Parfit: A Defense of Consequentialism?

Consequentialism:
> C1: There is one ultimate moral aim: that outcomes be as good as possible.
> C2: What each of us ought to do is to make the outcome best.
> C5: The best possible motives are those which it is true that, if we have them, the outcome will be best.

T is *indirectly collectively self-defeating* when it is true that, if several people tried to achieve their T-given aims, these aims will be worse achieved.

C2 may be indirectly collectively self-defeating. If we were all *disposed to try* to make the outcome as good as possible—were *pure do-gooders*—then we might produce a worse outcome than we would if some of us had different dispositions.

Why?
> 1. Happiness comes from having and acting on strong desires: love for other people, desires for work, desires for play. Pure do-gooders either cannot have, or must often suppress, these desires.
>
> 2. Violating a commonsense rule like, "Don't kill," "Don't lie," etc. may, in rare circumstances, produce the best outcome. But we may well be mistaken, or give in to wishful thinking, about when such circumstances obtain. If have strong aversions to violating these rules, which allow us to violate them only when we believe that they will produce *much better* outcomes, then we may avoid such mistakes and so produce better outcomes.

However, this does not mean that, if any of us becomes a pure do-gooder, he or she will produce a worse outcome. So long as enough others are *not* pure do-gooders, becoming a pure do-gooder may produce the best outcome.

Does the fact that C2 is indirectly collectively self-defeating mean that C2 is incoherent? No. C2 tells us (other things equal) to acquire the best motives, as described by C5. What we have seen is that sometimes, the best motives do not involve a disposition to try to follow C2. In such cases, C2 will simply tell us *not* to become disposed to try to follow C2.

Suppose that my becoming a pure do-gooder *will* produce a worse outcome than remaining someone who loves his family, etc. Then sometimes, because I favor my family, I may knowingly produce a worse outcome.
- On the one hand, I act wrongly.
- On the other hand, it would be wrong of me to lose the motives that led me to act wrongly.

Although my action is immoral, my character is not, and it would be immoral to try to change it. If we blame people for their character, then I should not be *blamed*. I am a *blameless wrongdoer*.

This might allow us to reconcile the theoretical appeal of utilitarianism with commonsense intuitions (as Darren suggested).

It might even produce the best outcome *not to believe* C2. In this case, C2 would tell us (roughly) not to believe it! C2 would be *self-effacing*.

Would it be an objection to C2, if it were self-effacing?