Chapter 9: Against the Standard Conception of Idealization

Chapter 8 outlined the standard conception of idealization. In this chapter, we shall reject it. Recall that the standard conception has four dimensions: rationality, information, coherence and reasonableness. Above (Ch.8:III) I declined to criticize the idea of the reasonable within public reason, so this chapter criticizes the standard conception of rationality, information and coherence. But it offers these criticisms against the three dimensions as a whole—what I shall refer to as the ‘standard conception’ of idealization. I shall offer three arguments against the standard conception: (i) it is incoherent, (ii) it violates integrity, incurring first-personal costs and (iii) it artificially restricts reasonable pluralism, incurring second-personal costs. Like my arguments in Chapter 6, these three arguments are comparative in that they show that moderate conceptions of idealization should be ranked above the standard conception. Chapter 9 has five sections. Section I outlines the argumentative structure of the chapter. Section II develops the claim that the standard conception of idealization is incoherent because agents with extreme cognitive powers will face a number of processing obstacles. Section III shows how the standard conception of idealization incurs substantial first-personal costs with respect to less radical forms of idealization and Section IV shows the same for second-personal costs. Section V concludes.

Section I: The Argumentative Strategy

In Chapter 6 I both dispelled the benefits of the standard conception of public reasons and showed how it incurs higher first and second-personal costs than alternative conceptions. To show this, I ranked conceptions vis-à-vis one another according to the first and second-personal metrics. In short, I argued that for any conception of public reasons that contains accessibility, shareability or symmetry requirements, there exists a conception of public reasons without the requirements that ranks above it on first and second-personal grounds. Thus, there always exists an eligible conception of public reasons without accessibility, shareability or symmetry that ranks above any conception that contains them. As a result, there is decisive reason to reject each requirement.

This argumentative strategy must be modified if we are to reject the standard conception of idealization. Determining the first and second-personal costs associated with the standard
conception is difficult because it is unclear how agents idealized in the standard way could function. In response, I begin the main argument of the chapter by turning this challenge into a criticism of the benefits of standard conception. The primary reason to adopt the standard conception is that it approximates the ideal of rationality at the core of public reason liberalism. If adding information, cognitive power and coherence to a belief-value set removes errors in reasoning, then public reason liberals should hardly resist the relevant upgrades. Yet if the upgrades are so extreme as to make agent cognition opaque, the benefits of radical idealization are undermined. Radically idealized agents may face insoluble decision-making problems. Consequently, it will not form part of a viable version of public reason liberalism. Thus the first stage of the argument dispels the benefits of the standard conception by showing that it runs into problems. It will thereby lack net benefits even without appealing to first and second-personal costs.

The second stage of the argument delineates the costs of the standard conception vis-à-vis more moderate conceptions of idealization. It proceeds in two steps, first citing the first-personal costs of the standard conception and then citing the second-personal costs. The first-personal costs arise from revisions to an agent’s belief-value set. I shall argue that revisions should be understood as having pro tanto first-personal costs. Consequently, as revisions increase, the first-personal costs of the conception of idealization that requires such revisions increase. Conversely, revision can increase an agent’s liberty to act on her projects and plans since correcting incoherence, irrationality and ignorance can increase an agent’s ability to successfully achieve her aims. For this reason, laws passed in accord with some models of idealization can extend liberty. That said, when increasing idealization leads to the revision of citizens’ core beliefs and values, the benefits of revision for liberty vanish and the costs of revision explode.

Second-personal costs arise to the extent that idealization reduces reasonable pluralism. Public reason liberalism is motivated by the recognition of reasonable pluralism as a pervasive feature of modern social life. Conceptions of idealization that minimize reasonable pluralism in order to generate agreement will therefore threaten one of public reason liberalism’s core features. Further, while many public reason liberals suppose that idealization increases the degree of agreement, it may not (Ch.6:V). Increasing idealization might increase the disagreement amongst parties to a social decision. Although it is no gain to public reason if idealization increases reasonable pluralism beyond its present levels, should idealization significantly reduce the degree of reasonable pluralism among the relevant bargainers or deliberators then idealization incurs second-personal costs. Thus, either radical idealization generates no useful increase in determinacy—in which case it
is pointless—or it increases determinacy at the expense of reasonable pluralism, thereby incurring second-personal costs. If the standard conception reduces reasonable pluralism, then it should be ranked below conceptions of idealization that do not. Consider, though, that idealization might reduce reasonable pluralism enough to generate determinate principles of justice but only reduce reasonable pluralism to a moderate degree. In this case, public reason liberals could use such a conception of idealization to get traction of the problem of public justification while still recognizing reasonable pluralism. If so, such a conception of idealization would have net second-personal benefits. In fact, this is a live possibility but we shall see below that such conceptions are likely to be moderate conceptions rather than standard conceptions.

Thus, the main argument of Chapter 9 has three aims: (i) dispel the benefits of the standard conception by showing that it risks incoherence, (ii) demonstrate the first-personal costs of the standard conception, and (iii) demonstrate the second-personal costs of the standard conception. By the end of the chapter, I hope to have shown that in accomplishing these aims, we have produced reason to rank moderate conceptions of idealization over the standard conception. The stage will then be set for Chapter 10.

Section II: Arguments from Incoherence

Agents idealized in the standard way cannot represent real-world individuals because they are too cognitively distinct. While idealized agents are not identical to real-world agents, they must still confront and solve the same types of problems in order to generate reasons for action for the citizens they model. An idealized agent is only a plausible model of reason attribution if its reasons rank choices vis-à-vis a similar set of alternatives. Accordingly, I will argue that idealized agents can make choices over highly individuated actions but that real-world cognition often requires choices over plans or groups of actions. Further, idealized agents make choices by selecting options from a ranking of a maximal set of available alternatives; real-world individuals select options from a more limited set of options. Thus when idealized in the standard way, agents are not confronted with the same problems as real-world agents. Since idealized agents face different challenges and alternatives, they cannot make recommendations for how a real-world agent should make choices and so cannot provide real-world agents with bona fide reasons for action. Idealized agents generate reasons to take certain actions that real-world agents could never take or could take only irrationally. In this
way, the standard conception of idealization cannot achieve the purpose for which it is employed and is thereby incoherent.

II.i John Pollock’s Critique of the Standard Conception

To demonstrate the incoherence of the standard conception, I will rely on arguments against classical decision theory advanced by epistemologist John Pollock. Classical decision theory concerns fully informed and fully rational choices given an internally coherent set of preferences. Since classical decision theory bears significant analogies to agent choice under the standard conception, Pollock’s criticisms of the former should apply to the latter. Pollock’s criticism of classical decision theory holds that it rarely applies to the decision-making of real world agents. His arguments therefore provide evidence of a gulf between the cognition of real-world agents and their idealized counterparts. This distinction motivates our general conclusion that the standard conception of idealization is incoherent because it cannot count as an idealization of a real-world agent at all.

Classical decision theory is equivalent to what Pollock calls the “optimality prescription”, the view that a rational agent is one who always makes an optimal choice or chooses an action with the highest expected utility. Given that agents idealized in the standard way have perfect reasoning power, full information and coherent preferences and beliefs, they cannot help but make optimal choices. So why should unidealized agents care about the recommendations of the optimality requirement? Pollock suggests a possible reply:

Although theories of ideal agents are not directly about how real agents should solve decision problems, a plausible suggestion is that the rules of rationality for real agents should be such that, as we increase the reasoning power of a real agent, insofar as it behaves rationally its behavior will approach that of an ideal rational agent in the limit.

Ideal agents always make what Pollock calls warranted choices, choices that are justified after an agent completes all possible relevant reasoning (Ch.7:IV). Warranted choices look normative for real-world agents because they are made on the basis of better reasoning.

---

1 Pollock 2006, p. 15.
2 Ibid., p. 4.
Pollock maintains that real agents cannot and should not always aim to make warranted choices. One reason for this is that reasoning for cognitively sophisticated agents like human beings is “non-terminating”.

Because agents must act, they never complete all relevant reasoning. Further, real agents must reason “defeasibly”, meaning that the future acquisition of additional information or the performance of additional reasoning may rationally necessitate the agent’s changing its mind. Perfectly idealized agents do not reason defeasibly because they can complete all relevant reasoning given all relevant information. Since real-world agents are constrained by time, we can only require that they perform a “respectable amount” of reasoning before they act. Thus the recommendations of idealized agents will be rooted in reasoning and information that real-world agents are too constrained to utilize. This implies that real-world agents would often be irrational to make warranted choices. Pollock offers the following example to illustrate. Assume that P and Q are logically equivalent but that a real-world agent does not know this given her present reasoning. Now assume that the agent has reason to believe that the probability that P is true is 66%. Suppose that choosing P is both justified and warranted. Also assume that the agent is unable to determine the probability of Q, as she has not completed enough reasoning to even have a justified belief about the probability of Q. For this reason, “it would be irrational for the agent to accept a bet that Q is true at 2:1 odds. That choice would not be justified” despite the fact that the choice is warranted since Q is logically equivalent to P. Therefore the rational actions of real agents and ideal agents diverge.

While many believe that the rationality of real agents or “real rationality” should approximate ideal rationality, Pollock argues that a theory of rational choice for a real agent is only correct if it describes how a real rational agent should act.

Second, the cognitive limitations of real agents require that we make choices about plans, rather than actions. We often imagine that our regular decision-making ranges over acts, but Pollock denies that this is always true. In many cases, we choose among various plans and leave open the details about the acts involved in order to begin implementing the plan. We do this in part because we are often unaware of the conditions we will face when we implement the plan. Pollock argues:

For example, I don’t normally plan ahead about which traffic light to use on a particular leg of my route. I decide that in light of the flow of traffic around me as I drive. Decisions about

---

3 Ibid.
4 Ibid.
5 Ibid., p. 5. Also see Gaus 2010, pp. 246-9.
6 Pollock 2006, p. 5.
the details of my plans are often best left until the last minute when I will know more about the circumstances in which I am executing the plan. To accommodate this, plans are typically somewhat schematic.\(^7\)

Further, agents may be required to choose before they can figure out the details of the plan. Pollock gives the example of being invited to give a talk in Copenhagen in nine months. He may have to make a fast decision despite having no idea what flight he will take, for instance; he must decide to work out the details later when he has both “more time and a lighter cognitive load.”\(^8\) For this reason, real agents should adopt plans and fill out the details later, but ideal agents can immediately complete the relevant cognitive task. Real agents thus make choices over plans, not over actions and need reasons to choose one plan over another, rather than reasons to select an optimal act.

Real agents also cannot make optimal choices like ideal agents can. The optimality prescription requires choice over the set of all possible actions, but it does not specify how to determine the set of alternatives. A fully rational and fully informed agent \textit{automatically} detects the set of alternatives and ranks them, but real agents must expend resources delineating the set. One specific obstacle to determining the set is determining what is possible to do. The optimality prescription presumes that we can perform actions “infallibly” or with “action omnipotence”.\(^9\) Idealized agents possess action omnipotence because they automatically know what they are able to do. But real agents cannot always do what they think they can do. Instead, real agents formulate choices between what they should \textit{try} to do, not what they \textit{can} do with perfect reliability. A critic could reply that ideally rational agents will face the same problem, but this is only so if the ideally rational agent has limited information. The ideally rational agent will know the probability of failure and integrate these probabilities into her expected utility function. Real agents cannot do this.

A final difficulty for the optimality prescription is that it presupposes that real agents can choose actions or plans in isolation from one another. For ideal agents, actions and plans can be disaggregated and ranked, but real agents cannot make such determinations. To illustrate, imaging trying to implement a simple plan; to make a rational choice, we must compare the expected utility of the plan to relevant alternatives. But what is the alternative? One might think that an agent should compare implementing the plan to not implementing it. But the real world agent rarely faces a binary option as she is always implementing a host of other plans. Pollock notes that “trying to execute

\(^7\) Ibid., p. 11.
\(^8\) Ibid.
\(^9\) Ibid., p. 144.
previous plans may affect both the probabilities and the utilities employed in computing the expected utility of the new plan.” For instance, achieving one plan may diminish the expected utility of another. We must choose whether to implement plans in the context of the real-world decision maker. Thus the probabilities and utilities of plans are conditional upon the situation of the decision-maker. As Pollock argues, “there isn’t just one possible situation the decision maker might be in, because the other plans will normally have their results probabilistically.” Thus an agent may have to choose to implement a plan not even knowing what plan should be implemented as an alternative. Pollock argues that “local plans” should be organized in accord with “master plans”. Master plans are not universal plans for all circumstances of life. Instead, they’re plans for achieving higher-order values. Pollock thinks these plans are general enough to compare in terms of expected utility. We do not optimize over master plans because there are too many. Real decision-makers must start choices “by finding good plans” that can later be improved upon. Even one’s master plan changes over time. The rules of real rationality are rules directing the evolution of plans, “not rules for finding a mythical endpoint.” Real agents must be evolutionary planners not action optimizers.

Real-world agents are differ dramatically from the idealized gods of the standard conception. They would often be irrational to follow the recommendations of an ideal agent given their current knowledge. The ideal agent would make choices over actions rather than plans and so provide irrelevant recommendations that range over actions of too fine a grain. Ideal agents optimize but real agents often should not; ideal agents are action omnipotent, and real agents are not. Further, real agents live in a world of plans that cannot be clearly disaggregated and compared. They thus make choices based on conditional probabilities of staggering complexity. Thus, it seems clear that while the idea of an ideally rational agent is coherent, taking the standard of the ideally rational agent as normative for a real-world agent is wrong-headed. In fact, it runs the risk of incoherence since the context of choice for the ideally rational agent is substantially different from the contexts of choice for the real agent. We can draw a useful analogy from economist F. A. Hayek. Hayek argued that the central problem within economics is figuring out how to cope with limited information and cognitive powers. Mainstream neoclassical economics, on Hayek’s view, assumes that individuals have knowledge of each other’s preferences and how to meet them. In other words, equilibrium

---

10 Ibid.
11 Ibid.
12 Ibid., p. 184.
13 I shall bracket the complex matter of whether satisficing or settling for a less than optimal choice is in fact a method of optimization given cognitive limitations.
14 Hayek 1945.
modeling is rooted in the assumptions of full rationality and full information. But for Hayek, humans are essentially ignorant of important facts about others and the world. Economics should investigate the conditions of real human beings with limited cognitive capacities and recommend institutions that will allow them to achieve mutual benefits from cooperation. Neoclassical economics simply fails to ask the most foundational question: how do individuals of limited rationality, information and time cooperate effectively? Similarly, the optimality prescription and thus classical decision theory do not help real agents deal with uncertainty, ignorance and cognitive limitations encountered in the real world. Part of being a good real-world agent is coping with these problems. In fact, one of the points of publicly justifying moral norms is that they help to solve these problems for real world agents. Thus, generating citizens’ reasons in accord with the standard conception of idealization misses perhaps the most crucial set of problems that morality is supposed to solve—how moral norms help us to deal with uncertainty, mistrust, ignorance and irrationality or, again, how to cooperate in such a world for mutual benefit. Thus reason-attribution within public reason should not be rooted in the standard conception.

II.i: Ideal Advisor Accounts as an Alternative

A critic could reply to the arguments of II.i by advancing an ideal *advisor* account of reasons. The ideal advisor possesses all the features of the standard ideal agent, but instead of representing citizens, she simply *advises* them about what their reasons for action are. She generates an agent’s reasons by asking what she would do were she in its circumstances. Peter Railton has defended an ideal advisor account of reasons. Reasons are rooted in what Railton calls “objectified subjective interests”, interests articulated by an agent idealized in the standard way; call this agent A+, an adviser to unidealized agent A. The objectified subjective interest is “not what *be* [A+] currently wants, but what he would want his nonidealized self A to want—or, more generally, to seek—were he to find himself in the actual condition and circumstances of A.”\(^\text{15}\) One might wonder why we cannot ask an idealized A what *he* wants. In short, who cares about A+? The problem is that if we idealize A, the interests he endorses “might be quite different owing to the changes in idealization.”\(^\text{16}\) This is just what II.i claims. However, Railton admits that A+ might have similar problems advising A, since A+ may not be able,

\(^{15}\) Railton 2003, p. 11.

\(^{16}\) Ibid., p. 37.
… to set aside entirely his desire in his present circumstances with regard to himself or to A in considering what he would want to want were he to be put in the place of his less-than-ideal self. … His interests may perturb his psychology in ways that alter the phenomenon we wish to observe.\textsuperscript{17}

Railton seems to recognize the types of challenges discussed in II.i but he does not think they are “sufficient ground for skepticism”; instead, we need a good “perturbation theory” to account for such effects.\textsuperscript{18} But since Railton does not address the deep problems raised by Pollock, I submit that the problems raised in II.i do indeed provide sufficient ground for skepticism.

One might reply that the ideal advisor would recognize the Pollock problems and correct for them. Thus she could avoid the optimality prescription in the hopes of better taking A’s perspective. The ideal agent would not merely imagine herself in the same environmental circumstances as her nonideal counterpart, but in the same cognitive circumstances as well. There is a problem for this reply. If A+ is indeed idealized in the standard way, it looks like she would be irrational to place herself in the cognitive circumstances of A when advising A. Why shouldn’t she appeal to all the reasoning available to her? One might respond that she would otherwise give A bad advice, but A+ will certainly disagree since by definition she can see a vastly large range of considerations than A. Isn’t getting optimal (or at least good) advice precisely the reason we should appeal to A+ in the first place? If A+ submits to the pressure to ignore this extra reasoning, I submit that she makes herself a spare wheel. If she thinks the way A thinks, then it seems obvious that she will simply recommend that A do whatever he would do otherwise. I submit, therefore, that the ideal advisor account of reasons cannot save the standard conception.

Section III: The First-Personal Costs of the Standard Conception

With the benefits of the standard conception dispelled, we may outline its costs. Section III focuses specifically on showing how the standard conception imposes integrity-costs. I will first disaggregate the first-personal costs and benefits of idealization and then show how both are functions of increasing idealization. First-personal costs of idealization increase as idealization increases, whereas

\textsuperscript{17} Ibid.
\textsuperscript{18} Ibid.
the first-personal benefits of idealization increase as idealization increase up to a point and then begin to decline. When the two cost curves are combined, they show how the net benefit of idealization pushes away from the standard conception and towards a more moderate conception of idealization. To optimize the first-personal net benefit of idealization, the standard conception must be rejected.

III.i: First-Personal Costs and Benefits

I argued above (Ch.4:V.iii) that a first-personal cost is imposed upon citizens when they are denied opportunities to act in accord with the core first-personal reasons, those tied to their projects and principles. But idealization does not bar action; rather, it attributes reasons to citizens. How can idealization impose first-personal costs if does not obstruct action? Idealization specifies what reasons citizens have which seems conceptually prior to determining whether citizens may act on their reasons. Nonetheless, idealization can be understood as imposing first-personal costs to the extent that it requires revision of one’s belief-value set. Revising one’s beliefs and values should be understood as costly because the point of idealizing is to treat citizens in accord with their reasons and thereby extend their ability to act on their projects and plans and cooperate with one another. Idealization must be somehow tied to citizens’ real-world interests. The more citizens’ reasons are altered the less it seems they should be concerned with the recommendations made by their idealized counterparts. Following Gilbert Harman, I shall suppose that there is a presumption against revising one’s belief-value set. Harman suggests the following related principles of belief revision:

*Clutter Avoidance.* One should not clutter one’s mind with trivialities.

*Interest Condition (on theoretical reasoning):* One is to add a new proposition P to one’s beliefs only if one is interested in whether P is true (and it is otherwise reasonable for one to believe P).\(^{19}\)

We should avoid acquiring trivial beliefs and accept a presumption against belief formation that can only be met by an interest because “one can handle only a limited amount of … tentative acceptance

---

\(^{19}\) Harman 1986, p. 55.
since one can engage in only a limited amount of inquiry at one time.” Harman’s conception of cognizing as a real agent is similar to Pollock’s conception, as both want to understand good reasoning in terms of boundedly-rational agents solving concrete problems in real-time. The Clutter Avoidance Principle and Interest Condition are justified by an appeal to economizing on costs of cognition for just these sort of agents. In this case, interests are curiosities motivated by reasons to meet particular goals. For instance, one can have an interest in one’s environment just from wanting to familiarize one’s self with it. More relevant for our purposes, Harman postulates an “Interest in Facilitating Practical Reasoning” which holds that “If one desires E and believes M’s being true would facilitate or hinder E, one has a reason to be interested in whether M is true.” We can understand this interest as an interest in extending the effectiveness of one’s ability to act with integrity. For similar reasons, we have an interest in avoiding relevant inconsistencies. Harman’s Interest Condition already requires making minimal changes to one’s belief-value set, that is, of revising one’s beliefs only as much as one’s interests require. Due to its similarity with the Pollockian conception of real-world reasoning, Harman’s principles of belief revision seem in need of little additional defense. Agents with limited cognitive and informational resources have an interest in belief revision but while minimizing the cost to changing their cognitive faculties. Thus, to return to our central topic, one challenge for idealization is to show that the first-personal costs associated with belief revision are worth paying.

The first-personal benefits of idealization are the extension of an individual’s ability to act on her core projects and plans, or to live an integrated life. In Harman’s terms, the benefit is meeting the interest in facilitating practical reasoning. In many cases, contradictory beliefs, ignorance and poor reasoning frustrate citizens’ ability to achieve their ends. They will develop views and take themselves to have reasons that undermine their values and projects. The benefit of revising beliefs is that the integrity of citizens is extended. Thus, moving from populism to more moderate idealization values has significant first-personal benefits. On the other hand, this benefit of idealization will at some point begin to diminish and eventually become a cost itself. Imagine our system of beliefs and values linked in a Quinean web, with core beliefs and values at the center, and other beliefs located at the periphery. I take idealization to begin revision at the periphery, restructuring beliefs and values that are of less central importance to an individual’s life and those that contradict beliefs and values of greater weight. But as idealization increases and the

20 Ibid., p. 53.
21 Ibid., p. 55.
harmonization of one’s belief-value set increases, we proceed towards the core of the web and may begin to alter and disrupt core beliefs. Perhaps at high degrees of idealization, some individuals will realize that their religion is based on faulty evidence, all things considered. A conception of idealization radical enough to undermine the core projects associated with one’s religious faith would thus attribute reasons to citizens that would directly contradict their real-world core values and principles. It is such revisions that raise the objections of figures like Wolterstorff and Gauthier. Such revisions are so significant and radical that it is unclear whether the real-world individual has any reason to identify with her idealized counterpart as the counterpart could have projects and plans diametrically opposed to her own.

Perhaps core projects, principles and values are retained at high levels of idealization. But I shall argue in Section IV that increasing idealization often leads to normalization, that is, it can lead to wider agreement among agents. In fact, this is one of the motivations for radical idealization in the first place. If so, radical idealization threatens to destroy reasonable pluralism in part by annihilating the diverse values and aims of a free and equal citizenry. By idealizing in the standard way, citizens’ integrity will be largely annihilated as their projects and plans will be homogenized with the projects and plans of others. The benefits of extension become a cost. Thus, the revision costs of idealization steadily increase as we idealize, whereas the benefits increase to a certain margin and then start to decrease. At some point the costs and benefits will intersect, but they will arguably intersect at a degree of idealization far below the standard conception. We can now model the first-personal costs and benefits graphically.

III.ii: Optimizing First-Personal Benefits

As we have seen (III.i), the first-personal costs of idealization are the costs of revising one’s belief-value set. We need a positive reason to change our beliefs; the default position is to retain our existing commitments. The first-personal benefits of idealization are that they extend our ability to act on our projects and plans. Idealized agents are employed by philosophers to attribute reasons to citizens to act consistently with their projects and plans. If we act on our reasons, we will therefore have more integrity not less. Treating citizens in accord with such reasons will increase their opportunities to develop their integrity, not decrease them. However, this is true only to a point. I will now represent the first-personal costs and benefits of idealization, as displayed in Figure 1.
Figure one displays two curves. The curve in blue is the first-personal cost curve. It represents revision costs as a function of the degree of idealization. At populist levels of idealization on the left of the curve, there are minimal or no revision costs. At moderate levels of idealization, there is a slow up-tick of revision costs as peripheral beliefs and values are revised. But as revision mechanisms work their way into citizens’ core beliefs and values, revision costs skyrocket. When a core value or belief is revised, it reverberates throughout one’s peripheral beliefs and values. Further, it substantially changes the identity of a citizen, displacing vast sets of reasons. The curve in red is the first-personal benefit curve. The first-personal benefits represent the benefits of extending a citizen’s reasoning as a function of idealization. As idealization increases, reasons at the periphery of an agent’s belief-value set are revised, making them more consistent with her core values and principles, which is an increase in benefit. However, beyond moderate levels of idealization, core beliefs and values are revised. Consequently, the benefit of increasing a citizens’ ability to act on her real-world projects and plans is sharply diminished as she is idealized so as to lack them.

I assume that the marginal benefits of idealization first increase and then decrease, whereas the marginal costs of revision consistently increase. As a result, the cost and benefit curves will
intersect. The intersection point is the point where one additional unit of benefit achieved from idealization is no longer worth paying one additional unit of costs. At this point, the marginal benefits of idealization equal the marginal costs. The intersection point arguably lies substantially to the left of the standard conception of idealization, as the standard conception has both revision and extension costs. Nonetheless, the benefits of extension push away from populism as well, moving the intersection point to the right, or towards the standard conception. In this way, optimizing the benefits of idealization relative to its costs shows the superiority of a moderate conception of idealization over both populism and the standard conception. I will leave it until Chapter 10 to outline just what moderate idealization involves but for now we can see how to deal with the concerns raised by Wolterstorff and Gauthier in Chapter 8. In short, when idealized in the standard way, citizens have no reason to care about the reason-attributions made by their idealized counterparts. Recall Wolterstorff’s question: “What does the fact that a person would not believe what he does believe if he were fully rational have to do with how he should actually be treated in the political sphere?” At the standard margin of idealization, the answer is “nothing”. Yet at lower margins of idealization the objection evaporates. The reason that citizens should identify with the reason attributions derived from their moderately idealized counterparts is because their moderately idealized counterparts generate reasons that, if acted upon, will extend their real-world projects and plans and advance their real-world values more effectively.

Section IV: The Second-Personal Costs of the Standard Conception

Note that idealization is typically employed by public reason liberals in order to generate determinate principles of justice. For instance, both Rawls and Gauthier use the standard conception of idealization to generate their preferred principles of justice. Without full idealization, the parties and bargainers might disagree so much that convergence on principles of justice could not be reached. Thus idealization, public reason liberals assume, will lead to agreement or to what Fred D’Agostino called “normalization”. D’Agostino remarks that Rawls’s original position reasoning is normalization in the sense that members of the public find a way to choose a collectively binding mechanism of ranking proposals about how to shape the basic structure of their society. If individuals were not idealized in the standard way, they might know their identities. If they did, “it is

---

23 D’Agostino 2003, p. 100.
unlikely, in the extreme, that any agreement on substantive matters would be possible. The preferences of diverse individuals must be “translated” into a collective ranking of options that makes the problem of political justification tractable.

The second-personal argument against the standard conception takes the form of a dilemma. If the standard conception normalizes too much, it will have net second-personal costs, that is, the costs of reducing reasonable pluralism will exceed the benefits of determinacy. On the other hand, if the standard conception fails to produce normalization, then the motivation for idealizing in the standard way evaporates.

IV.i: Second-Personal Benefits

Before proceeding with the dilemma, we must ask whether idealization has any second-personal benefits. The only benefit discussed in the literature that can be construed as second-personal is the putative increase in determinacy. By yielding determinate political principles, a conception of public reason enables individuals to produce a society where all can freely advance their values; individuals can converge on norms that can coordinate their behavior and help them secure the benefits of cooperation. Determinacy is a massive benefit because it is crucial to solving the problem of social cooperation among diverse individuals with differing ends, the central problem of social philosophy. But to justify the radical degree of normalization generated by the standard conception, we must show that determinacy can only be secured by the standard conception. The public reason literature lacks an argument to the effect that determinacy can only be achieved in this way. Further, a recent alternative conception of public reason developed by Gerald Gaus employs evolutionary mechanisms to generate determinate political principles from the reasoning of moderately idealized citizens. I cannot comment on the viability of this project but the Gaussian alternative suggests that idealizing in the standard way may not be the only way to achieve determinacy. In Chapter 10, I will sketch a number of ways in which moderate idealization can achieve normalization. In this way, the second-personal benefits of normalization can be dispelled.

IV.ii: Second-Personal Costs

---

24 Ibid.
If standard conception of idealization does not produce normalization, it has no benefits. I have already suggested some reasons why the standard conception of idealization may not lead to normalization (Ch.6:V); in short, the burdens of judgment imply that reasonable pluralism will arise even among agents idealized in the standard way. We have also seen that the standard conception may not have unique benefits over more moderate conceptions of idealization as it may not be the only way to achieve determinacy. To establish the second-personal dilemma for the standard conception, we now only need to show that normalization generates problematic second-personal costs.

As we saw above (Ch.8:I.i), Rawls’s conception of idealization normalizes the parties to the original position until they agree. The obvious worry about such normalization is that it generates political principles by artificially restricting the considerations that are significant to citizens. There are real dangers to restricting reasonable pluralism in this way. Christopher Eberle worries that any proposed idealization threatens to attribute reasons to citizens based on the biases of the theorist whose conception of idealization is being employed. For instance, Eberle suggests that Rawlsian idealization opens the door to Rawlsians imputing their preferred political positions to citizens who would stridently oppose them. He compares appeals to improved rationality and information as appeals to “false consciousness and ideological blindness” that have been historically abused by those who cannot gain the actual assent of the coerced. Almost hyperbolically, Eberle suggests that,

… one can also image the justificatory liberal assuring the fundamentalist parent that she’d find acceptable a policy that “encourages” her school-age children to be more open-minded about other religious if only she were slightly better informed about what’s good for her children.26

Idealization risks permitting the public reason liberal to “purify” the belief-value set of citizens in ways that suit her political agenda. Eberle sees no nonarbitrary way to avoid this problem,

Any claim about what idealized citizens would believe is uneliminably dependent on an implicit appeal to what those in much less than ideal conditions would believe, and any such appeal can—and will—be met by competing claims about what idealized citizens would believe, which are in turn dependent on an implicit appeal to a different set of claims formed.

in less than ideal conditions. The appeal to what idealized citizens would believe is, in short, a circuitous way for nonidealized citizens to express their disagreements with one another.  

The danger of normalization, on Eberle’s view, is not merely that reasonable pluralism is reduced but that it is reduced in a biased and oppressive fashion. When figures like Rawls idealize members of the public, they must make decisions about how to alter belief-value sets and these decisions risk being arbitrary. As a result, the public reason liberal will simply build her values into the values of those she is idealizing; this might lead to the sort of pernicious oppression Eberle points to above. Recall that Isaiah Berlin expressed Eberle’s concern long before:

For—so Hegel, Bradley, Bosanquet have often assured us—by obeying the rational man we obey ourselves: not indeed as we are, sunk in our ignorance and our passions, weak creatures afflicted by diseases that need a healer, wards who require a guardian, but as we could be if we were rational. … In this way the rationalist argument … has led by steps which, if not logically valid, are historically and psychologically intelligible from an ethical doctrine of individual responsibility and individual self-perfection to an authoritarian State obedient to the directives of an elite of Platonic guardians.  

Both Eberle and Berlin see second-personal costs accompanying the standard conception of idealization. If the standard conception leads to normalization, it not only restricts reasonable pluralism, but turns liberal political theory into a rationale for authoritarianism. The second-personal cost of normalization is not merely that reasonable pluralism will be restricted but that such restriction will lead citizens to be coerced by the state in ways that are obviously unacceptable to them. If Berlin and Eberle are correct, then conceptions of public reason that idealize in the standard way can hardly count as conceptions of public reason liberalism at all.

The main motivation for idealizing in the standard way is determinacy, but the second-personal cost of the normalization implied by the standard conception is too high. If the standard conception does not lead to normalization—and it very well might not—then there is no reason to radically idealize members of the public. Thus either the standard conception imposes significant second-personal costs or it is unmotivated. That said, merely citing second-personal costs is not

---

27 Ibid., p. 233.
enough to defeat the standard conception of idealization. We must look to the relevant alternatives to show that a viable conception of idealization is in the offing. In my view, the only serious alternative to the standard conception is some form of moderate idealization. I will show in Chapter 10 that while moderate idealization gives up some determinacy, the risk of overly restricting reasonable pluralism is significantly reduced. If I am correct, the standard conception of idealization will rank below more moderate conceptions of idealization on the second-personal metric. Coupled with the other criticisms advanced in this chapter, there will be significant reason to prefer a moderate conception of idealization over the standard conception.

Section V: Conclusion

After reviewing the argumentative strategy of this chapter in Section I, Section II dispelled the benefits of idealizing in the standard way by showing that it risks incoherence. Section III argued that the standard conception of idealization imposes significant first-personal costs on citizens; consequently, there is strong first-personal reason to rank the standard conception of idealization below more moderate conceptions. Finally, Section IV argued that idealizing in the standard way imposes significant second-personal costs on citizens, thus the second-personal metric supports moderate conceptions of idealization over the standard conception as well. When applied to the problem of idealization, the dissertation’s argumentative strategy provides very strong reason to reject the standard conception of idealization so long as a moderate theory of idealization is viable.

Chapter 9 is the analogue of Chapter 6—it rejects the standard conception of idealization, where Chapter 6 rejected the standard conception of public reasons. In doing so, it too sets the stage for the next chapter. We saw in Chapter 7 that the convergence conception of reasons avoids the problems leveled against the standard conception of public reasons. Chapter 10 will develop a conception of moderate idealization that avoids the problems associated with the standard conception of idealization. Thus Chapter 9 is the penultimate piece of the puzzle of the dissertation. In Chapter 10, we develop the final element of the alternative conception of public reason promised in Part I. The theory of moderate idealization will complete our attempt to reconcile social morality and individual ideal in the role of religion debate.
References


