

# Revising Sarah Conly's "Coercive Paternalism"

## 1 Introduction

In *Against Autonomy*, Sarah Conly (2012) makes a provocative argument for coercive paternalism, i.e. "laws that force people to do what is good for them" (p. 3). Conly's paternalism has an important limitation: paternalist legislation must be designed to "help us reach *our own* goals" (p. 12, emphasis added). The government does not dictate what those goals must be according to some objective value; rather, legislation accords with the goals and values that people choose for themselves, or what Conly has termed their 'subjective welfare.' Of foremost importance is that paternalist legislation respect people's desired ends while regulating means. If Sue desires long life, and if not wearing her seatbelt undermines that desire, the government will legislate the means (seatbelt laws) to Sue's ends (long life). In order to make a case for subjective welfare as the basis for coercive paternalism, Conly criticizes views about objective value, especially the view she labels 'welfare perfectionism.'

This paper argues that Conly's 'subjective welfare' view can be subjected to the very criticisms she levels against welfare perfectionism. Thus, if welfare perfectionism is incompatible with coercive paternalism, so is Conly's view of subjective welfare. In order for coercive paternalism to be viable, Conly will need an account of welfare that accords with paternalism, or else she will need to revise her thesis about paternalist legislation. In section 2, I proceed by first examining welfare perfectionism's compatibility with coercive paternalism. I give three options for how the coercive paternalist might account for welfare perfectionism. In section 3, I exposit Conly's summary of and reasons for rejecting welfare perfectionism as a basis for coercive paternalism. Section 4 gives Conly's proposed alternative, subjective welfare. In section 5, I argue that given Conly's reasons for rejecting welfare perfectionism,

her position is untenable. Conly cannot maintain a commitment to subjective welfare and coercive paternalism – as she understands it – at the same time. I conclude with ways she might revise her account.

## **2 Welfare Perfectionism and Coercive Paternalism**

On Conly's interpretation, welfare perfectionists maintain that there are objective goods that must be obtained for a person to achieve human excellence. If a person fails to achieve these standards of human welfare, then his welfare is defective, even if he is satisfied with his life. Usually, welfare perfectionists base their notions of objective goods on the nature of humankind (p. 105). Because it is human nature to be sociable and desire long life, some common examples of objective goods are fulfilling relationships and good health. Now think of the stoner alone in his basement. He may find his life satisfactory and pleasurable, but the welfare perfectionist would argue that he is lacking some important aspects of a good life. If the stoner has no fulfilling relationships, or if his health is poor (perhaps due to overuse of marijuana, overeating, or some other health issue), then his welfare is limited, even if his life brings him intense pleasure.

When considering the welfarist view vis-à-vis Conly's coercive paternalism, the implication is that if an activity or good is truly necessary for human welfare, the paternalist government must legislate that such things be pursued – even if people hate those things. In other words, the government determines a person's ends and also legislates the means to those ends, no matter what the person actually desires.

Many will find such a conclusion objectionable; some will not. But supposing that we are dedicated to coercive paternalism, there are at least three options to consider:

- 1) Accept welfare perfectionism and its implications for paternalism, including the implication that the government legislate the pursuit of objective goods (even if people hate those goods);

2) Qualify coercive paternalism so as to complement welfare perfectionism in a way that would not require people to pursue goods they hate, or in a way that limits any unpalatable implications;

3) Reject welfare perfectionism as a basis for coercive paternalism and substitute something else.

Conly argues for the third option. She finds the implication of the first option counter-intuitive, in that forcing people to do something they hate seems wrong. She does not seriously consider the second option; I will outline my own arguments for it at the end of this essay. In the next section, I will consider Conly's reasons for rejecting welfare perfectionism.

### **3 Conly's Objections to Paternalist Welfare Perfectionism**

Conly gives a number of objections to welfare perfectionism in general, including the objection that it fails the test for intuitive plausibility (p. 108). Though Conly does not spell out exactly what such a test entails, it seems to be straightforward: if a view seems intuitively implausible, it should not be seriously considered. She prescribes this test because she believes that since we all disagree about value, we should only consider those which have intuitive appeal to the majority of people. Presumably (Conly says little on this point), welfare perfectionism does not pass this test for precisely the reasons the test was prescribed – we disagree about value, and it is unlikely that an objective account of welfare can be given. We will see whether her account of subjective welfare passes the test for intuitive plausibility in a later section. For now, we will focus on objections that Conly raises against welfare perfectionism – not in itself, but as a basis for coercive paternalism.

She levels two such objections. First, Conly believes that because people disagree in reasonable ways about what welfare consists in, imposition of legislation that supports any one particular view of welfare will err (p. 112). That is, paternalist legislation that accords with any one particular account of

human welfare will get people's welfare wrong in some cases. But Conly is not exactly clear what she means by 'err.' The most likely possibility is that she is making a kind of epistemological objection. Because human knowledge is limited, any list of proposed objective values is bound to get some of those values wrong. In other words, there is room for rational disagreement over what it means to be human and which activities should be pursued. For instance, Aristotle did not include humility on his list of human excellences, while philosophers like Aquinas do. Both had good reasons for their lists of goods. But as Conly comments, some lists will inevitably include the personal preferences of those making the list (p. 110). Therefore, if paternalist legislation is passed on the basis of such lists, some of that legislation might actually inhibit people's welfare by mandating goods that are not actually objective. To summarize the objection: welfare perfectionism remains inherently flawed because of widespread disagreement about what goods should count towards objective welfare. Paternalist legislation based on flawed, 'objective' lists will therefore err.

Her second objection is that coercive paternalism actually works against the aims of welfare perfectionism (p. 112). She argues that, on most perfectionist accounts, it is important that people actually have the right motives and affections for the end being pursued (p. 111). Within a system of coercive paternalism, the government would only be able to coerce certain types of behavior; it could not force people to enjoy those behaviors. Thus, a key component of perfectionism will be impossible for a paternalist government to achieve. In other words, her second objection to welfare perfectionism is that its aims cannot be reached through paternalist legislation, period. Notice, this argument says nothing about the overall strength of welfare perfectionism. Rather, applying coercive paternalism to welfare perfectionism as a means to establishing people's objective welfare is self-defeating. Instead of instilling attitudes and affections for the ends being sought through legislation, coercive paternalism would breed hatred for those ends (p. 107).

A simple reply from the welfare perfectionist might be: so much the worse for coercive paternalism (p. 106). But we must remember that if we suppose one's commitment to coercive paternalism should be prioritized over one's commitment to views about welfare, then Conly's argument implies that we must search for another view of welfare on which to base paternalism. Thus, a more nuanced counter-argument must be sought that takes coercive paternalism seriously. We will return to possible responses in section 5.

Since Conly is committed to coercive paternalism, she is led to reject welfare perfectionism as its basis for the two reasons mentioned. In its place, she offers subjective welfare. She believes that as a basis for paternalist programs, a theory of subjective welfare does better at honoring the needs, goals, and values of individuals (p. 112).

#### **4 Conly's Account of Subjective Welfare**

Given these objections, we must see if Conly's account of subjective welfare fares any better. But first, her view will require some unpacking. On her view, the standard of welfare should be the *desires* of the individual subject, or what each person thinks is good for him- or herself (p. 14). But these subjective desires are a standard only insofar as they relate to the goals and ends chosen by the individual. Conly is not exactly clear on what desires count as 'ends,' but she thinks these desires must be stable preferences of the individual – goals that are consistently sought, though the means of achieving those goals may be confused or variable (p. 124).

Furthermore, she seems to think that these desires must be informed in some sense, though she acknowledges the limits of information's influence on rational deliberation (p. 25). That is, Conly believes information is important, but no clear dividing line exists between informed desires and ignorant ones. Likewise, even with all of the experimental data on practical irrationality, "there is no

identifiable point at which we go from being purely rational thinkers to completely irrational ones” (p. 6). It is a difference of degree, not kind. Still, it is easy to see why information matters. Presumably, if a person knows for certain that she will get cancer and suffer a great deal from smoking before she even smokes her first cigarette, her desire to smoke will be greatly diminished. Yet, while information plays a role in our desires, it is not always the deciding factor, either. People may know that smoking cigarettes greatly increases their risk of disease, yet choose to smoke anyways.

Whatever counts as valid ends, Conly believes people tend to make the most mistakes when they deliberate over means. Thus, paternalist legislation will attempt to short-circuit cognitive errors so people will arrive at their desired goals by way of means that do not undermine their goals (p. 124). That is, the aim of coercive paternalism founded on subjective welfare is to help individuals achieve their own ends.

## **5 Critique of Subjective Welfare and its Consequences for Coercive Paternalism**

Conly, then, wants subjective welfare to be based on the desired ends of the individual. Yet, we have seen that she is not entirely clear about when a person’s desire counts as an appropriate ‘end.’ In trying to distinguish the ends relevant to a person’s welfare, a multitude of questions present themselves: Should we give more weight to the desires that an individual would have at the end of her life? Or, should we prioritize the desires that an individual currently endorses? Does the individual get to decide these questions, and if so, what if her answers change? If she does not get to decide, who does? More questions could be asked, but the point in raising these questions at this early stage is simply to prime us for the difficulties inherent in weighing which desires should count as ends.

Recall that Conly seems to say that those ends that promote a person’s stable preferences (health, long life, retirement savings, etc.) should always be prioritized over ends that only promote short-term

goals (e.g. enjoying an enormous Thanksgiving meal, the pleasure of smoking cigarettes for their own sake). Indeed, she usually construes short-term ends as means to long-term ones (Purshouse, 2014, p. 369-70). Further, she seems to prioritize informed desires, though she admits the limits of information. In any case, she seems to limit subjective welfare to a set of ends with which an individual may not agree in the moment. Sure, Sally wants health, but not if this means that she will not be able to enjoy her imprudently proportioned Thanksgiving meal for its own sake.

In any case, Conly's account lacks clarity. In the next two subsections, I attempt to clarify her account by considering two possible ways of interpreting her theory. The first will be an interpretation that equates subjective welfare with a qualified subset of a person's desires, or in some sense orders those desires for the person. Let us call this the *qualified-desires* interpretation. The second interpretation will honor the actual desires of a person as comprising the person's total subjective welfare. Let's call this the *actual-desires* interpretation.

I argue that the first interpretation of Conly's subjective welfare view fails to provide a sufficient basis for coercive paternalism for precisely the same reasons as welfare perfectionism. Conly leveled two objections against welfare perfectionism *as a basis for* coercive paternalism. One, perfectionism will incorrectly identify some goods as 'objective' because of the limits of human knowledge. Thus, legislation that aims for conformity with one particular perfectionist account will probably make mistakes in the cases of some proposed objective goods. Two, given perfectionism, coercive legislation seeking to impose objective goods on the governed will fail to be accepted and internalized by everyone; thus, it will not achieve its goal (namely, the internalization of those values). I argue that the second interpretation is no more intuitively plausible than objective welfare.

## **5.1 Subjective Welfare as a Qualified Subset of a Person's Desired Ends**

To see why Conly's theory cannot avoid her own critiques of welfare perfectionism, we might ask Conly to clearly distinguish which of a person's subjective ends count towards his welfare, and which ends do not. That way, a coercive paternalist can know which of the individual's ends must be taken seriously in legislation. But if Conly were to do so, she may undercut her theory before it even gets off the ground. If she limits a person's welfare to some subset of that person's desires – as Conly has suggested, perhaps those that are stable, well-informed preferences – then Conly's ordering of desires may in fact conflict with the person's actual, expressed ends. In other words, if Conly were to propose a standard that imposes an ordering on the set of Jim's actual, expressly desired ends, he may resent or disagree with such an ordering. In fact, such an ordering would seem to be an imposition of either: 1) Conly's personal preferences; or 2) Higher-ordered values which do not reflect Jim's *expressed* desires and values. Either way, if Conly accepts the *qualified-desires* interpretation of subjective welfare, she cannot to avoid the very critiques she levels against the welfare perfectionist, no matter the standard she uses to qualify those desires.

Though unlikely, if she admits to the first case, where this ordering is simply her personal preference (or when lived out paternalistically, the preferences of a government), then a number of consequences would seem to follow. The most obvious is that Jim may not agree with Conly's preferences, and resent any legislation that tries to tell him what his ordered ends should be.

For instance, say Jim is a smoker who loves smoking for its own sake. In other words, he does not use smoking as a means to some other end; rather, he loves the smell of a new cigarette, he loves the feeling of putting a cigarette to his lips, and he loves drawing deeply from it for that first puff. He would not want to give it up for anything. But Jim also has a desire to be healthy. He exercises, eats appropriately, and takes good care of himself. He desires both of these ends equally; we might further stipulate that he is well-educated about the effects of smoking. Nevertheless, especially in the case of

smoking, Conly argues that folks like Jim have not ordered their stable preferences correctly (p. 169). Thus, Jim needs help, even if he resents it. But Jim might very well argue that Conly is simply imposing her own personal preferences on him. Sure, smoking might shorten his life, but he'd rather live by the old adage: "life without risk is not worth living." In any case, if he were forced not to smoke, he would hate it. He could also argue that not being able to smoke actually hurts his welfare, because as Conly herself has admitted, he is the one who is best positioned to decide his ends.

It is more likely that Conly would argue that she is not simply forcing on him her own personal preference, but instead respond along the lines of option #2. She might say that Jim has made systematic cognitive errors that undermine his subjective welfare (p. 113-19). Perhaps she responds: "Look, Jim, even if you *really* want to smoke, it has the real potential to undermine your ability to pursue lots of other things you really value. And that's what justifies forcing you not to smoke." In other words, Conly might say she is doing what is best for Jim, given his entire set of desires. Notice, the good being promoted here is not some objective notion. Rather, it points to Jim's own desires and his own cognitive error, thus ordering his desired ends in a way that he has not been able to.

But this response still does not answer the central question being asked: if Jim orders his own ends differently than someone else does on his behalf, on what grounds can that person insist that Jim is wrong about his ordering? Jim does not have a higher-order desire to want *not* to smoke, so he does not order health above smoking. Conly has appealed to some of Jim's other desired ends and claims that Jim has made cognitive errors. But Jim is intelligent; he knows the risks. If Jim is aware of his own thoughts on this matter and is presented with the evidence, yet still chooses to smoke, it would seem that the discussion must end. Indeed, if he were not allowed to smoke, it would keep him from pursuing his own ends. Yet, especially in the case of smoking, Conly seems unwilling to end the discussion there. But if she stands firm here and argues that Jim should not be allowed to smoke despite his expressed desires,

she would probably have to make at least an implicit appeal to some values outside of Jim, perhaps empirical evidence. That is, she would need to appeal to subject-neutral evidence and assign that evidence value that is not subject-dependent. But these values would begin to look suspiciously like objective values, thus making the qualified-desires type of subjective welfare subject to the very arguments she made against the welfare perfectionist. Of course, Conly's other option would be to let Jim's values stand as they are, honoring his welfare for what it is. In other words, she may choose not to limit Jim's welfare to some subset of his desires. We will consider this second type of subjective welfare in a moment.

But before doing so, we will notice that Jim's counter-arguments in this section are the very arguments that Conly uses against the welfare perfectionist. Jim believes he will hate being forced not to smoke cigarettes. In other words, he believes he will not internalize the values being forced upon him. Secondly, Jim argues that in his case, imposition of values alien to him actually hurts his welfare.

But Conly might counter that although these consequences *are* regrettable, they can be avoided. Indeed, she has given herself enough leeway to make this move. But in so doing, she has also provided an escape for the welfare perfectionist. In chapter two, she argues that part of the strength of paternalism is that it uses laws to habituate people to choose good ends. She uses Aristotle's discussion of virtue and education to argue that people must be trained to do "the right sorts of thing" (p. 70). People have a natural ability to be virtuous and vicious; therefore, they need good teachers – good laws – to help them develop proper habits of character. Once their tastes for doing right are developed, they are less inclined to be tempted by alternatives. Although it remains a possibility that they might be led astray, these laws will continue to function as guides to right conduct.

Returning to Jim, it may be too late for him since he seems to have developed vicious habits of decision-making, i.e. he may never form a desire not to smoke. But this does not mean that legislation

against smoking will necessarily fail to cause internalization in other people of the values it seeks. In fact, it may serve as a means to inculcate these values in the next generation, so that this generation will not be tempted to desire deviant ends. But even if they are so tempted, legislation will remain in place to keep them from acting on those temptations. But not that if this response to Jim's objection is successful, it is equally so for the welfare perfectionist. Indeed, the welfare perfectionist can argue that, in general, coercive legislation successfully instills in the population the values it seeks to uphold.

But how can Conly avoid Jim's second argument that imposition of alien (objective) values actually hurts his welfare? As we have already seen, if she simply claims that it does not, she would be arguing along the lines of the welfare perfectionist. Instead, she might simply bite this bullet and claim that, although Jim's welfare *is* hurt by this imposition of values, these values are actually good for *most* people. Without getting into the details of her proposed cost/benefit analysis and utilitarianism, we simply note that this option is available to Conly. In other words, some type of utilitarian thinking may be used to bolster her argument in favor of imposing values that might hurt some people's welfare.

But practically speaking, there may be little observable difference between the welfare perfectionist and the subjective welfarist (of the *qualified-desires* type) when applying such thinking to coercive paternalism. That is, taking such a line may greatly weaken her argument against the perfectionist. Take the case of cigarette smokers like Jim. Presumably, the perfectionist would appeal to some type of scientific evidence or objective value to justify legislating against smoking. Notice, these are the same values to which Conly may appeal. Like Conly, the perfectionist can argue that such legislation is worth it. Though the perfectionist would ultimately only need to appeal to an objective list rather than utilitarian thinking or cost/benefit analysis, the results would be the same: people like Jim will claim that their welfare is being damaged. Just as Conly admits that such legislation might indeed hurt Jim's welfare, welfare perfectionists can be open to the possibility that they have gotten an

objective good wrong. Maybe not smoking is *not* an objective good. But for the perfectionist and the subjective welfarist alike, the objection of individuals will not be the deciding factor.

## 5.2 Subjective Welfare as a Person's Actual Desired Ends

We have seen that if Conly tries to limit a person's desires to a qualified subset, such ordering looks eerily similar to the reasoning of the perfectionist. In any case, subjective welfare encounters the same problems as the welfare perfectionist when taken as a basis for coercive paternalism. But Conly may choose simply to honor a person's desired ends for what they are, without qualification. Is this understanding of welfare intuitively plausible?

Consider the case of Jill. Say Jill has a desire to consume alcohol every evening to the point where she gets drunk. Jill is inherently antisocial and feels anxious around people, so she usually drinks alone. She believes that drinking calms her nerves, and loves the feeling of losing control when drinking. She believes life is something to be endured, and has no desire to live long. At the age of 40, she gets cirrhosis of the liver and dies. On her death bed, she says that it was all worth it – no regrets. If Jill's welfare consists of her actual desires and expressed ends, then it would seem Jill lived a good life. But this seems like an intuitively implausible view of Jill's welfare.

To see why this might be so, we can consider what might happen if Jill were to be shown what her life could have been without the deleterious effects of alcohol. When shown the meaningful relationships she might have had, the successful career, the positive feelings, and other things that she gave up for alcohol, she herself might be persuaded that her life with alcohol was not as good as she thought. She might be like an Ebenezer Scrooge, who felt quite fine with his desire for money, but when shown what he might miss out on, repented of his desires and changed his life. But if this is the case, then a person's welfare is not simply bound up with her actual expressed desires in the moment. That leads back to the qualified-desires option.

But perhaps seeing what her life may have been like will *not* be enough for Jill, just like it was not enough for Jim to consider not smoking. Still, most people would probably see a life of alcoholism and depression as something that undermines Jill's welfare, no matter what she thinks. In other words, considering Jill's actual desired ends as the only things conducive to her welfare seems to fail Conly's test for intuitive plausibility. But I hasten to add that Conly does not spell out exactly what she means by a "test for intuitive plausibility." If Jill's example does not show how an actual-desires account of subjective welfare fails this test, Conly will need to offer more explanation as to why this is so. But whatever Conly might mean, it does seem true to our intuitions to say that Jill's obtainment of her actual desired ends is not sufficient to establish her welfare. But I do not expect to settle the issue by a simple appeal to intuition; Conly does not either (p. 193). My point is only to show that if welfare perfectionism is to be rejected because it is intuitively implausible, so too might the *actual-desires* interpretation of subjective welfare be rejected.

## 6 Conclusion

In conclusion, Conly's account of subjective welfare seems unable to successfully evade the very critiques she levels against welfare perfectionism. Rather, where welfare perfectionism encounters problems, so does the subjective welfare view. I argued for this view by offering two interpretations of Conly's theory: either Conly accepts an *actual-desires* account of subjective welfare, or she accepts a *qualified-desires* account. On the one hand, an *actual-desires* account seems intuitively implausible, just as Conly claims of welfare perfectionism. On the other hand, a *qualified-desires* account leads to an imposition of alien values and failure to achieve its goals, much like welfare perfectionism. The upshot is that whatever account one might choose to adopt, subjective welfare is liable to the same critiques as objective welfare.

Perhaps there is another, more viable view of welfare that may serve as a basis for coercive paternalism. If so, the closing suggestion of this essay is for Conly to argue for a more viable view of welfare that might act as a basis for coercive paternalism (which would still remain faithful to the third option given in section 2 of this paper). But as far as I can tell, there is not a more viable view. If not, this would seem to raise a serious issue for the plausibility of coercive paternalism as Conly presents it.

But this leads us back to Conly's unexplored option #2. A better route for Conly may be to qualify coercive paternalism in a way that would make welfare perfectionism more palatable as its basis. Specifically, it should not legislate ends that are questionable. 'Questionable' ends are ends for which we have no empirical reasons for correlating with individual welfare. These empirical reasons would come from the realms of what Dan Haybron (2008) calls 'prudential psychology.' With the advancements being made in prudential psychology, social scientists are beginning to learn a lot more about human welfare and happiness. In short, I am proposing that paternalist governments would use this evidence as an 'objective' standard for legislating ends. Indeed, Haybron argues this research may have some promising implications for public policy and government legislation. A positive argument for this view will need to wait for future papers.

Still, citizens being forced to do what they hate may be unavoidable on any view of government we take. If this is the case, and if Conly is right, the best alternative would seem to be to force people to do what is good for them. We may sometimes get these things wrong, but progress in the social and psychological sciences may be a better foundation on which to base legislation than Sally Citizen's views about her own subjective welfare. Thus, welfare perfectionism based on empirical science may be Conly's best hope for grounding coercive paternalism.

## Bibliography

- Conly, Sarah. 2012. *Against Autonomy: Justifying Coercive Paternalism*. Cambridge: Cambridge University Press.
- Conly, Sarah. 2014. "Against Autonomy: Response to Critics." *Journal of Medical Ethics* 40, (5): 354-356.
- Haybron, Daniel. 2008. *The Pursuit of Unhappiness: The Elusive Psychology of Well-Being*. Oxford; New York: Oxford University Press.
- Oshana, Marina. 2014. "Against Autonomy: Justifying Coercive Paternalism." *Ethics* 124, (2): 392-397.
- Pugh, Jonathan. 2014. "Coercive Paternalism and Back-door Perfectionism." *Journal of medical ethics* 40, (5): 350-350.
- Purshouse, C. 2014. "Against Autonomy: Justifying Coercive Paternalism." *PHILOSOPHY* 89, (2): 367-371.
- Wertheimer, Alan. 2014. "Against Autonomy?" *Journal of Medical Ethics* 40, (5): 351-351.