, Glen Newey argues that 'a life without government is not worth living' is the central message of the *Leviathan* (Newey 2008, 1). The revolutionary presupposition here is that facts about what makes life worth living for all human individuals are even *relevant* to the authority of rulers over their subjects.

Williams was convinced that moral theory needs just this sort of story, and as I have already suggested, most contemporary Kantians share this idea. It's not hard to see why: for Kant, the authority of morality derives from the fact that it is *self-addressed*, that it is already nascent in the practical consciousness of all human beings (Kant 2011, 4:431–435). Thus, the Kantian must show how the answer to what I am calling the *justification problem* might harmonize with the actual motivations and dispositions characteristic of a liveable human life (those who find talk of 'harmonization' infuriatingly vague will find a more complete

https://doi.org/10.1080/00455091.2017.1415105 Published online by Cambridge University Press

account in what follows). This is the problem of integration, and Williams would later conclude that it could not be solved by any ethical theory:

> My own view is that no ethical theory can render a coherent account of its own relation to practice: it will always run into some version of the fundamental difficulty that the practice of life, and hence also an adequate theory of that practice, will require the recognition of what I have called deep dispositions; but at the same time the abstract and impersonal view that is required if the theory is to be genuinely a *theory* cannot be satisfactorily understood in relation to the depth and necessity of those dispositions. (Williams 2009, 295–296)

As we will see, many philosophers believe that Kant's system *does* have such an account, and I shall come to their arguments shortly. However, the point is that we must not lose sight of the fact that OTTMP is best seen as embedded in a larger set of concerns about integration and authority, concerns that animated much of Williams' work. Having laid out the issue as I see it, I will now begin by outlining the ways in which Louden and Baron establish two important theses that are required for a Kantian solution to the justification problem.

The justification problem

Since Kantian morality is generally well-understood, I will here merely assemble a list of relevant theses which Kant and his modern proponents share. *Qua* deontological moral theory, Kant's theory supplies us with *duties* that we must fulfill. He believed that all of our particular moral duties can be traced, in some way, to a single *Ur*-obligation, which is to obey the Categorical Imperative. Moreover, our particular duties come in several varieties; for my purposes here, I will only note that for Kant, we have a *direct* duty to perform some action when it is the only way to avoid violating the Categorical Imperative (as was allegedly the case with truth-telling), and we have an *indirect* duty to perform some type of action or to enact some general policy when it aids us in pursuing our direct duties (such as maintaining our physical health).

So, how might someone operating under this basic paradigm respond to Williams' suggestion that they cannot make sense of the husband's rather obvious right to save his wife without thinking? The solution offered by Baron and Louden involves the conjunction of two theses, neither of which is sufficient, on its own, to solve the problem:

> The *Permissibility Thesis*: In situations relevantly similar to Fried's, Kant's moral theory *permits* the rescuing husband to assign special priority to his wife.

> The *Indirect Duty Thesis*: Kant's moral theory establishes an *indirect duty* to form and maintain loving relationships such as the one that exists between the rescuing husband and his wife in Fried's scenario.

https://doi.org/10.1080/00455091.2017.1415105 Published online by Cambridge University Press

If only the second thesis were true, then we would not have ruled out the possibility of a direct duty to act impartially (for example, to decide by flipping a coin), and since Kant's indirect duties have no weight when they conflict with direct duties, we could not conclude that the rescuing husband was justified in automatically saving his wife. Conversely, if only the first thesis were true, it would be a hollow victory for Kantian morality, since we would not be able to make much sense of why the husband is *positively justified* in saving his wife, and not merely permitted to do so. I will now turn to the ways in which Baron and Louden defend these two theses.

The permissibility thesis

Baron argues that there is nothing particularly wrong, from a Kantian point of view, with according a certain priority to particular persons in certain kinds of situations. Consider the maxim: *in certain dire situations, provide special care and devote significant amounts of energy or resources to your loved ones.* It is both coherent and possible for an agent to will that this should become a universal law. There is no contradiction, either in conception or in willing, as Kant himself explicitly notes (*MM* 6:452). If the maxim directed agents to devote every ounce of time and energy they have into promoting the welfare of certain 'significant others', it might generate a contradiction in willing, but no such maxim is required in order for the rescuing husband to save his wife (Baron 2008, 255).

It is worth noting that neither Baron nor Louden seek to establish a stronger thesis to the effect that saving one's spouse in such situations is *required*. It is perfectly possible and feasible to will a *contrary* of the maxim just cited, one that directs agents to treat all persons impartially in extreme situations of this sort. That is to say, the robotic, coin-flipping husband may be somewhat distasteful, but he is not, speaking in a strict sense, violating Kantian morality. If he had some *generalized* maxim of impartiality, he might risk a contradiction in willing, since it is a requirement of human society that its members be raised by those who pay special care and attention to their well-being. But the robotic husband need not operate under any such maxim, he need only think that in certain difficult situations it is best to remain wholly impartial. So, while we cannot establish a strong or unconditional *requirement* to rescue one's loved ones, the permissibility thesis is comparatively easy to establish, and for brevity's sake I will not discuss it further. Instead, I will move on to the more challenging demonstration offered by Baron and Louden.

The indirect duty thesis: Baron and Louden

Mere permissibility aside, can a Kantian moral theorist make sense of our *positive* reasons to protect and cherish those we love? Baron certainly thinks so. Kant, she writes, 'recognizes that we have special duties to particular others. Why would anyone think otherwise (Baron 2008, 252)?'[6] In a footnote, she elaborates: '[t]he

https://doi.org/10.1080/00455091.2017.1415105 Published online by Cambridge University Press

special duties he recognizes are duties to friends; one surmises that if he had a section in the *Tugendlehre* on familial relationships, as he does on friendship, he would recognize special duties to family, as well' (Baron 2008, fn 13). However, she admits that Kant says very little about this issue, and this is problematic, because we are now asking what *grounds* Kant can have for affirming positive moral duties of care towards particular others.[7] It seems reasonably clear that these can only be *indirect* duties. If this isn't clear, consider the following argument:

(1) Kantian duties are either direct or indirect.[8]
(2) Direct duties are those which *must* be performed in order to avoid violating the Categorical Imperative, while indirect duties are only required as means to a final end: the more efficient or reliable performance of our direct duties.
(3) Since we do not *violate* the CI by remaining perfectly impartial in rescue cases, acts of partial care cannot be direct duties.[9]
(4) Therefore, duties of partiality, if they exist, must be indirect duties.
(5) It follows that acts of partial care can only be positively justified, on the Kantian picture, as a means to our own direct moral duties.

It is no accident that Louden's defense of Kant affirms precisely this conclusion. His own reply to the OTTMP begins with the suggestion that Kant has a very easy time accounting for the priority we assign to non-moral projects such as personal relationships. He writes,

> Such criticisms [as OTTMP] lose their force when an act conception is replaced by a broader agent conception; for the latter requires us to ask how a person's life in general and overall is going. Someone whose nonmoral personality is empty or who typically carries around excessive cognitive baggage in situations that will not tolerate it cannot be said to have a good life, for such a person is going to fail morally in too many cases. (Louden 1992, 33)

Thus, anyone who is as mechanically detached from his relationships as the coin-flipping husband would not be living well. So far, so good, but notice that for Louden, 'living well' is living a life *that does not fail morally*. He places great emphasis on the fact that non-moral projects (such as personal attachments, careers and hobbies) will provide human beings with the psychological well-being they need to conform to their direct duties. Thus, Louden concludes, we must have indirect duties to form and maintain such projects, and we therefore have strong positive practical reasons to do so.[10]

Since we now know why the rescuing husband is positively justified in saving his wife, we are in possession of a solution to what I have called the justification problem. However, we must now proceed to the integration problem, which, I think, is the *real* issue, here. As Louden's emphasis on moral failure suggests, Kant's model of personal relations is heavily *moralized*, in the sense that it aims at a sympathetic union between good wills.[11] Since the robotic, coin-flipping version of the rescuing husband does not violate the categorical imperative,

https://doi.org/10.1080/00455091.2017.1415105 Published online by Cambridge University Press

the value of particular acts of care cannot be explained by their being strictly *required*, or by their being *direct* duties. This leaves only the purely instrumental category of indirect duties, and this is why Louden is forced to admit that Kant can only have an instrumental account of the value of human relationships. They are valuable *only* insofar as they are (in Kant's words) 'cultivator[s] of virtue and a preparation for its surer practise (*LE* 27:420).'

It is here that the integration problem begins to loom. We began with the permissibility thesis in order to pave the way for a broader understanding of the positive value of partial care. We aimed to secure this further understanding via the indirect duty thesis, which claims that caring for particular others is required because it makes us better moral agents. Now, it might seem as though we now have a complete story to offer the loving, rescuing husband (or anyone else, for that matter) about why Kantian morality does not interfere with the normal functioning of human relationships. However, as Williams would have asked, can we *live* with these ideas? That is to say, given what we know about human beings and about how their relations with one another are structured, should we expect agents to incorporate this justificatory model into their practical lives? Do the considerations that we have marshalled in order to secure the permissibility and indirect duty theses sit well with the ways in which we can expect the rescuing husband to view his life? I now turn to a fuller description of this requirement, after which I will examine Herman's attempt to show that Kant's theory meets it.

The integration problem

What exactly does integration require? What does it mean for theoretical considerations to *harmonize* with the way an agent sees his or her practical situation? Here, it will help to consider other sorts of practical activity which might be informed by theory. First, consider an airline pilot preparing for takeoff. The pilot's overall goal is to become safely airborne, and this requires the completion of several sub-tasks. He or she is expected to run down a checklist, checking wind conditions and fuel levels, reviewing the flight plan aloud, and so on. However, for special reasons, it is *not* desirable for the pilot to internalize these directives and act on them 'unthinkingly'. Because the cost of failure is so high, the currently accepted theory of safe flight not only specifies various items on the checklist, it also contains an explicit imperative: when reviewing the flight for takeoff, *never* act unthinkingly or instinctually, *always* run over the checklist consciously and deliberately.

By contrast, consider tennis. A theory of how to play tennis begins with the final end of winning tennis matches, which naturally breaks down into discrete sub-tasks. Yet, here, things are more subtle than they are for taxiing airline pilots, since a tennis player who tried, during a match, to consciously apply the theoretical lessons they had learned would probably lose. What matters during the game is their ability to play, and not whether they *think* correctly about playing.[12]

https://doi.org/10.1080/00455091.2017.1415105 Published online by Cambridge University Press

Training thus involves the inculcation of various physical and perceptual dispositions that enable the player to respond in the way that they ought to, given the dictates of tennis theory and the overarching goal of winning. For example, a player may simply *go to the net* at a certain point in the match without exercising reflective control over this disposition, and without calling to mind any of the more general theoretical considerations that justify their going to the net. So, while self-conscious application of tennis theory during a match would involve one, indeed *several* thoughts too many, surely no one would suggest that this renders the authority of the theory problematic. This shows that a genuinely powerful version of the OTTMP cannot just involve the accusation that Kant's theory forces us to *think* too much. But what is the problem supposed to be?

It is, I claim, fundamentally about the impossibility of integration. For reasons of clarity, let me explicitly define the terminology which can help us to see just what integration amounts to:

> **Instrumental End**: A sub-end which is necessary (or important) for achieving a final end, given an agent's practical context.
>
> **Final End**: The end which supplies instrumental ends with the normative importance or significance.
>
> **Motivating Thought**: The agent's sense of what considerations justify an action they are currently performing.
>
> **Theory**: A body of propositions which specifies final and instrumental ends for a given activity.

A theory is *integrated* with its activity when the Final Ends it specifies are identical to (or perhaps very similar to) at least one of the primary motivating thoughts that agents can be reasonably expected to have when they pursue the Instrumental Ends which (the theory claims) are necessary or important for the proper pursuit of the activity. If that test seems like a mouthful, just consider the way in which tennis theory easily passes it. When a good tennis player acts on the internalized disposition to *go to the net*, she is pursuing the proper Instrumental End that is specified by the theory. Moreover, that end is said, by the theory itself, to be justified by the Final End of winning the match. And since tennis players are constantly trying to win while playing, this is precisely the end that the player can be expected to have at the forefront of their mind as they go to the net (it need not be the *only* aim they have in mind, but this is not what integration requires). In other words, integration prevents a certain sort of evaluative fragmentation: an agent is not expected to see the value of their action in two distinct or contrary ways at various points in their life.

Put another way, in tennis theory, the reasons that one has for inculcating various dispositions in training are the same as the reasons one has when one automatically expresses those dispositions in an actual game. The question of the theory's *authority* over one's tennis-related practical activity has a neat answer: since you will want to win when you are playing, you had better learn

https://doi.org/10.1080/00455091.2017.1415105 Published online by Cambridge University Press

to do as the theory tells you. So it is with a huge number of practical theories; the reasons you will have for doing as the theory instructs are the same reasons you have when you learn and internalize the theory. Such theory has a simple solution to the integration problem, since there is no *extra* justifying thought, no mismatch between the final values, goals or ends which justify instrumental ends and those which the agent will have in mind when acting as the theory requires. The integration problem for a *moral* theory, then, is this: can the final ends specified by the theory harmonize with the goals or aims present in agent's motivating thoughts at the time of action?

Herman's account of integration

In defending Kant's ethics from Williams, Herman shows an admirable sensitivity to the ways in which Kant's theory must be made to square with social and psychological reality. As we have already seen, she believes that a Kantian theorist must provide an account of how a moral agent is supposed to incorporate Kant's system into his or her practical activity. And we now know that Kant's theory both permits the rescuing husband to save his wife and supplies us with an account of why it is good that he does so. How are these ideas supposed to regulate the practical lives of real agents?

Herman's account centrally involves a distinction between *primary* and *secondary* motives. For Herman, a motive is best described as 'the way [the agent] takes the object of his action to be good, and hence reason-giving' (Herman 1985, 36). Furthermore, a primary motive supplies the agent directly with the motivation to do some act (Williams, remember, called this a *motivating thought*), whereas a secondary motive merely provides limiting conditions on what may be done. Secondary motives serve a 'regulative' function, and that they operate 'in the background', whereas primary motives are capable of producing action all on their own, and are prominent in the phenomenology of agency (Herman 1985, 35).[13] As an example of a secondary motive, Herman argues that the motive of *economy* is often merely regulative: rather than directly move an agent to action, a commitment to economy constrains the set of actions that an agent is prepared to consider.

Her solution to the integration problem is this: being *motivated* by the categorical imperative does not imply that the husband has to have this principle as his particular consciously willed *end*. This is because it need only show up in his *secondary* motive. The husband's *primary* motive, however, will be the love he has for his wife. Since he has previously acted on his (indirect) duty to form and maintain loving relationships, the husband will naturally save his wife out of love. In Herman's words, this is a situation where 'being a moral person involves the recognition of the limits of the moral: when moral reasons are not the appropriate reasons to act on' (Herman 1985, 42). Thus, the moral law need only play a regulative role in the rescue case, such that it is counterfactually true

https://doi.org/10.1080/00455091.2017.1415105 Published online by Cambridge University Press

of the rescuing husband that he would not have immediately jumped in to save his wife *if* the corresponding maxim were not universalizable.[14] This regulative motive should (in this case) be quite far from the husband's consciousness, but this does not mean that it is not playing an important role in his moral agency. To put this solution in the terms of a distinction emphasized by Onora O'Neill: the agent's subjective *motive* is to save his wife, but this does not at all preclude the possibility that his action has the *form* of a universal law, or that, in Kant's own words, his maxim's moral value 'lies objectively in the rule and the form of universality, which makes it capable of being a law' (Kant 2011, 4:431. See O'Neill 1989, 132).

It is now worth asking whether we have a solution to the integration problem. As it turns out, according to the neo-Kantians we are surveying, morality is more like tennis than takeoff, since many dispositions we express in performing permissible and obligatory actions will be justified only by reference to a theory that is quite far from the agent's mind at the time of action. At the level of the primary motive, it is clear that the rescuing husband is *not* meant to think of his acting from love as contributing to moral ends: this is just what Herman means when she says that he may act for non-moral reasons. This seems to neatly describe how an admirable husband thinks and feels *at the moment of rescue*, and to that extent, it represents progress on the integration problem. But our deliberative lives contain many more moments than this, since we also stop and think about the normative significance of our motives and actions after the fact. Once his wife is safe and dry at home – and the poor stranger's funeral is over – can the husband reflect on the newly refined version of Kant's theory and feel as though he has a coherent understanding of why this theory ought to legislate to his sentiments?

As the case of tennis theory showed, if this is at all possible, it must be the case that the goals or aims which drive the rescuing husband at the time of action harmonize with the Final End(s) that are specified by the theory. Here, the *indirect duty thesis* makes serious trouble. Williams reminded us that, in situations where someone we care deeply about is seriously threatened, our motivating thought will simply be to save or aid *that person*. That, in Herman's terms, will be our primary motive. Yet, according to the indirect duty thesis, we ought to inculcate this type of motivating thought *in order to make ourselves more able to perform our direct moral duties*. Remember, this is just what 'indirect duty' means for Kant: an action which has only instrumental value in virtue of contributing to our capacity to conform to our direct duties (Timmermann 2009, 36).

On this account, it is good that the husband has developed a disposition to love his wife, because he is thereby better able to perform his direct duties: to avoid lying, stealing, murdering. Yet, these final ends are quite different from any motivating thoughts that the husband in Fried's case will certainly have in mind, and this opens up reflective space for questions about authority. Even if the husband accepts the Kantian theory, he must confront in himself the

https://doi.org/10.1080/00455091.2017.1415105 Published online by Cambridge University Press

disposition to reject it, to feel as though certain actions have unqualified significance, a significance which is not explained in terms of the action's contributing to the formation of the Kingdom of Ends. Moreover, Williams would caution us against a moralistic mistake here, which is to think that this is due to some weakness, self-centeredness or failure of imagination on the husband's part. The theory *itself* is directing us to form loving relationships with particular persons, and a necessary condition on forming such relationships is to have those very motivating thoughts.[15] These reflective questions are not being asked by some Calliclean immoralist who stands outside of the moral system and sneers at it; they are being asked by an agent whose actions and dispositions are (according to the various accounts provided by Baron, Louden and Herman) *precisely* as they should be.

Thus, even within the confines of a sophisticated Kantian theory, there is a deep mismatch between the rescuing husband's sense of justification and the aims and goals which ultimately permit him to feel justified.[16] This is just what Williams meant when he claimed that 'such things as deep attachments to other persons will express themselves in the world in ways which cannot at the same time embody the impartial view, and they also run the risk of offending against it' (Williams 1981b, 18). In short, the integration problem remains unsolved, and I conclude that the most sophisticated versions of Kantian theory on offer *do* require one thought too many. As Michael Stocker would say, we are meant to embody a 'schizophrenic' attitude towards these same commitments, at one moment seeing them as grounded in (and constrained by) the moral law, at another moment seeing them as possessing self-standing normative significance.

Importantly, though, we do not need to follow Stocker in claiming that this schizophrenia is unhealthy, or 'a malady of the soul' (Stocker 1976). Nor do we need to follow many readers of Williams who mistakenly identify this as a worry about 'demandingness'.[17] Rather, we need only see that it re-poses the Hobbesian question of authority, the same question which greatly vexed Williams. The tennis player who asked this question about his practical theory received a tidy answer: your aim is to win, so do and think as the theory tells you, both in the moment and when you are training for the match. The Kantians surveyed in this paper have no such answer to offer the rescuing husband, who may well be troubled by the clash between the justificatory story and the sense of justification he feels in the moment of action. He might ask: 'which of these two psychological moments is to have priority in determining the *true* significance of my loving commitments?' And can there be an argument that shows him that it would necessarily be a mistake to think that the *partial* moment reveals the whole (decidedly non-Kantian) truth about what he values? That is the 'One Thought Too Many' problem. If Kant is right about morality, each of us is doomed to an evaluative fragmentation that leaves questions about the authority of morality unanswered.[18,19]

https://doi.org/10.1080/00455091.2017.1415105 Published online by Cambridge University Press

Now, lest anyone think the integration bar is being set impossibly high, it is worth pointing out that there are moral theories that have a much easier time clearing it. Consider W.D. Ross' pluralist theory of prima facie duties. On this model, ordinary moral consciousness is directly aware of several distinct duties, each of which generates strong yet defeasible practical reasons. When these duties conflict, there is nothing much that can be said about how to systematically resolve the conflict; this is left to the refined judgment of good sense. The model is messy and relatively unsystematic, and might not really count as a moral *theory* in any strict sense, but this is because it mirrors, more or less, the messiness and unsystematicity of ordinary morality. It is open to Ross to say, as he basically does, that the rescuing husband shows good sense in privileging his special obligations over his more general duty of beneficence.[20] Furthermore, Ross can claim, with great plausibility, that the overwhelming importance of the rescuing husband's special obligations *defeats* the reasons generated by his more general duty of beneficence, rendering them null and void. He need not say that the special obligations are only justified inasmuch as they promote the observance of any other duties whatsoever. This, I think, is a story that might be nicely reflected (in perhaps a less sophisticated form) in the husband's state of mind at the time of action, and so Ross's theory is comparatively well-positioned with respect to the integration problem.

By contrast, Kant's belief that there is a *single* principle that can (both in theory and in practice) uniquely determine the content of all our moral obligations creates enormous difficulty when cases like Fried's arise. This is so even if we follow most contemporary Kantians in embracing the more humanistic *Doctrine of Virtue* as our best guide to Kant's ethics. Williams' worries about integration – and, by extension, about authority – remain, because there is a fundamental mismatch between the values that guide actual decision-making and the values that are supposed to justify our actions.

Finally, we should ask what happens when the Kantian adopts a defensive tactic hinted at earlier. Perhaps the Kantian, qua *moral* theorist, need only rest easy with the *permissibility thesis*, with the claim that the rescuing husband does not violate the categorical imperative. We do not, on this view, have any moral *duties* whatsoever with respect to our nearest and dearest; acts of love have only non-moral value, so long as they are permissible. This Kantian would not need to pursue any defence of the *indirect duty thesis*, and there would be no complex justificatory story that might conflict with the sense of justification possessed by agents who act out of love. For what it's worth, I think that this is the most promising route for the Kantian, but it faces several independent problems. First, it is doubtful that Kant would have taken it: as Baron notes, Kant is clear that we have positive *moral* duties towards our close friends, indirect duties which derive their normative force from the categorical imperative.[21] Second, it is highly counterintuitive to say that the rescuing husband's action has no positive *moral* content, and as generations of critics have charged, any theory which

https://doi.org/10.1080/00455091.2017.1415105 Published online by Cambridge University Press

forces us to deny that bonds of family and friendship are the source of moral reasons seems woefully impoverished (Held 2007; Noddings 1998; Slote 2013; Williams 1972). Finally, this sharp distinction between moral and non-moral ends might actually make it *harder* to answer questions about the authority of Kantian ethics. The indirect duty thesis, as described and defended by Baron and Louden, at least had the virtue of giving us something to say to the loving husband about the value of their personal commitment. It turned out that we were saying the wrong *sort* of thing, but at least we were saying *something*. According to this proposal, morality has nothing at all to say about the practical significance of love, personal commitment or other 'non-moral' projects. It was Williams' view that this sort of maneuver would make questions about the authority of morality more difficult to answer, since the conflicts between the two spheres of value are not only possible but almost inevitable (see 'Moral Luck' in Williams 1981b, 19–39).

Conclusion

Baron, Louden and Herman are each engaged in an important reconciliation project, one that aims to bring Kant's theory closer to common moral thought. This project seeks to go beyond the rigid formalism of mid-twentieth century Kantian scholarship (as represented by Fried) and to embrace the more humanistic picture of Kant that arises from his later writings. The relevance of this work to the 'One Thought Too Many' problem is now clear. In establishing the *permissibility* and *indirect duty* theses, they have shown how Kantian moral theory can make sense of the importance of human relationships, and they have provided a solution to the justification problem. However, while this solution to the justification problem is broadly plausible, I have suggested that it does not give us the resources we need to solve the integration problem. There remains a critical incongruence between the justificatory story itself and the motivations of the loving, rescuing husband. As I have argued, *this* is the problem that really worried Williams, because it lead directly to questions about the authority of moral theory, about its right to legislate to our sentiments. Once we are prepared to admit that there are practically significant situations in which our natural sense of justification conflicts with the justifications provided by a moral theory, we are left with inevitable questions about why we should think of moral laws as having the ubiquitous authority which Kant and Kantian alike are united in thinking that they do.

Notes

1. See, for example, Lazari-Radek and Singer (2016). To the extent that a consequentialist avoids this route, however, she does render herself vulnerable to OTTMP.

https://doi.org/10.1080/00455091.2017.1415105 Published online by Cambridge University Press

2. For a classic treatment of this sort of case from a non-Kantian perspective, see Taurek (1977).
3. For a subtle and creative attempt to direct our attention away from this, see Wolf (2012).
4. The famous Internal Reasons Thesis (Williams 1981b, 101–113) does not involve the claim that desires are *sufficient* for practical reasons. However, Williams was inclined to say things like 'desiring to do something is of course a reason for doing it' (Williams 1985, 19).
5. The *locus classicus* here is Robert Filmer's *Patriarcha*. See Filmer and Sommerville (1991).
6. Now, this is not *quite* fair. As we have seen, Charles Fried seems to have thought otherwise, on basically Kantian grounds. Kantians in Fried's day tended to ignore the later *Metaphysics of Morals,* and while we may fault them for this, one can hardly blame them for deriving strict impartialist principles from the *Groundwork*, where Kant emphatically denies that contingent empirical attachments can form the basis of morally worthy action. See Kant (2011, 4:426.3).

   In this paper I use the following abbreviations for Kant's work: G = *Groundwork for the Metaphysics of Morals*, CPrR = *A Critique of Practical Reason*, MM = *The Metaphysics of Morals*, LE = *Lectures on Ethics*. See table of abbreviations at the end for full bibliographic information (Kant 1998).
7. 'Those looking for explicit indications from Kant as to how much one must do for strangers compared to how much one should do for acquaintances, and how much for those one loves, will be disappointed. In general, Kant offers little by way of guidelines for deciding whom to help and how. We take it that this is not an oversight, but simply something on which he does not believe that people need him, or other ethicists, to provide advice or direction' (Baron and Fahmy, 223).
8. Kant's discussions of this distinction are at Kant (G 4:399 and MM 6:388). For a longer discussion, see (Timmermann 2006).
9. Once again, no Kantian theorist has argued that it is *wrong* to save the stranger in Fried's case. A world full of such agents might be distasteful to us, but it is hard to argue that it is either logically impossible or that it necessarily involves a contradiction in willing.
10. This account bears a striking resemblance to Philip Pettit's *standby consequentialism*, which directs us to avoid explicitly consequentialist reasoning most of the time. We are, on this view, to activate such reasoning only when contextual features of our situation alert us to the possibility that we ought to do so (Pettit 2015). I thank an anonymous reviewer for pointing this out.
11. This is why, in the passages on friendship cited by Baron, Kant writes that 'friendship cannot be a union aimed at mutual advantage, but must rather be a purely moral one' Kant (1996, 6:470).
12. It might be thought that this is only due to contingent facts about our limited ability to process information, and it might be concluded that there is no real distinction between piloting a plane and playing tennis in this respect. However, for my purposes here, this does not matter, since I am merely trying to show that *even if* a theory cannot be consciously applied, it may still be integrated into our activities in a way that preserves the theory's authority over that activity.
13. This combination of characteristics might be unstable, since motives of which we are not conscious are plainly capable of directly moving us to action, as social psychologists are forever reminding us. However, we might admit that there is a fairly intelligible distinction between merely regulative and effective motives.

https://doi.org/10.1080/00455091.2017.1415105 Published online by Cambridge University Press

14. It's worth noting that Baron is in full agreement on this basic point (Harman 1999, 121–128).
15. David Velleman's 'Love as a Moral Emotion' is, I take it, an attempt to deny these observations about love. He claims that what we really perceive in experiencing the emotion of love is our beloved's 'rational self-governing will'. I join many others, however, in finding such claims basically implausible. For philosophical criticism, see Jeanette Kennett (2008) and Edward Harcourt (2009). For empirical evidence against Velleman's developmental-psychological story, see M. L. Hoffman (1990) and J. G. Smetana and J. L. Braeges (2000).
16. This is not merely because the husband lacks the time to deliberate on these justificatory considerations, for this result obtains even if we give him a little more time, say, by giving him sixty seconds to decide who he will save. This husband will spend those sixty seconds in anguish, continuing to feel as though he must save his wife *because* she is his wife.
17. These are very common readings of Williams' objection (see Scherkoske 2013). For my part, I have never been able to see why such objections should have force against Kantian ethics, which is resolutely non-eudaimonist. Kant is absolutely clear on this point (MM 6:331, 6:378), and so it is unclear why the fact that morality reduces one's personal flourishing (or indeed that it is *difficult* or *alienating* or a *malady of the soul*) should be relevant to the Kantian enterprise. If this reduction in personal flourishing were such that it significantly undermined one's ability to carry out one's moral duties, then Kant would surely be concerned (see Kant 2011, 4:399), but neither Stocker nor the proponents of the 'demandingness' objection have shown anything quite so strong as *that*.
18. A closely related argument against consequentialism is offered by Paul Hurley in *Beyond Consequentialism*. See Hurley (2009).
19. I suspect that at this stage, some Kantian theorists will be tempted to declare that the Categorical Imperative is the product of *reason*, whereas one's valuation of a loved one is the product of mere *emotion*. But the task of saying what is actually meant by such descriptions – of how we might go about verifying that they are true – is fantastically difficult. Suppose I claim that the opposite is true, or that the fully impartial agent is moved by some unconscious fear or shame, whereas the partial agent is moved by an immediate rational perception of evaluative reality (Arpaly 2003, 20). Who is right about rationality, and how are we to decide?
20. Ross is clear that we *do* have special obligations to those that are close to us, and he does not attempt to derive this obligation from any more basic duty. He claims that I have such basic obligations towards those who 'stand to me in the relation of promisee to promiser, of creditor to debtor, of wife to husband, of child to parent, of friend to friend, of fellow countryman to fellow countryman, and the like' (Ross 2002, 19).
21. He also directly states that we have duties of love to our parents at Kant (1996, 6:390).

Notes on contributors

Nicholas Smyth is a postdoctoral researcher at Simon Fraser University. He works in Moral Philosophy, focusing on issues in moral epistemology, meta-ethics and moral psychology. For related work, see 'The Inevitability of Inauthenticity', forthcoming in the collected volume, *Ethics Beyond the Limits* (Routledge).

https://doi.org/10.1080/00455091.2017.1415105 Published online by Cambridge University Press

References

Arpaly, Nomy. 2003. *Unprincipled Virtue: An Inquiry Into Moral Agency*. Oxford: Oxford University Press.

Baron, Marcia. 2008. "Virtue Ethics, Kantian Ethics, and the 'One Thought Too Many' Objection." *Kant's Ethics of Virtue* 1 (2): 245–277.

Filmer, Robert, and Johann P. Sommerville. 1991. *Sir Robert Filmer: Patriarcha and Other Writings*. Cambridge, MA: Cambridge University Press.

Fried, Charles. 1970. *An Anatomy of Values*. Cambridge, MA: Cambridge University Press.

Harcourt, Edward. 2009. "Velleman on Love and Ideals of Rational Humanity." *The Philosophical Quarterly* 59 (235): 349–356.

Harman, Gilbert. 1999. "XIV-Moral Philosophy Meets Social Psychology: Virtue Ethics and the Fundamental Attribution Error." *Proceedings of the Aristotelian Society* 99 (3): 315–331.

Held, Virginia. 2007. "Feminism and Moral Theory." *Bioethics :An Introduction to the History, Methods, and Practice*, edited by Nancy Jecker and Albet Jonsen, 158–163. Sudbury: Jones and Bartlett.

Herman, Barbara. 1985. "The Practice of Moral Judgment." *The Journal of Philosophy* 82 (8): 414–436.

Hoffman, M. L. 1990. *Empathy and Moral Development: Implications for Caring and Justice*. Cambridge: Cambridge University Press.

Hurley, Paul. 2009. *Beyond Consequentialism*. Oxford: Oxford University Press.

Kant, Immanuel. 1996. *Kant: The Metaphysics of Morals*. Cambridge, MA: Cambridge University Press.

Kant, Immanuel. 1998. *Critique of Pure Reason*, edited by Paul Guyer. Cambridge, MA: Cambridge University Press.

Kant, Immanuel. 2011. *Immanuel Kant: Groundwork of the Metaphysics of Morals: A German–English edition*. Cambridge, MA: Cambridge University Press.

Kennett, Jeanette. 2008. "True and Proper Selves: Velleman on Love." *Ethics* 118 (2): 213–227.

Korsgaard, Christine M. 2009. *Self-Constitution: Agency, Identity, and Integrity*. Oxford: Oxford University Press.

Lazari-Radek, Katarzyna de, and Peter Singer. 2016. *The Point of View of the Universe: Sidgwick and Contemporary Ethics*. Oxford: Oxford University Press.

Louden, Robert B. 1992. *Morality and Moral Theory: A Reappraisal and Reaffirmation*. Oxford: Oxford University Press.

Newey, Glen. 2008. *Routledge Philosophy Guidebook to Hobbes and Leviathan*. New York, NY: Routledge.

Noddings, Nel. 1998. "Thinking, Feeling, and Moral Imagination." *Midwest Studies in Philosophy* 22 (1): 135–145.

O'Neill, Onora. 1989. *Constructions of Reason: Explorations of Kant's Practical Philosophy*. 41 vols. Cambridge: Cambridge University Press.

Pettit, Phillip. 2015. "The Inescapability of Consequentialism." *Luck, Value, and Commitment: Themes from the Ethics of Bernard Williams*, edited by Ulrike Heuer and Gerald Lang, 41–70. Oxford: Oxford University Press.

Ross, W. D. 2002. *The Right and the Good*. Oxford: Oxford University Press.

Scherkoske, Greg. 2013. "Whither Integrity II: Integrity and Impartial Morality." *Philosophy Compass* 8 (1): 40–52.

Slote, Michael A. 2013. *From Enlightenment to Receptivity: Rethinking Our Values*. Oxford: Oxford University Press.

https://doi.org/10.1080/00455091.2017.1415105 Published online by Cambridge University Press

Smetana, J. G., and J. L. Braeges. 2000. "The Development of Toddlers' Moral and Conventional Judgments." *Merrill-Palmer Quarterly* 36: 329–346.
Stocker, Michael. 1976. "The Schizophrenia of Modern Ethical Theories." *Journal of Philosophy* 73 (14): 453–466.
Taurek, John M. 1977. "Should the Numbers count?." *Philosophy & Public Affairs* 6 (4): 293–316.
Timmermann, Jens. 2006. "Kant on Conscience, "Indirect" Duty, and Moral Error." *International Philosophical Quarterly* 46 (3): 293–308.
Timmermann, Jens. 2009. *Kant's' Groundwork of the Metaphysics of Morals': A Critical Guide*. Cambridge: Cambridge University Press.
Williams, Bernard Arthur Owen. 1981b. *Moral Luck: Philosophical Papers, 1973-1980*. Cambridge: Cambridge University Press.
Williams, Bernard. 1972. "Morality and the Emotions." In *Problems of the Self*, 207–229. Cambridge: Cambridge University Press.
Williams, Bernard. 1981a. "Persons, Character, and Morality." In *Moral Luck*, edited by James Rachels, 1–19. Cambridge: Cambridge University Press.
Williams, Bernard. 1985. *Ethics and the Limits of Philosophy*. London: Fontana.
Williams, Bernard. 2009. "The Point of View of the Universe: Sidgwick and the Ambitions of Ethics." In *The Sense of the Past: Essays in the History of Philosophy*, 277–296. Princeton, NJ: Princeton University Press.
Wolf, Susan. 2012. "One Thought too Many: Love, Morality, and the Ordering of Commitment." In *Luck, Value, and Commitment: Themes From the Ethics of Bernard Williams*, edited by Ulrike Heuer and Gerald Lang, 71–92. USA: Oxford University Press.

https://doi.org/10.1080/00455091.2017.1415105 Published online by Cambridge University Press